

UC Merced

UC Merced Previously Published Works

Title

US Valuation of Health Outcomes Measured Using the PROMIS-29

Permalink

<https://escholarship.org/uc/item/8zt2p7vv>

Journal

Value in Health, 17(8)

ISSN

1098-3015

Authors

Craig, Benjamin M
Reeve, Bryce B
Brown, Paul M
et al.

Publication Date

2014-12-01

DOI

10.1016/j.jval.2014.09.005

Peer reviewed

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

US Valuation of Health Outcomes Measured Using the PROMIS-29

Benjamin M. Craig, PhD^{1,*}, Bryce B. Reeve, PhD², Paul M. Brown, PhD³, David Cella, PhD⁴, Ron D. Hays, PhD^{5,6}, Joseph Lipscomb, PhD⁷, A. Simon Pickard, PhD⁸, Dennis A. Revicki, PhD⁹

¹Health Outcomes and Behavior, Moffitt Cancer Center and University of South Florida, Tampa, FL, USA; ²Department of Health Policy and Management, Gillings School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; ³School of Social Sciences, Humanities and Arts, University of California, Merced, Merced, CA, USA; ⁴Department of Medical Social Sciences, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA; ⁵Division of General Internal Medicine and Health Services Research, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA; ⁶RAND, Health Program, Santa Monica, CA, USA; ⁷Department of Health and Policy Management, Rollins School of Public Health, Emory University, Atlanta, GA, USA; ⁸Department of Pharmacy Systems, Outcomes and Policy, University of Illinois at Chicago, Chicago, IL, USA; ⁹Outcomes Research, Evidera, Bethesda, MD, USA

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 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 article introduces a novel approach to valuing HRQOL for economic evaluations using paired comparisons and provides a tool to translate PROMIS-29 responses into quality-adjusted life-years.

Keywords: discrete choice experiments, patient-reported outcomes, quality-adjusted life-years.

Copyright © 2014, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

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]. This study, however, 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.

Because of its chronologic reference, outcomes evidence is more informative than health status evidence, yet outcomes evidence alone may not be 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

Conflicts of interest: The authors have indicated they have no conflicts of interest with regard to the content of this article.

* Address correspondence to: Benjamin M. Craig, Moffitt Cancer Center, 12902 Magnolia Drive, MRC-CANCONT, Tampa, FL 33612.

E-mail: benjamin.craig@moffitt.org.

1098-3015/\$36.00 – see front matter Copyright © 2014, International Society for Pharmacoeconomics and Outcomes Research (ISPOR).

Published by Elsevier Inc.

<http://dx.doi.org/10.1016/j.jval.2014.09.005>

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, such as 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 QALY). 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 outcome (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 trade-off between losses in HRQOL and lifespan. For example, the paired comparison shown in Figure 1 involves a trade-off between 10 years sometimes depressed and a loss of 1 QALY. Using responses to 541 pairs like this one, this study directly estimates the value of PROMIS outcomes on a QALY scale.

Using a data set with both PROMIS scores and EuroQol five-dimensional questionnaire (EQ-5D) responses, Revicki et al. [7] derived regression equations that mapped PROMIS scores to QALYs. 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 one item per domain (e.g., the EQ-5D) or that reduce evidence from multiple items to one attribute per domain (e.g., the six-dimensional health state short form [derived from short-form 36 health survey]). The use of one 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 29-item PROMIS (PROMIS-29), which includes four items on seven 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 PROMIS-29, this is the first national study that uses online DCE for health valuation. Previously, Bansback et al. [9] 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 et al. [10] recruited members of the Toluna United States panel to value 12-Item Short Form Health Survey (SF-12) Version 1 and the six-dimensional health state short form (derived from short-form 36 health survey) responses, but this national sample was heavily skewed toward older white women. Viney et al. [11] 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. The objective of this project was to estimate values for 10-year losses in HRQOL on a QALY scale described by PROMIS-29 on the basis of the perspective of adult members of the US population.

Given the extent of study details and the need to limit article 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 in Supplementary Materials found at <http://dx.doi.org/10.1016/j.jval.2014.09.005>, we provide Stata code, log, and data to allow reproducibility of the results within this article.

