US Valuation of Health Outcomes Measured Using the PROMIS-29

Benjamin M. Craig, PhD, **Bryce B. Reeve, PhD**, **Paul M. Brown, PhD**, **David Cella, PhD**, **Ron D. Hays, PhD**, **Joseph Lipscomb, PhD**, **A. Simon Pickard, PhD**, and **Dennis A. Revicki, PhD**
Health Outcomes and Behavior, Moffitt Cancer Center and University of South Florida

University of North Carolina at Chapel Hill, North Carolina

University of California, Merced, California

Northwestern University, Chicago, Illinois

University of California, Los Angeles, California

Emory University, Atlanta, Georgia

University of Illinois at Chicago, Illinois

Outcomes Research, Evidera, Bethesda, Maryland

Abstract

Objectives—Health valuation studies enhance economic evaluations of treatments by estimating the value of health-related quality of life (HRQoL). The Patient-Reported Outcomes Measurement Information System® (PROMIS) includes a 29-item short-form HRQOL measure, the PROMIS-29.

Methods—To value PROMIS-29 responses on a quality-adjusted life year (QALY) scale, we conducted a national survey (N=7557) using quota sampling based on the US 2010 Census. Based on 541 paired comparisons with over 350 responses each, pair-specific probabilities were incorporated into a weighted least-squared estimator.

Results—All losses in HRQoL influenced choice; however, respondents valued losses in physical function, anxiety, depression, sleep, and pain more than those in fatigue and social functioning.

Conclusions—This paper introduces a novel approach to valuing HRQoL for economic evaluations using paired comparisons and provides a tool to translate PROMIS-29 responses into QALYs.

Keywords

Quality-adjusted life years; discrete choice experiments; patient-reported outcomes

Address correspondence to: Benjamin M. Craig, PhD, Moffitt Cancer Center,12902 Magnolia Drive, MRC-CANCONT,Tampa, FL 33612-9416. Phone: (813) 745-6710; Fax: (813) 745-6525; benjamin.craig@moffitt.org.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

INTRODUCTION

To inform resource allocation decisions and patient guidelines, comparative effectiveness research (CER) aims to "provide evidence on the effectiveness, benefits, and harms of different treatment options" including the differences in health outcomes.(1) Other events may coincide with health outcomes, such as economic (e.g., cost), clinical (e.g., disease), and humanistic outcomes (e.g., privacy).(2) However, this study solely focuses on enhancing the measurement and valuation of health outcomes for CER.

All measures of health outcomes record duration (e.g., In the past 7 days, I felt depressed). Although health status (e.g., Do you currently feel depressed?) may be useful to diagnose a disease, formulate a prognosis, or indirectly capture health outcomes, health status does not quantify the burden of an outcome without further information. Continuing the example for CER, the likelihood of reporting current depression may be different between two interventions, but this prevalence information may not be decision-relevant unless the duration and frequency of depressive symptoms can be taken into account.

Due to its chronologic reference, outcomes evidence is more informative than health status evidence, yet outcomes evidence alone may be not sufficient to inform decisions, particularly when alternative treatments have distinct advantages. To resolve such dilemmas, an understanding of outcome value is required (i.e., preference-based weights or tariffs). Discrete choice experiments (DCEs) enhance our understanding of health outcomes by asking respondents to choose between alternatives (e.g., 1 week of depression vs. 1 week of pain). Such choices define the relative value of treatment outcomes and facilitate treatment recommendations. This study expresses the value of health outcomes along a common metric for CER, namely quality-adjusted life years (QALYs).

Among health outcomes, a QALY represents a year with no health problems and serves as the fundamental unit of measurement in outcomes research. All other health outcomes represent a loss in health-related quality of life (HRQoL) from this standard, inherently reducing a person's quality-adjusted life. Valuation studies, like this one, are typically designed to identify the value of outcomes in terms of lost QALYs (e.g., a year feeling sometimes depressed equals a loss of 0.26 QALYs). The debate over this numéraire began with its introduction in 1970(3) and remains heated,(4) particularly in the United States.(5) Nevertheless, no other numéraire has achieved comparable notoriety in the summary of health outcomes, such as patient-reported outcomes (PRO) measures.

The Patient-Reported Outcomes Measurement Information System® (PROMIS) includes publicly available generic profile HRQoL measures.(6) These standardized PRO measures complement clinical findings on patient health (e.g., blood pressure) and epidemiologic evidence in the community setting (e.g., viral infection rates) for CER. The PROMIS measures provide scores for multiple HRQoL domains; however, they do not summarize outcomes across domains. By incorporating DCE evidence, health valuation studies summarize outcomes across domains by weighing losses in HRQoL in terms of their influence on choice. In addition, such preference elicitation tasks can ask respondents about the tradeoff between losses in HRQoL and lifespan. For example, the paired comparison in

Figure 1 involves a tradeoff between 10 years sometimes depressed and a loss of 1 QALY. Using responses to 541 pairs like this one, this study directly estimate the value of PROMIS outcomes on a QALY scale.

Using a dataset with both PROMIS scores and EQ-5D responses, Revicki et al. derived regression equations that mapped PROMIS scores to QALYs.(7) This indirect approach is analogous to predicting the value of a house from past sales in the neighborhood. No study has directly elicited preferences for any PROMIS outcomes to derive values on a QALY scale. Furthermore, most health valuation studies focus on instruments that have 1 item per domain (e.g., EQ-5D) or that reduce evidence from multiple items to 1 attribute per domain (e.g., SF-6D). The use of 1 attribute per domain simplifies the valuation task, but this reduction sacrifices the psychometric advantage of improving measurement reliability.(8) Ideally, a health valuation study will directly assess preferences (i.e., no mapping) and summarize all measured outcomes (i.e., no reduction). To exemplify this, this study values the entirety of the PROMIS-29, which includes 4 items on 7 domains as well as an 11-level Pain Intensity scale. Never before has such a large instrument been valued.

