Published in final edited form as:

Qual Life Res. 2015 July ; 24(7): 1759–1765. doi:10.1007/s11136-014-0897-z.

Do health preferences contradict ordering of EQ-5D labels?

Benjamin M. Craig,
Health Outcomes and Behavior, Moffitt Cancer Center, University of South Florida, 12902 Magnolia Drive, MRC-CANCONT, Tampa, FL 33612-9416, USA

A. Simon Pickard, and
Department of Pharmacy Systems, Outcomes and Policy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, USA

Kim Rand-Hendriksen
Department of Health Management and Health Economics, University of Oslo, Oslo, Norway

Benjamin M. Craig: benjamin.craig@moffitt.org

Abstract

Purpose—The aim of this study was to test whether the ordering of item labels in EQ-5D instruments disagrees with the preferences of US adults.

Methods—A preference inversion occurs when “worse” health along a scale or score is preferred. As a sub-study of the 2013 United States Measurement and Valuation of Health Study, we tested for 33 EQ-5D preference inversions using paired comparisons with unique samples of 50 or more US adults, aged 18 or older. Specifically, we tested whether health preferences contradicted ordering of EQ-5D labels.

Results—The EQ-5D-3L and EQ-5D-Y item labels had no significant preference inversions. The EQ-5D-5L version had preference inversions between Levels 4 and 5. For example, 30 out of 59 respondents (51 %) preferred being “extremely” over “severely anxious or depressed,” contrary to the ordering of labels for that item.

Conclusions—Preference inversions between Levels 4 and 5 on the EQ-5D-5L were tested and confirmed; therefore, valuation studies may find that Levels 4 and 5 have the same value. To mitigate such inversions, labels could be revised or a 4-level version could be considered.

Keywords

EQ-5D; QALY; Health preferences; Psychometrics; Item response; Scales

Introduction

The EQ-5D is perhaps the most widely used measure of health-related quality of life (HRQoL). This 5-item instrument now exists in three versions: (1) the initial version—now called EQ-5D-3L—consists of five health dimensions described at three levels

Correspondence to: Benjamin M. Craig, benjamin.craig@moffitt.org.

Conflict of interest There are no conflicts of interest.

corresponding roughly to no, some, and extreme problems; (2) a more recent version—EQ-5D-5L—with five levels; and (3) the EQ-5D-Y (youth version), which is a three-level version intended for adolescents aged 7–12 years. An important challenge in designing any HRQoL measure is the selection of labels that describes the extent of problems along each dimension.

A preference inversion is when a respondent states a preference that contradicts the ordering of labels in an item. A high prevalence of preference inversions within a valuation study can lead to indifference between levels (i.e., coefficients equal zero) or even an increase in value (i.e., positive coefficient sign). For instance, such a high prevalence conceivably occurred along the dimension of cognition for the Health Utilities Index Mark 2/3 between Level 2 and 3 based on the coefficient estimates [1]. Evidence of indifference may have affected the UK valuation of the SF-6D, where lack of statistical significance and/or the sign of coefficient estimates for certain models were important criteria in eventual model selection [2]. Subsequently, the US SF-6D valuation study bounded the coefficients to be nonzero due concerns of preference inversions [3].

Based on the many EQ-5D-3L valuation studies that have been conducted [4, 5], there is substantial evidence to indicate that EQ-5D-3L item labels are ordered and few (if any) preference inversions occur. On the other hand, EQ-5D-5L valuation studies to date indicate that general population respondents appear to assign similar values to Levels 4 and 5 of several items; and it has been suggested that the labels used to describe Levels 4 and 5—“severe” and “extreme” problems, respectively—may be too similar in magnitude to work well. For example, Xie et al. [6] found that the Level 4 on anxiety/depression has nearly the same value as Level 5 (0.244 and 0.254, respectively). In fact, the study found that Level 4 was higher than Level 5 for Usual Activities (0.129 and 0115, respectively; i.e., value inversion). The lack of preference evidence at the time of EQ-5D-5L item labeling may also have contributed to these suspicions of EQ-5D-5L preference inversions, which motivated this study.

Successful integration of psychometric and econometric methods is a fundamental challenge within outcomes research [7–9]. To be fit for purpose, HRQoL measures should be precise and decision relevant. In psychometrics, item responses from a validated instrument describe a health domain (e.g., depression) commonly under the assumption that “better” health outcomes are preferred (even with little or no supporting evidence). In econometrics, HRQoL instruments are evaluated based on whether they measure decision relevant outcomes, not based on their measurement properties. These fields must integrate their methods for each to succeed. After items and labels are proposed, preference inversion studies can be conducted using a variety of valuation approaches such as discrete choice experiments (DCE). These experiments (Fig. 1) basically show respondents' two alternative outcomes and ask, “Which do you prefer?” For example, is “severely anxious or depressed” preferred over “extremely anxious or depressed”?

In this study, we exemplify this process by examining whether the item labels in EQ-5D instruments disagreed with the preferences and whether US adults are indifferent to “worse” health along each item. This was accomplished by conducting a sub-study as part of the

2013 United State Measurement and Valuation of Health Study (2013 US MVH), in which we examined 33 EQ-5D preference inversions using unique samples of 50 or more US adults (i.e., each respondent answered only one preference inversion paired comparison).

Methods

EQ-5D

The EQ-5D is a generic measure of HRQoL that has been widely used in clinical and economic evaluations of health care as well as to capture the health of populations [10]. The original version included three levels (EQ-5D-3L); however, two more recent versions were released: a youth-friendly 3-level version (EQ-5D-Y) and a 5-level version (EQ-5D-5L) [11–14]. In each version, the measure included five items: Mobility, Self-Care, Usual Activities, Pain/Discomfort, and Anxiety/Depression. Each version was designed for self-completion, has a low respondent burden, and has been administered using a variety of modalities (e.g., postal surveys, online).