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 was introduced in 2008 to describe the relationship between health episodes and individual choices, particularly ranking tasks [12]. Episodic random utility model specifies that the utility of a health episode is a function of 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 because of 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, 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). Instant random utility model describes the relationship between health states and choices and becomes unstable when the duration becomes small (e.g., $t \rightarrow 0$) [12,14].

Although the concepts are sometimes used interchangeably [15], this article differentiates between value and utility. Value refers to a preference-based measure representing the choices of

| Which health scenario do you prefer? | |
|---|---|
| No health problems (except depression) for 10 years then die | No health problems for 9 years then die (1 year less than 10 years) |
| Sometimes feel worthless, helpless, depressed, and hopeless | Not feel worthless, helpless, depressed, and hopeless |

Fig. 1 – Example of paired comparison.

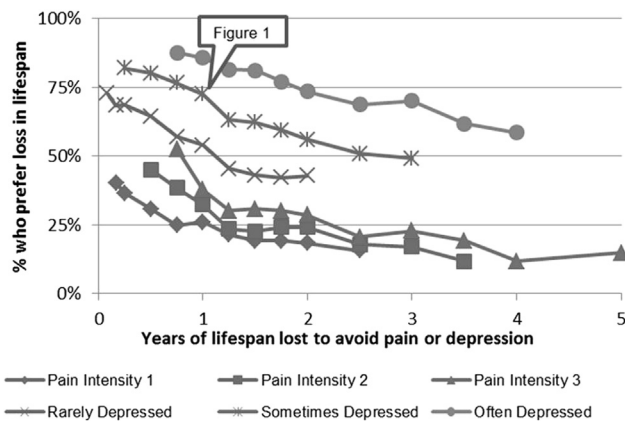


Fig. 2 – Proportion who prefer a loss in lifespan over pain or depression for 10 years. Pain Intensity was measured on an 11-point scale from no pain (0) to worst imaginable pain (10). Each point represents a pair-specific sample and sample sizes range from 711 to 772, except the first and last pairs on Pain Intensity 3 (571 and 282, respectively) due to a coding error.

a group of individuals, V , and utility is a random latent trait at the individual level that governs a person's choice (i.e., episodic random utility model). 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; that is, $V(\text{sometimes depressed}, 10 \text{ years}) = V(\text{no health problems}, 7.4 \text{ years})$. 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 QALY). 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; that is, $1/(2.6 + 1)$. Looking at the empirical data (Fig. 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 ; that is, $1/(d_A + 1) = 28\%$ 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 nonadditive, depending instead on a CDF—for example, $d_B/(d_A + d_B)$. 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 PROMIS-29 on a QALY scale.

Health Outcomes

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 short-form 36 health survey [18]. 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 seven 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, that is, $(7 \times 16) + 10$.

Survey Panels

This project recruited US respondents from multiple panel vendors, with each panel recruiting 1000 respondents with completed surveys [21]. We chose to use multiple vendors to assess and compare costs, services, responsiveness, and quality of data. We separated survey hosting from recruitment activities to use 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 on the basis of survey responses or alter or autogenerate 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 no. Pro00000076) and are described in greater detail in a report posted online and in the Appendix in Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2014.09.005> [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's 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 to compare these demographic data with survey responses.

Survey Design

Pretesting 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 three paired comparisons, the DCE component consisted of 30 paired comparisons distributed over four sections [10,30].

The primary difference between the four DCE sections was their pivots. A pivot is the set of attributes in common for both alternatives in a pair (aka holdouts) [10,31]. Within a DCE section, each pair had the same pivot, which was modified by adding two compensating attributes [30]. For the six lifespan pairs, the pivot was 10 years with no health problems followed by death (see Appendix Fig 1 in Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2014.09.005>). For the eight health pairs in the next three sections, the pivot was 10 years in good, fair, and poor health followed by death, respectively (see Appendix in Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2014.09.005>). 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 years). A loading animation required that at least 8 seconds be spent on each comparison to ensure 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 on the basis of his or her demographic characteristics (reported in survey and verified by vendor) to guarantee that each pair-specific sample corresponded to demographic quotas [8].

The six lifespan pairs directed respondents to choose between episodes with either reduced lifespan or one of six “health problems” for 10 years, including three levels of depression (rarely, sometimes, or often feeling worthless, helpless, depressed, and hopeless) and three 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 with 10 losses in lifespan, creating 60 pairs with 100 responses for each pair from each panel (1000 respondents \times 6 responses/60 pairs = 100 responses per pair), except for the third pain intensity level, which was compared with 11 losses in lifespan because of a coding error (Figure 2 and Appendix in Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2014.09.005>).