Aside from being the first to value the PROMIS-29, this is the first national study that uses online DCE for health valuation. Previously, Bansback and colleagues conducted an online DCE to value the EQ-5D responses by recruiting from the IPSOS Canadian panel, but excluded French-speaking Canadians (e.g., Québécois).(9) Craig and colleagues recruited members of the Toluna United States panel to value SF-12v1 and SF-6D responses, but this national sample was heavily skewed toward older White women.(10) Viney and colleagues used an online best-worst scaling task (including death and a survival attribute) to value the EQ-5D from the perspective of the Australian general population.(11)The objective of this project was to estimate values for 10-year losses in HRQoL on a QALY scale described by the PROMIS-29, based on the perspective of adult members of the US population.

Given the extent of study details and the need to limit paper length, we provide a didactic appendix that reviews terminology in paired comparisons for health valuation, adjectival statements, pair selection (with all results) and an overview of econometric concepts. In complement to this appendix, we also provide STATA code, log, and data to allow reproducibility of the results within this paper.

METHODS

Theory Underlying Health Outcomes and Choice

A health episode is a description of HRQoL over a period of time and typically includes many health-related events (e.g., child birth) and outcomes (e.g., 1 week feeling sometimes depressed). The episodic random utility model (ERUM) was introduced in 2008 to describe the relationship between health episodes and individual choices, particularly ranking tasks. (12) ERUM specifies that the utility of a health episode is a function of the health-related quality and quantity of life with an additive error term, $U(h,t) + \varepsilon$ where h is HRQoL and t is duration (t>0). The probability of a choice between two independent episodes, A and B, depends on individual understanding of HRQoL domains and durations and may vary due to intrapersonal variability or respondent heterogeneity.(9, 12, 13) Alternatively, some studies,

particularly those based on time trade-off (TTO) tasks, have applied the instant random utility model (IRUM), which first divides HRQoL by duration before including an additive error term ($U(h/t) + \varepsilon$), simplifying episodes to health states (i.e., an instantaneous experience, h/t). IRUM describes the relationship between health states and choices and becomes unstable when duration becomes small (e.g., $t \to 0$).(12, 14)

While the concepts are sometimes used interchangeably,(15) this paper differentiates between value and utility. Value refers to a preference-based measure representing the choices of a group of individuals, V, and utility is a random latent trait at the individual level, which governs a person's choice (i.e., ERUM). The two concepts are linked, because value is inferred from the choices from multiple individuals. Specifically, episodes A and B have the same value ($V_A = V_B$) if and only if exactly half choose A instead of B (i.e., switching A and B has no effect on the aggregate's choice probability).

In Figure 1, respondents were asked to choose between 10 years sometimes depressed followed by death and fewer years with "no health problems" followed by death. As shown in Figure 2 (i.e., where the starred line crosses the 50% mark), the probability reaches 50% at around 2.6 years (i.e., V(sometimes depressed,10 years)=V(no health problems, 7.4 years)). This 50% point implies that the loss in HRQoL (sometimes depressed for 10 years) equals a loss of 2.6 QALYs. When more or less respondents choose A instead of B, this imbalance implies the extent of difference in value.

Application of DCE in Health Valuation

For the purposes of this study, all values are expressed on a QALY scale. Differences in QALYs are directly linked to choice probabilities using a cumulative density function (CDF): knowing a difference in QALYs predicts the choice probability, and knowing a choice probability predicts the difference in QALYs.

Continuing the example, Figure 1 includes a loss in HRQoL (d_A=sometimes depressed for 10 years) and a loss in lifespan (d_B=1 QALYs). Suppose you want to predict the choice probability in Figure 1 using the QALY results in Figure 2 and the CDF = $d_B / (d_A + d_B)$, where d_h is the decrement in value associated with the alternative, h. According to Figure 2, sometimes depressed for 10 years equals a loss of 2.6 QALYs. Therefore, placing this result in the CDF predicts that 27% prefer feeling sometimes depressed over losing 1 QALY (i.e., 1/(2.6+1)). Looking at the empirical data (Figure 2), the sample probability for this pair is actually 28%. Likewise, knowing a sample probability predicts the difference in QALYs. If the sample probability in Figure 1 is 28%, we can solve for d_A (i.e., 1/(d_A+1) =28% or d_A = 2.57 QALYs). The next, more challenging task is to combine evidence from multiple pairs.

All losses in HRQoL can be expressed as decrements in value on a QALY scale, d_h, using a multi-attribute utility (MAU) regression. By definition, each decrement, d_h, decreases the likelihood of choosing a particular health episode; however, its effect on choice is non-additive, depending instead on a CDF (e.g., $d_B / (d_A + d_B)$). For this study, we assumed that choice depends solely on the differential attributes between A and B (d_A and d_B), not on the attributes that they share (i.e., "pivot" or "scope"; see Appendix). Building from this

theoretical framework, this study was designed to estimate the independent values of the losses in HRQoL captured by the PROMIS-29 on a QALY scale.

Health Outcomes

The PROMIS-29 is quickly becoming a standard for PRO research and practice and recommended for initial outcome assessment.(16, 17) Studies continue to support its construct validity and feasibility;(18, 19) in fact, one study stated that it may be superior to the SF-36.(18) The PROMIS-29 includes seven HRQoL domains (Physical Functioning, Anxiety, Depression, Fatigue, Sleep Disturbance, Social Functioning, and Pain), and the pain domain has two subdomains (interference and intensity). Each of the 7 domains has four 5-level items (i.e., 16 decrements each). In addition to these items, pain intensity is assessed using a single 11-point numeric rating scale anchored between no pain (0) and worse imaginable pain (10), adding 10 additional decrements.(20) For use in DCE, PROMIS responses (e.g., sometimes depressed) were expressed as losses in HRQoL lasting 10 years followed by death and parameterized as 122 decrements in value on a QALY scale (i.e., $(7 \times 16) + 10$).

Survey Panels

This project recruited US respondents from multiple panel vendors, with each panel recruiting 1000 respondents with completed surveys.(21) We chose to employ multiple vendors in order to assess and compare costs, services, responsiveness, and quality of data. We separated survey hosting from recruitment activities to utilize multiple panels and to mitigate potential conflicts of interest. In an effort to maintain control over data quality, no panel vendor was allowed to host the survey; therefore, vendors were not able to invite respondents based on survey responses or alter or auto-generate responses. A single hosting company was used for all respondents, regardless of panel. All study procedures were approved by the University of South Florida Institutional Review Board (IRB # Pro00000076) and are described in greater detail in a report posted online and in the Appendix.(8)

Each panel company sent its members a generic e-mail invitation containing payment information and a member-specific hyperlink that provided immediate access to the survey informed consent page. Once a respondent clicked on the link, the member-specific data (e.g., birth date) were "passed through" and captured by the survey software in order to compare these demographic data with survey responses.

Survey Design

Pre-testing at Moffitt Cancer Center and the University of South Florida, as well as pilot work in health valuation using online DCEs, verified the feasibility and methodological approach for the study.(8, 22) Furthermore, these preliminary studies enabled understanding of the issues surrounding task complexity and the appropriateness of the attribute/levels. (23-25) After the consent page, respondents completed the screener, health, DCE, and follow-up components of the survey. Respondents were not allowed to proceed to the next page unless all questions on a page were answered. In the screener component, consenting respondents were asked 10 questions about their demographic, geographic, and

socioeconomic characteristics. If a respondent belonged to a filled demographic quota or met any of the four termination criteria (invalid country or state; discordant demographic responses; use of a proxy server; JavaScript disabled), he or she was disqualified from further participation.

The valuation of health outcomes requires an experimental design that accounts for the natural complexity of health and cognitive considerations of subjects. In this study, value was quantified by the likelihood of preference and was estimated using choice data on stated preferences over health outcomes. The health component included 49 questions derived from PROMIS items, which were modified and used with permission of the PROMIS Health Organization and the PROMIS Cooperative Group.(26, 27) To reduce response error, direction of health was fixed; best health was always placed on the left-hand side of the page.(28, 29) After a brief introduction of 3 paired comparisons, the DCE component consisted of 30 paired comparisons distributed over 4 sections.(10, 30)

The primary difference between the four DCE sections was their pivots. A pivot is the set of the attributes in common for both alternatives in a pair (a.k.a. holdouts).(10, 31) Within a DCE section, each pair had the same pivot, which was modified by adding two compensating attributes.(30) For the 6 lifespan, the pivot was 10 years with no health problems followed by death (see Appendix, Figure 1). For the 8 health pairs in the next 3 sections, the pivot was 10 years in Good, Fair, and Poor health followed by death, respectively (see Appendix). The duration of 10 years is conventionally used in TTO tasks as a compromise between avoiding proximal mortality (i.e., not too soon) and promoting realism for older respondents whose life expectancy may not exceed 10 years (e.g., age 100). A loading animation required that at least 8 seconds be spent on each comparison to assure sufficient time for page loading and to force respondents to spend a minimum duration on each page.

The follow-up component included 33 health, socioeconomic, and survey feedback questions and an open-text box for comments. Aside from dropping out of the survey (e.g., losing internet connection), respondents were terminated if JavaScript failed or if 2 or more hours passed since entry.

Pair Selection and Assignment

Each respondent in a panel was randomly assigned 1 of 1000 unique sequences of lifespan pairs and 24 health pairs based on his/her demographic characteristics (reported in survey and verified by vendor) to guarantee that each pair-specific sample corresponded to demographic quotas.(8)

The 6 lifespan pairs directed respondents to choose between episodes with either reduced lifespan or 1 of 6 "health problems" for 10 years, including 3 levels of depression (rarely, sometimes, or often feeling worthless, helpless, depressed, and hopeless) and 3 levels of mild pain (1, 2, or 3 on a pain scale from 0 [no pain] to 10 [worst pain imaginable]). Assigned in random sequence, these problems were selected to be severe enough to be worth a loss of lifespan (with "no health problems"), yet mild enough to not imply problems on other HRQoL domains. Each problem was compared to 10 losses in lifespan creating 60

pairs with 100 responses for each pair from each panel (1000 respondents × 6 responses/60 pairs=100 responses per pair), except for the third pain intensity level, which was compared to 11 losses in lifespan due to a coding error (Figure 2 and Appendix).

Assigned in random sequence, attribute order, and horizontal arrangement, the 24 health pairs were taken from a set of 256 item pairs and 224 domain pairs (see Appendix). Each item pair directs respondents to choose between a decrement in 1 item and a decrement in another item within the same domain (e.g., rarely hopeless vs. rarely helpless). Domain pairs trade a decrement in all items in 1 domain (e.g., Depression) for a compensating decrement in all items for another domain (e.g., Fatigue). The domain pairs inform the value of the domain decrements, and the item pairs allocate this value across the specific items within the domain. Under this approach, the addition of items to a domain has no impact on the value of the domain (i.e., no double counting). In this design, each of the 480 health pairs (i.e., 256 item and 224 domain pairs) has 50 responses per panel (1000 respondents × 24 responses/480 pairs=50 responses per pair; see Appendix for more details on pair selection).

Econometrics

Each of the 226,710 DCE responses (N=7,557 respondents × 30 responses) was incorporated into the calculation of the 541 pair-specific probabilities, $p_1 \ldots p_{541}$ (i.e., 61 lifespan and 480 health pairs). Given that we attempted to select pairs with population probabilities between 0.1 and 0.9 and pair samples were large (over 350 responses per pair), each sample probability is approximately normally distributed with standard error, $\sigma = sqrt(p \times (1-p) /n)$.(32, 33). Specifically, the standard error of each sample probability ranges from 0.016 to 0.026.

To estimate the 122 decrements in the MAU regression, d_h, we minimized the sum of squared error surrounding these sample probabilities, $\sum_{k=1}^{541} (P(A_k > B_k) - p_k)^2 / \sigma_k^2$, where P(.) is a CDF. Two specifications of P(.) were tested: $\ln(P/(1-P)) = \theta(d_B - d_A)$ and $\ln(P/(1-P)) = \ln(\theta d_B) - \ln(\theta d_A)$. The former specification is a logit model with a rescaling parameter, θ, and the latter is a relativity model, $P = d_B/(d_A + d_B)$, which has the advantage that θ factors out. These two specifications are compared based on their ability to predict the pair-specific probabilities in terms of least squared error (see STATA data, code, and log). Confidence intervals are estimated by percentile bootstrap with pair stratification and 1000 resampling iterations.