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 in Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2014.09.005>). Each item pair directs respondents to choose between a decrement in one 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 one 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 \times 24 responses/480 pairs = 50 responses per pair; see Appendix in Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2014.09.005> for more details on pair selection).

Econometrics

Each of the 226,710 DCE responses ($N = 7557$ respondents \times 30 responses) was incorporated into the calculation of the 541 pair-specific probabilities, $p_1 \dots p_{541}$ (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 (more than 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 multiattribute utility 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 $P(\cdot)$ is a CDF. Two specifications of $P(\cdot)$ 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, $P = d_B/(d_A + d_B)$, that has the advantage that θ factors out. These two specifications are compared on the basis of their ability to predict 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 with 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 those in 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 weighted and unweighted probabilities were small (<0.004); therefore, only unweighted results are shown. Compared with the relativity specification, the logit produced greater squared error (6519 vs. 2403) and more negative decrements (36 vs. 0);

Table 1 – Respondent characteristics by completion and compared with 2010 US population.*

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

Note. Values are n (%) unless indicated otherwise.

* 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.

therefore, all results shown are based on the relativity specification.

Tables 2 and 3 provide multiattribute utility 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 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 ten-year decrements (94.58 QALYs). Under the same constant proportionality and no discounting assumptions, 1 year in pits represents a reduction of 9.458 QALY (i.e., – 8.458 QALYs; 1 – 94.58/10) from full health. Therefore, the range of 1-year values based on 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 fourth pain interference item, not “your enjoyment of life.” For illustrative purposes, the

Table 2 – Valuation of the PROMIS-29: Seven 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 |
| 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 |

CI, confidence interval; d_h , decrement; PROMIS-29, 29-item Patient-Reported Outcomes Measurement Information System; QALY, quality-adjusted life-year.

responses are assumed to be the same (excluding or including this fourth item had no noticeable effect on Fig. 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 on the basis of 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 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.

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 y | 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 |

CI, confidence interval; d_h , decrement; PROMIS-29, 29-item Patient-Reported Outcomes Measurement Information System; QALY, quality-adjusted life-year.

* Same results as last column in Figure 3.

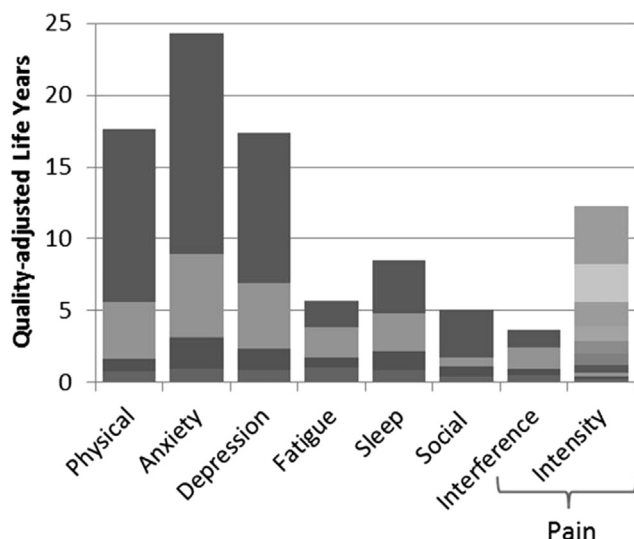


Fig. 3 – Losses in QALYs associated with health problems for 10 years described by PROMIS-29. Cut points in the bars represent the losses in QALYs associated with an increase in severity of a health problem (i.e., Level 1 to 2...Level 10 to 11). The full bar represents the loss in QALYs associated with 10 years with the health problem at its worst level of severity. QALY, quality-adjusted life year.

Although 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 to 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 [36], the EQ-5D [37], and the Quality of Well-Being Self-Administered questionnaire [38]. 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 MAU 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 MAU, estimation with both score shifts and decrements may test whether the whole is greater than the parts.