RESULTS

Between March 2012 and July 2012, we recruited 29,031 respondents across the 50 States and Washington, DC. Among the 29% who met the survey requirements (e.g., respondents were excluded once quotas were filled), 90% completed the survey with a median duration of 20 minutes (interquartile range of 16-28 minutes). Compared to the 90% who completed the online survey, the 10% with incomplete responses were younger, less educated, and more likely to be Black/African American (Table 1). Respondent characteristics in the analytic sample were largely similar to the 2010 Census, except for higher educational attainment.(34) Even though we did not use geographic quotas, the analytic sample includes

respondents from all 50 states and their proportions largely agreed with the 2010 US Census (Lin concordance 0.97). Across the 541 pairs, the differences between the weighted and unweighted probabilities were small (<0.004); therefore, only the unweighted results are shown. Compared to the relativity specification, the logit produced greater squared error (6519 vs. 2403) and more negative decrements (36 vs. 0); therefore, all results shown are based on the relativity specification.

Tables 2 and 3 provide the MAU estimates, including 122 decrements (i.e., decreases in QALYs attributable to losses in HRQoL over 10 years) and their confidence intervals. These decrements are non-negative and largely increase from best to worst, suggesting decrement acceleration. Figure 3 summarizes these decrements in terms of domain values (i.e., sum of all decrements within a domain). For Fatigue, Sleep, and Social Functioning, a shift from Level 1 (best) to Level 5 (worst) is less than 10 QALYs; however, such shifts in Physical Functioning, Anxiety, Depression, and Pain (interference and intensity) were largely considered worse than 10 QALYs.

The value of 10-year losses in HRQOL on a QALY scale can be calculated by adding together the 10-year decrements for the PROMIS-29 responses. For example, the mildest loss is no problems on all items, except "pain interferes a little bit with work around the home" (a decrement of 0.06 QALYs over 10 years). If we assume constant proportionality in time (with no health problems and with health problems) as well as no discounting, this mildest loss for 1 year has a value of 0.006 QALYs (0.06/10 years or 2.2 quality-adjusted days). In other words, such a year has a QALY value of 0.994 (1-0.006). On the contrary, 10 years with the worst responses on all items (i.e., pits) equals the sum of all 122 10-year decrements (94.58 QALYs). Under the same constant proportionality and no discounting assumptions, 1 year in pits represents a reduction from full health of 9.458 QALY (i.e., −8.458 QALYs; 1-94.58/10). Therefore, the range of 1-year values based on the PROMIS-29 is from 1 to −8.458 QALY.

To illustrate the distribution of 1-year values, we applied the 10-year decrements to PROMIS-29 responses from the health component of the survey and assumed constant proportionality and no discounting to produce the 1-year estimates. The colors indicate the distribution by self-reported general health: Excellent, Very Good, Good, Fair and Poor. It is important to note that the health component of the survey describes health for a week (not 1 year) and included "chores" as the 4th pain interference item, not "your enjoyment of life." For illustrative purposes, the responses are assumed to be the same (excluding or including this 4th item had no noticeable effect on Figure 4). Figure 4 also shows the mean, standard deviation, percent positive, median, and interquartile range. Clearly, the overall distribution is skewed with 28.2% below 0 and 10.6% below −1. Among those in fair and poor health, 31.6% and 74.0% are below −1, respectively.

DISCUSSION

This is the first study to directly value health outcomes based on the PROMIS measures. The PROMIS initiative has advanced the science of PRO measurement through instrument development using both qualitative and quantitative methods and application of modern

measurement theory methods. This study incorporated general US society perspectives using DCE methods to value multiple items within seven HRQoL domains of the PROMIS-29. On a QALY scale, respondent values suggest that physical function, anxiety, depression, sleep, and pain are more detrimental than fatigue and social functioning. In most cases, the worst decrement in each item was greater than all other decrements combined, emphasizing the importance of measuring poor health over good health.

While interview-based tasks (e.g., TTO) remain commonplace in health valuation, these tasks include an adaptive DCE process ending in a statement of indifference.(35) DCEs without adaption were applied in this study as an attempt build from valuation studies in other fields (e.g., conjoint analysis) and to measure health preferences in the community using the internet.

The approach to valuing multiple items per domain undertaken in this study provides an alternative to the development of HRQoL instruments specifically for health valuation, such as the Health Utilities Index Mark 3 (HUI-3),(36) EQ-5D,(37) and the Quality of Well-Being-Self Administered (QWB-SA) questionnaire.(38) The PROMIS-29 also differs from these preference-based instruments in conceptual framework and health domains covered. Although these instruments share a comparable construct of overall health cross-sectionally, (39) their variability in coverage likely influences their QALY predictions. (40, 41)

The multiple items per domain and calibration to the larger domain item banks create the possibility of incorporating more advanced psychometric scores directly into the MUA regression of the health valuation study. Score shifts may represent changes in the latent domain as a whole, and decrements of each item may represent the parts. Like incorporating interaction terms in the MUA, estimation with both score shifts and decrements may test whether the whole is greater than the parts.

The relationship between QALYs derived here for the PROMIS-29 and those of existing preference-based measures is unknown. More work is needed to demonstrate the advantages for CER of the potential improvements in measurement reliability and greater number of domains. For example, the EQ-5D includes mobility, self-care, usual activities, pain/ discomfort, and anxiety/depression, while the PROMIS-29 also includes a broader assessment of physical function, social function, sleep disturbance, and fatigue. In contrast, the HUI-3 takes a different perspective and includes attributes of vision, hearing, speaking, ambulation, dexterity, emotion, cognition, and pain. Confounding between domains may also cause double counting (e.g., sleep and fatigue) in health valuation, similar to the use of multiple correlated items within a domain. In this study, such confounding was controlled through the use of domain pairs: comparing bundles of attributes between domains so that the number attributes within the domains has no effect on the estimates.(42)

This study focused on valuing 10-year PROMIS-29 outcomes using online DCE and panels of US adults. All decrements in health lasted 10 years, a conventional duration used for TTO tasks; future studies should examine shorter and longer durations, as research suggests the respondent's age and duration of time horizon systematically impact valuations.(40, 43, 44) Great care was taken to verify the respondent qualifications as US adults (e.g., verifying

pass-through data, IP geolocation, and concordance of age/birth date responses), and we applied quotas at the pair-level to assure demographic representation of each pair-specific probability. However, unobservable characteristics concerning participation in panels may introduce biases, similar to other recruitment methods including random digit dialing, door-to-door interviewing, and postal invitations. In this study, we observed selection toward higher educated respondents compared to the US Census (Table 1).(34) Still, it is unclear whether these potential selection issues introduced bias in the decrement estimates (e.g., Is education related to pain preferences?). Future valuation studies may examine additional PROMIS items or domains; nevertheless, this study establishes a methodological foundation to examine expeditiously US health preferences and may be adapted to explore new populations, durations, and items.

The valuation results from this study have implications for the use of PROMIS for CER. In addition to identifying the effectiveness and costs of treatments and procedures in practice as opposed to clinical trials, CER can be used to ascertain whether the treatments and procedures are worth the expense. To achieve this goal, researchers and policymakers need to understand the value that people place on the health outcomes. Consistent with previous research, extreme forms of depression, anxiety and physical functioning are ranked as highly detrimental episodes of health.(45, 46) Likewise, social functioning and mild outcomes (e.g., walking up and down stairs) are less important compared to other domains and levels. The evidence from this study is a step toward developing a systematic way for researchers to assess the effectiveness of alternative interventions based on the value gained from improved health outcomes as assessed by PROMIS measures. This will greatly enhance our understanding of the relative merit of treatments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

REFERENCES

1. Agency for Healthcare Research and Quality (AHRQ). [cited December 3, 2012] What is Comparative Effectiveness Research. Available from: http://effectivehealthcare.ahrq.gov/index.cfm/what-is-comparative-effectiveness-research1/

2. Kozma CM, Reeder CE, Schulz RM. Economic, clinical, and humanistic outcomes: a planning model for pharmacoeconomic research. Clin Ther. 1993; 15(6):1121–32. Epub 1993/11/01. [PubMed: 8111809]

3. Fanshel S, Bush JW. A Health-Status Index and its Application to Health-Services Outcomes. Operations Research. 1970; 18(6):1021–66.

4. Lipscomb J, Drummond M, Fryback D, Gold M, Revicki D. Retaining, and Enhancing, the QALY. Value in Health. 2009; 12:S18–S26. [PubMed: 19250127]

5. Neumann PJ, Greenberg D. Is the United States ready for QALYs? Health affairs (Project Hope). 2009; 28(5):1366–71. Epub 2009/09/10. [PubMed: 19738253]

6. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol. 2010; 63(11):1179–94. [PubMed: 20685078]

7. Revicki DA, Kawata AK, Harnam N, Chen W-H, Hays RD, Cella D. 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. Quality of Life Research. 2009; 18(6):783–91. [PubMed: 19472072]

8. Craig, B.; Reeve, BB. Methods Report on the PROMIS Valuation Study: Year 1. 2012. Available from: http://labpages.moffitt.org/craigb/Publications/Report120928.pdf

9. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. J Health Econ. 2012; 31(1):306–18. [PubMed: 22197308]

10. Craig BM, Pickard AS, Stolk E, Brazier JE. US valuation of the SF-6D. Medical decision making : an international journal of the Society for Medical Decision Making. 2013; 33(6):793–803. Epub 2013/05/01. [PubMed: 23629865]

11. Viney R, Norman R, Brazier J, Cronin P, King MT, Ratcliffe J, et al. An Australian Discrete Choice Experiment to Value EQ-5D Health States. Health Economics. 2014; 23(6):729–42. [PubMed: 23765787]

12. Craig BM, Busschbach JJ. The episodic random utility model unifies time trade-off and discrete choice approaches in health state valuation. Popul Health Metr. 2009; 7:3. Epub 2009/01/16. [PubMed: 19144115]

13. Hadorn DC, Hays RD, Uebersax J, Hauber T. Improving task comprehension in the measurement of health state preferences - a trial of informational cartoon figures and a paired-comparison task. J Clin Epidemiol. 1992; 45(3):233–43. [PubMed: 1569420]

14. Craig BM, Busschbach JJV. Revisiting United States valuation of EQ-5D states. J Health Econ. 2011; 30(5):1057–63. [PubMed: 21835477]

15. Keeney, RL.; Raiffa, H. Decisions with Multiple Objectives: Preferences and Value Trade-offs. Cambridge University Press; New York, NY USA: 1993.

16. Adams K, Bayliss E, Blumenthal D, Boyd C, Guralnik J, Krist AH, et al. Universal Health Outcome Measures for Older Persons with Multiple Chronic Conditions. J Am Geriatr Soc. 2012; 60(12):2333–41. [PubMed: 23194184]

17. Forrest CB, Bevans KB, Tucker C, Riley AW, Ravens-Sieberer U, Gardner W, et al. Commentary: The Patient-Reported Outcome Measurement Information System (PROMIS) for Children and Youth: Application to Pediatric Psychology. J Pediatr Psychol. 2012; 37(6):614–21. [PubMed: 22362923]

18. Hinchcliff M, Beaumont JL, Thavarajah K, Varga J, Chung A, Podlusky S, et al. Validity of Two New Patient-Reported Outcome Measures in Systemic Sclerosis: Patient-Reported Outcomes Measurement Information System 29-Item Health Profile and Functional Assessment of Chronic Illness Therapy-Dyspnea Short Form. Arthritis Care Res. 2011; 63(11):1620–8.