The relationship between QALYs derived here for PROMIS-29 and those of existing preference-based measures is unknown. More work is needed to demonstrate advantages of potential improvements in measurement reliability and greater number of domains for CER. For example, the EQ-5D includes mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, while PROMIS-29 includes a broader assessment of physical function, social function, sleep disturbance, and fatigue. In contrast, the Health Utilities Index Mark 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 of 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 because research suggests that the respondent's age and the 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 ensure demographic representation of each pair-specific probability. Unobservable characteristics concerning participation in panels, however, 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 with the US Census (Table 1) [34]. Still, it is unclear whether these potential selection issues introduced bias in 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 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 with 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 on the basis of the value gained from improved health outcomes as assessed by PROMIS measures. This will greatly enhance our understanding of the relative merit of treatments.

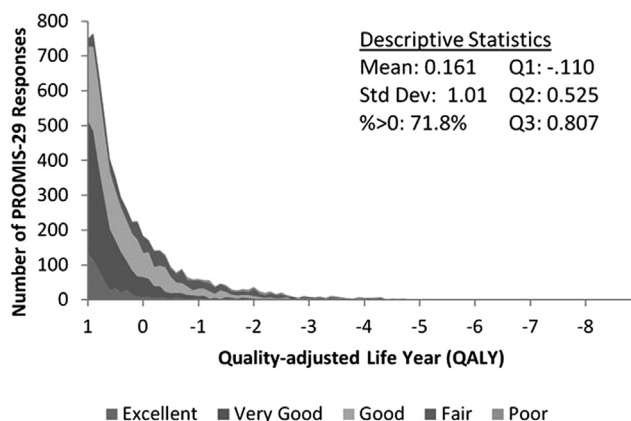


Fig. 4 – Histogram of 1-year values on a quality-adjusted life year scale by self-reported general health.

Acknowledgments

We thank Michelle Owens, Carol Templeton, and Shannon Runge at Moffitt Cancer Center for their contributions to the research

and creation of this article. We also greatly appreciate the external review comments from Dennis Fryback and David Feeny on the study methodology.

Source of financial support: Funding support for this research was provided by a National Cancer Institute R01 grant (1R01CA160104). Ron D. Hays was supported in part by grants from the National Institute on Aging (P30-AG021684) and the National Institute on Minority Health and Health Disparities (P20MD000182).

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at <http://dx.doi.org/10.1016/j.jval.2014.09.005> or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