19. Selewski DT, Collier DN, MacHardy J, Gross HE, Pickens EM, Cooper AW, et al. Promising insights into the health related quality of life for children with severe obesity. Health Qual Life Outcomes. 2013; 11

20. PROMIS-29 Profile v1.0. 2008-2012 PROMIS Health Organization and PROMIS Cooperative Group; Available from: https://www.assessmentcenter.net/ac1//files/pdf/44b7636201a34267a9213db7f69f2c6d.pdf [June 06, 2012]

21. Craig B, Hays R, Pickard AS, Cella D, Revicki D, Reeve B. Comparison of US Panel Vendors for Online Surveys. Journal of Medical Internet Research. 2013; 15(11):e260. Epub published online Nov 29, 2013. [PubMed: 24292159]

22. Craig B, Owens MA. Methods Report on the Child Health Valuation Study (CHV): Year 1. 2013 Available from: http://labpages.moffitt.org/craigb/Publications/CHVMethods_130917.pdf.

23. Ryan, M.; Gerard, K.; Amaya-Amaya, M., editors. Using Discrete Choice Experiments to Value Health and Health Care [electronic resource]. Springer; Dordrecht: 2008.

24. Kahneman D. A perspective on judgment and choice - Mapping bounded rationality. American Psychologist. 2003; 58(9):697–720. [PubMed: 14584987]

25. Bridges JFP, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint Analysis Applications in Health-a Checklist: A Report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value in Health. 2011; 14(4):403–13. [PubMed: 21669364]

26. PROMIS Assessment Center. [August 08, 2012] PROMIS Terms and Conditions. 2012. Available from: https://www.assessmentcenter.net/documents/PROMIS%20Terms%20and%20Conditions%20v8%20July10_2012.pdf

27. National Institutes of Health. [June 27, 2012] PROMIS: Dynamic Tools to Measure Health Outcomes from the Patient Perspective. Available from: http://www.nihpromis.org/#1

28. Alexandrov A. Characteristics of Single-Item Measures in Likert Scale Format. The Electronic Journal of Business Research Methods. 2010; 8(1):1–12.

29. Swain SD, Weathers D, Niedrich RW. Assessing three sources of misresponse to reversed Likert items. J Mark Res. 2008; 45(1):116–31.

30. Chrzan K. Using partial profile choice experiments to handle large numbers of attributes. International Journal of Market Research. 2010; 52(6):827–40.

31. Louviere J, Lancsar E. Choice experiments in health: the good, the bad, the ugly and toward a brighter future. Health Economics, Policy and Law. 2009; 4:527–46.

32. Urban, FM. The Application of Statistical Methods to the Problems of Psychophysics. Psychological Clinic Press; Philadelphia: 1908.

33. Urban, FM. Urban's Solution (Minimum Normit X^2). In: Bock, RD.; Jones, LV., editors. The Measurement and Prediction of Judgment and Choice. Holden-Day; San Francisco: 1968. p. 33-49.

34. Liu HH, Cella D, Gershon R, Shen J, Morales LS, Riley W, et al. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. J Clin Epidemiol. 2010; 63(11):1169–78. [PubMed: 20688473]

35. Luo N, Li M, Stolk EA, Devlin NJ. The effects of lead time and visual aids in TTO valuation: a study of the EQ-VT framework. The European journal of health economics : HEPAC : health economics in prevention and care. 2013; 14(Suppl 1):S15–24. Epub 2013/08/09. [PubMed: 23900661]

36. Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI (R)) system for assessing health-related quality of life in clinical studies. Ann Med. 2001; 33(5):375–84. [PubMed: 11491197]

37. Brooks, R.; Rabin, R.; De Charro, F. The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective: Evidence from the EuroQol BIO MED Research Programme. Springer; 2003.

38. Andresen EM, Rothenberg BM, Kaplan RM. Performance of a self-administered mailed version of the quality of well-being (QWB-SA) questionnaire among older adults. Med Care. 1998; 36(9): 1349–60. [PubMed: 9749658]

39. Fryback DG, Palta M, Cherepanov D, Bolt D, Kim JS. Comparison of 5 Health-Related Quality-of-Life Indexes Using Item Response Theory Analysis. Medical Decision Making. 2010; 30(1):5–15. [PubMed: 19843961]

40. Feeny D, Spritzer K, Hays RD, Liu HH, Ganiats TG, Kaplan RM, et al. Agreement about Identifying Patients Who Change over Time: Cautionary Results in Cataract and Heart Failure Patients. Medical Decision Making. 2012; 32(2):273–86. [PubMed: 22009666]

41. Kaplan RM, Tally S, Hays RD, Feeny D, Ganiats TG, Palta M, et al. Five preference-based indexes in cataract and heart failure patients were not equally responsive to change. J Clin Epidemiol. 2011; 64(5):497–506. [PubMed: 20685077]

42. Bateman I, Munro A, Rhodes B, Starmer C, Sugden R. Does part-whole bias exist? An experimental investigation. Econ J. 1997; 107(441):322–32.

43. Craig BM, Busschbach JJ, Salomon JA. Modeling ranking, time trade-off, and visual analog scale values for EQ-5D health states: a review and comparison of methods. Med Care. 2009; 47(6):634–41. Epub 2009/05/13. [PubMed: 19433996]

44. Craig BM, Reeve BB, Cella D, Hays RD, Pickard AS, Revicki DA. Demographic Differences in Health Preferences in the United States. Med Care. 2013 Epub 2014/01/01.

45. Sullivan PW, Ghushchyan V. Preference-based EQ-5D index scores for chronic conditions in the United States. Medical Decision Making. 2006; 26(4):410–20. [PubMed: 16855129]

46. Tengs TO, Wallace A. One thousand health-related quality-of-life estimates. Med Care. 2000; 38(6):583–637. [PubMed: 10843310]

Figure 1.
Example of Paired Comparison

Figure 2.
Proportion who prefer a loss in lifespan over pain or depression for 10 years.*
* Pain Intensity was measured on an 11-point scale from no pain (0) to worst imaginable pain (10). Each point represents a pair-specific sample and sample sizes range from 711 to 772, except the first and last pairs on Pain Intensity 3 (571 and 282, respectively) due to a coding error.
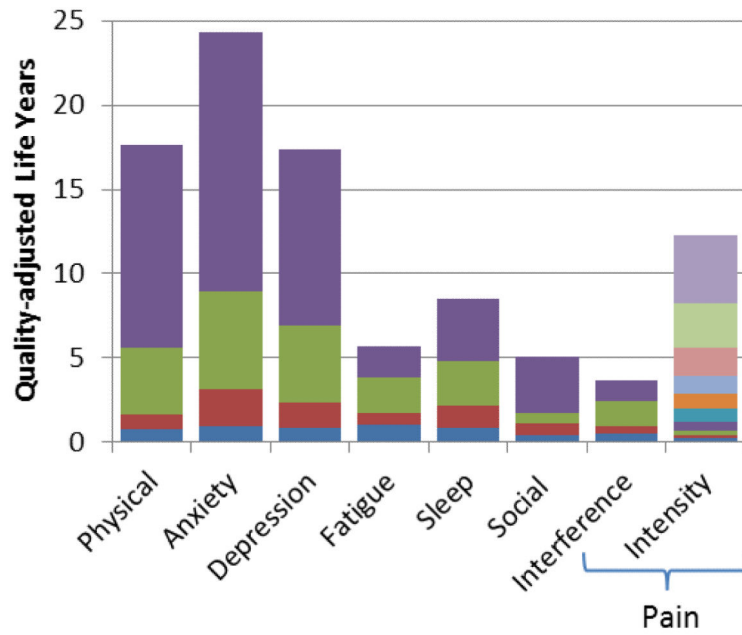
Figure 3.
Losses in QALYs associated with health problems for 10 years described by PROMIS-29*
QALY=quality-adjusted life year
*Cut points in the bars represent the losses in QALYs associated with an increase in severity of a health problem (i.e., Level 1 to 2…Level 10 to 11). The full bar represents the loss in QALYs associated with 10 years with the health problem at its worst level of severity.

Figure 4.
Histogram of 1-year values on a quality-adjusted life year scale by self-reported general health

Table 1

Respondent Characteristics by Completion and Compared to 2010 US Population[*]

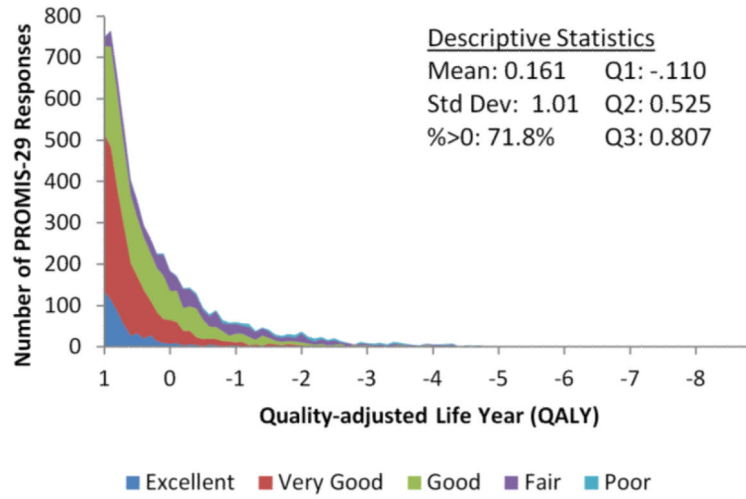|	Dropout N=386 % (#)	Terminated N=456 % (#)	Completed N=7557 % (#)	p-value	US 2010 Census %
Age in years					
18 to 34	30.05 (116)	30.92 (141)	28.12 (2125)	0.006	30.58
35 to 54	39.12 (151)	40.57 (185)	35.87 (2711)		36.70
55 and older	30.83 (119)	28.51 (130)	36.01 (2721)		32.72
Sex					
Male	46.11 (178)	46.71 (213)	48.39 (3657)	0.552	48.53
Female	53.89 (208)	53.29 (243)	51.61 (3900)		51.47
Race					
White	77.33 (290)	78.83 (350)	84.47 (6195)	<0.001	74.66
Black or African American	20.8 (78)	18.47 (82)	12.09 (887)		11.97
American Indian or Alaska Native	0.27 (1)	0.45 (2)	0.72 (53)		0.87
Asian	1.07 (4)	1.58 (7)	2.25 (165)		4.87
Native Hawaiian or other Pacific Islander	0.53 (2)	0.68 (3)	0.46 (34)		0.16
Some other race	-	-	-		5.39
Two or more races	2.93 (11)	2.70 (12)	3.04 (223)		2.06
Hispanic ethnicity					
Hispanic or Latino	13.21 (51)	13.16 (60)	12.86 (972)	0.966	14.22
Not Hispanic or Latino	86.79 (335)	86.84 (396)	87.14 (6585)		85.78
Educational attainment among age 25 or older					
Less than high school	1.94 (7)	4.52 (19)	1.64 (115)	<0.001	14.42
High school graduate	19.11 (69)	22.86 (96)	17.83 (1252)		28.50
Some college, no degree	23.82 (86)	23.57 (99)	25.76 (1809)		21.28
Associate's degree	17.17 (62)	12.38 (52)	13.03 (915)		7.61
Bachelor's degree	34.90 (126)	33.33 (140)	37.83 (2657)		17.74
Graduate or professional degree	2.77 (10)	3.10 (13)	3.86 (271)		10.44
Refused/Don't know	0.28 (1)	0.24 (1)	0.06 (4)		-
Household income					
$14,999 or less	11.92 (46)	11.40 (52)	8.55 (646)	0.036	13.46
$15,000 to $24,999	11.14 (43)	13.38 (61)	10.80 (816)		11.49
$25,000 to $34,999	12.95 (50)	12.94 (59)	11.75 (888)		10.76
$35,000 to $49,999	13.73 (53)	14.69 (67)	16.75 (1266)		14.24
$50,000 to $74,999	22.28 (86)	17.54 (80)	19.88 (1502)		18.28
$75,000 to $99,999	11.4 (44)	10.75 (49)	11.98 (905)		11.81
$100,000 to $149,999	6.22 (24)	8.11 (37)	10.27 (776)		11.82
$150,000 or more	3.89 (15)	4.17 (19)	4.39 (332)		8.14
Refused/Don't know	6.48 (25)	7.02 (32)	5.64 (426)		-

[*] Age, sex, race, and ethnicity estimates for the US are based on 2010 Census Summary File 1.