- Agency for Healthcare Research and Quality. What is comparative effectiveness research. Available from: <http://effectivehealthcare.ahrq.gov/index.cfm/what-is-comparative-effectiveness-research/1/>. [Accessed December 3, 2012].
- Kozma CM, Reeder CE, Schulz RM. Economic, clinical, and humanistic outcomes: a planning model for pharmacoeconomic research. *Clin Ther* 1993;15:1121–32.
- Fanshel S, Bush JW. A health-status index and its application to health-services outcomes. *Oper Res* 1970;18:1021–66.
- Lipscomb J, Drummond M, Fryback D, et al. Retaining, and enhancing, the QALY. *Value Health* 2009;12(Suppl):S18–26.
- Neumann PJ, Greenberg D. Is the United States ready for QALYs? *Health Affairs (Project Hope)* 2009;28:1366–71.
- Cella D, Riley W, Stone A, 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;63:1179–94.
- Revicki DA, Kawata AK, Harnam N, et al. 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* 2009;18:783–91.
- Craig B, Reeve BB. Methods report on the PROMIS valuation study: year 1. 2012. Available from: <http://labpages.moffitt.org/craigb/Publications/Report120928.pdf>. [Accessed October 29, 2012].
- Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. *J Health Econ* 2012;31:306–18.
- Craig BM, Pickard AS, Stolk E, Brazier JE. US valuation of the SF-6D. *Med Decis Making* 2013;33:793–803.
- Viney R, Norman R, Brazier J, et al. An Australian discrete choice experiment to value EQ-5D health states. *Health Econ* 2014;23:729–42.
- 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.
- 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;45:233–43.
- Craig BM, Busschbach JJ. Revisiting United States valuation of EQ-5D states. *J Health Econ* 2011;30:1057–63.
- Keeney RL, Raiffa H. *Decisions with Multiple Objectives: Preferences and Value Trade-offs*. New York, NY: Cambridge University Press, 1993.
- Adams K, Bayliss E, Blumenthal D, et al. Universal health outcome measures for older persons with multiple chronic conditions. *J Am Geriatr Soc* 2012;60:2333–41.
- Forrest CB, Bevans KB, Tucker C, et al. Commentary: The patient-reported outcome measurement information system (PROMIS) for children and youth: application to pediatric psychology. *J Pediatr Psychol* 2012;37:614–21.
- Hinchcliff M, Beaumont JL, Thavarajah K, 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;63:1620–8.
- Selewski DT, Collier DN, MacHardy J, et al. Promising insights into the health related quality of life for children with severe obesity. *Health Qual Life Outcomes* 2013;11:29.
- PROMIS-29 Profile v1.0. 2008–2012 PROMIS Health Organization and PROMIS Cooperative Group. Available from: <https://www.assessment-center.net/ac1/files/pdf/44b7636201a34267a9213db7f69f2c6d.pdf>. [Accessed June 6, 2012].
- Craig B, Hays R, Pickard AS, et al. Comparison of US panel vendors for online surveys. *J Med Int Res* 2013;15:e260.
- 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. [Accessed October 5, 2013].
- Ryan M, Gerard K, Amaya-Amaya M, eds. *Using Discrete Choice Experiments to Value Health and Health Care* [electronic resource]. Dordrecht: Springer, 2008.
- Kahneman D. A perspective on judgment and choice—mapping bounded rationality. *Am Psychol* 2003;58:697–720.
- Bridges JFP, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health* 2011;14:403–13.
- PROMIS Assessment Center. PROMIS terms and conditions. 2012. Available from: https://www.assessmentcenter.net/documents/PROMIS%20Terms%20and%20Conditions%20v8%20July10_2012.pdf. [Accessed August 08, 2012].
- National Institutes of Health. PROMIS: dynamic tools to measure health outcomes from the patient perspective. Available from: <http://www.nihpromis.org/#1>. [Accessed June 27, 2012].
- Alexandrov A. Characteristics of single-item measures in Likert scale format. *Electr J Bus Res Meth* 2010;8:1–12.
- Swain SD, Weathers D, Niedrich RW. Assessing three sources of misresponse to reversed Likert items. *J Mark Res* 2008;45:116–31.
- Chrzan K. Using partial profile choice experiments to handle large numbers of attributes. *Int J Market Res* 2010;52:827–40.
- Louviere J, Lancsar E. Choice experiments in health: the good, the bad, the ugly and toward a brighter future. *Health Econ Policy Law* 2009;4:527–46.
- Urban FM. *The Application of Statistical Methods to the Problems of Psychophysics*. Philadelphia: Psychological Clinic Press, 1908.
- Urban FM. Urban's solution (minimum normit X^2). In: Bock RD, Jones LV, eds. *The Measurement and Prediction of Judgment and Choice*. San Francisco: Holden-Day, 1968.
- Liu HH, Cella D, Gershon R, et al. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. *J Clin Epidemiol* 2010;63:1169–78.
- 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. *Eur J Health Econ* 2013;14(Suppl. 1):S15–24.
- 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;33:375–84.
- 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*. Amsterdam, The Netherlands: Kluwer Academic Publishers, 2003.
- 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;36:1349–60.
- Fryback DG, Palta M, Cherepanov D, et al. Comparison of 5 health-related quality-of-life indexes using item response theory analysis. *Med Decis Making* 2010;30:5–15.
- Feeny D, Spritzer K, Hays RD, et al. Agreement about identifying patients who change over time: cautionary results in cataract and heart failure patients. *Med Decis Making* 2012;32:273–86.
- Kaplan RM, Tally S, Hays RD, et al. Five preference-based indexes in cataract and heart failure patients were not equally responsive to change. *J Clin Epidemiol* 2011;64:497–506.
- Bateman I, Munro A, Rhodes B, et al. Does part-whole bias exist? An experimental investigation. *Econ J* 1997;107:322–32.
- 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;47:634–41.
- Craig BM, Reeve BB, Cella D, et al. Demographic differences in health preferences in the United States. *Med Care* 2013;52:307–13.
- Sullivan PW, Ghushchyan V. Preference-based EQ-5D index scores for chronic conditions in the United States. *Med Decis Making* 2006;26:410–20.
- Tengs TO, Wallace A. One thousand health-related quality-of-life estimates. *Med Care* 2000;38:583–637.