Educational attainment and household income are based on 2010 American Community Survey 1-Year Estimates. Unlike the US Census, the American Community Survey excluded adults not in the community (e.g., institutionalized) and describes income by the proportion of households, not adults.

Table 2

Valuation of the PROMIS-29: 7 Domains with Four 5-level Items from Best (1) to Worst (5)

Loss in QALYs associated with health problems for 10 years		Level 1 to 2		Level 2 to 3		Level 3 to 4		Level 4 to 5	
		d_h	95% CI	d_h	95% CI	d_h	95% CI	d_h	95% CI
Physical Functioning	Chores	0.18	0.16 0.20	0.20	0.18 0.22	0.86	0.77 0.97	2.57	2.30 2.98
	Stairs	0.15	0.13 0.17	0.17	0.15 0.20	0.56	0.49 0.64	1.20	1.06 1.38
	Walk	0.25	0.22 0.27	0.23	0.20 0.25	0.93	0.85 1.05	2.59	2.31 2.96
	Errands	0.24	0.21 0.27	0.22	0.19 0.24	1.63	1.50 1.82	5.68	5.08 6.65
Anxiety	Fearful	0.25	0.23 0.28	0.52	0.47 0.59	1.78	1.61 2.01	5.38	4.67 6.45
	Focus	0.31	0.27 0.34	0.57	0.51 0.63	1.68	1.54 1.91	3.81	3.28 4.65
	Worries	0.27	0.24 0.30	0.74	0.67 0.83	1.61	1.47 1.80	4.13	3.63 4.89
	Uneasy	0.15	0.13 0.17	0.34	0.30 0.38	0.72	0.64 0.82	2.05	1.80 2.43
Depression	Worthless	0.22	0.21 0.25	0.39	0.35 0.43	1.07	0.98 1.19	2.69	2.42 3.09
	Helpless	0.18	0.16 0.20	0.29	0.26 0.32	0.79	0.72 0.89	1.62	1.42 1.89
	Depressed	0.25	0.22 0.27	0.49	0.44 0.54	1.52	1.40 1.68	3.50	3.19 3.96
	Hopeless	0.22	0.20 0.24	0.33	0.29 0.37	1.18	1.08 1.33	2.65	2.36 3.05
Fatigue	Fatigue	0.24	0.21 0.26	0.14	0.13 0.16	0.66	0.59 0.74	0.51	0.45 0.57
	Starting	0.32	0.28 0.35	0.15	0.13 0.16	0.48	0.43 0.54	0.38	0.34 0.43
	Run-Down	0.31	0.27 0.35	0.17	0.15 0.19	0.52	0.47 0.58	0.48	0.43 0.55
	Average Fatigue	0.21	0.18 0.24	0.17	0.15 0.19	0.51	0.44 0.59	0.48	0.40 0.57
Sleep	Quality	0.17	0.15 0.19	0.56	0.51 0.61	1.39	1.26 1.58	1.21	1.09 1.37
	Refreshing	0.19	0.17 0.21	0.37	0.34 0.41	0.26	0.23 0.29	1.65	1.51 1.84
	Problem	0.34	0.31 0.38	0.21	0.19 0.24	0.66	0.59 0.75	0.42	0.36 0.48
	Difficulty	0.19	0.17 0.21	0.17	0.15 0.18	0.35	0.31 0.39	0.34	0.31 0.39
Social Functioning	Amount	0.09	0.08 0.11	0.16	0.14 0.18	0.15	0.13 0.16	0.57	0.51 0.64
	Work	0.11	0.09 0.13	0.17	0.15 0.19	0.16	0.14 0.18	0.74	0.67 0.82
	Personal	0.12	0.10 0.13	0.19	0.17 0.21	0.15	0.14 0.17	0.95	0.86 1.07
	Routine	0.12	0.10 0.13	0.20	0.18 0.22	0.16	0.15 0.18	1.00	0.91 1.12
Pain Interference	Day-to-day	0.10	0.08 0.12	0.13	0.11 0.14	0.44	0.38 0.50	0.34	0.30 0.38
	Home	0.06	0.05 0.08	0.08	0.07 0.09	0.20	0.17 0.23	0.19	0.17 0.22
	Social Activities	0.11	0.09 0.13	0.07	0.06 0.08	0.20	0.18 0.23	0.19	0.17 0.22

Author Manuscript

Loss in QALYs associated with health problems for 10 years	Level 1 to 2		Level 2 to 3		Level 3 to 4		Level 4 to 5	
	d_h	95% CI	d_h	95% CI	d_h	95% CI	d_h	95% CI
Enjoyment	0.22	0.19 0.25	0.16	0.14 0.18	0.68	0.61 0.76	0.48	0.43 0.55

QALY=quality-adjusted life year; d_h=decrement; CI=confidence interval.

Table 3

Valuation of the PROMIS-29: Pain Intensity from No Pain (0) to Worst Pain Imaginable (10)

Loss in QALYs associated with pain intensity for 10 years	d_h*	95% CI	
Level 0 to 1	0.23	0.21	0.25
Level 1 to 2	0.21	0.19	0.23
Level 2 to 3	0.28	0.25	0.31
Level 3 to 4	0.53	0.41	0.67
Level 4 to 5	0.80	0.72	0.89
Level 5 to 6	0.80	0.70	0.90
Level 6 to 7	1.07	0.95	1.21
Level 7 to 8	1.69	1.52	1.89
Level 8 to 9	2.61	2.37	2.91
Level 9 to 10	4.10	3.56	4.81

QALY=quality-adjusted life year; d_h=decrement CI=confidence interval.

*
Same results as last column in Figure 3.