A preference inversion is when a respondent states a preference that contradicts the ordering of labels in an item (i.e., prefers a loss or decrement in health). For the EQ-5D items, health worsens with each step up the scale (e.g., 1–2, 2–3). The EQ-5D-3L and EQ-5D-Y versions each have ten 1-step and five 2-step decrements (e.g., 1–2 and 1–3, respectively). The EQ-5D-5L has 20 1-step, 15 2-step, ten 3-step, and five 4-step decrements. Excluding those decrements that include “no problems” or that are more than 2 steps, the 3 versions have a total of 35 decrements. Two decrements (moderate to extreme on Anxiety/Depression and Pain/Discomfort) are the same for the EQ-5D-3L and EQ-5D-5L; therefore, the 3 versions have 33 decrements (see Tables 1, 2). The purpose of this study is to test whether health preferences contradict ordering of EQ-5D labels for each of the 33 decrements.

2013 US MVH study

Between November 22, 2013, and December 22, 2013, 5,672 US adults, aged 18 and older, were recruited from a nationally representative panel to participate in a 25-min online survey. This study protocol was approved by the University of South Florida Institutional Review Board (USF #8236), and funding was subsidized through Dr. Craig's support account, a EuroQol Group grant, and the PROMIS valuation study (NCI; 1R01CA160104-01) [15]. The survey instrument had four components: Screener, Health, Paired Comparison, and Follow-up. Each respondent completed all items of the three versions of the EQ-5D in random order as part of the Health component prior to completing the paired comparisons. After viewing 3 examples, the respondent completed 30 pairs. Each pair was randomized in sequence, and attributes were randomized horizontally (i.e., left/right). Each pair had 50 respondents or more following 18 demographic quotas (all combinations of 2 genders, 3 age groups, 3 race/ethnicity groups) to promote concordance with the 2010 US Census. Further details about 2013 US MVH are provided online [16].

Sampling for preference inversion sub-study

As a sub-study of the 2013 US MVH, 2113 respondents were randomly selected and assigned 1 of the 33 preference inversion pairs; of those respondents, 1904 (90 %)

completed the survey (Tables 1, 2). Figure 1 provides an example of a preference inversion paired comparison. In addition, 347 of the 1904 respondents completed 1 of 6 version-comparison pairs (Table 3). No adjustments to the significance level were required because each respondent completed only 1 preference inversion pair; therefore, the samples are completely independent. These six pairs assessed the relationship between labels at the worst level across versions to investigate potential changes in the EQ-5D-5L that may mitigate preference inversion.

Econometrics

Tables 1, 2, and 3 show the pair-specific sample sizes, N, and the proportion (P) who prefer the "better" health ($P = N_A/(N_A + N_B)$) for each of the 39 pairs. The 1-sided P values [i.e., $\Phi((P–C)/\text{sqrt}(P(1 – P)/N))$] indicate whether each proportion is greater than median among the 39 proportions (C = 54/55 = 98.18 %).

Results

Table 1 shows the proportions of respondents whose preferences agreed with ordering of EQ-5D-Y and EQ-5D-3L labels. The proportions ranged from 98 to 100 %. In each pair-specific sample of 53–64 respondents, 1 respondent, at most, preferred the worst statement over the better statement. According to these EQ-5D-3L and EQ-5D-Y results, we found no evidence of preference inversion.

Table 2 shows the proportions for the items of the EQ-5D-5L version. Like Table 1, we found no evidence of preference inversions, except for EQ-5D-5L pairs between Levels 4 and 5. For these pairs, we found significant evidence of preference inversion (i.e., $P > 0.98$). For the pair comparing "severely" and "extremely anxious or depressed," we were unable to reject indifference ($P = 0.5$; Fig. 1).

Table 3 examines the relationship between the label at the worst level in Mobility, Pain/Discomfort, and Anxiety/Depression across the three versions. In all pairs, we rejected indifference; however, we found preference inversion between three of the six pairs at a significance level of 0.05. Specifically, some respondents prefer being "severely" or "extremely anxious or depressed" over being "very worried, sad, or unhappy." In addition, some respondents preferred being "confined to bed" over being "unable to walk about." The results show that the EQ-5D-3L version measured worse mobility than the EQ-5D-5L and that the EQ-5D-5L version measured worse Anxiety/Depression and Pain/Discomfort than the Y.

Discussion

Overall, the EQ-5D-3L and EQ-5D-Y versions of the EQ-5D displayed no significant preference inversions. However, preference inversions were observed on the EQ-5D-5L version between Levels 4 and 5, and we failed to reject indifference between "severely anxious or depressed" vs. "extremely anxious or depressed" (Fig. 1).

Item levels may disagree with changes along a latent domain due to issues with the labels attached to the scale [17]. Instruments commonly include labels as written descriptions of ordinal responses (i.e., adjectival scales). These labels may contradict order within the domain from the perspective of respondents; therefore, the responses fail to capture changes along the latent domain. One approach to test for the appropriate ordering of labels is to utilize item response theory and to examine the calibration of step-order labels, which was conducted and supported in a prototype of the EQ-5D-5L version [9]. For example, it may be unclear to some respondents whether “extreme” represents more depression than “severe,” because the former relates more to an individual extrema and the latter seems more clinically relevant.

Likewise, changes along a latent domain may disagree with preferences. As a working definition for preference inversion, “worse” health along a latent domain may be preferred. There may be a wide variety of reasons for such an inversion to occur, ranging from inter-individual differences in the social interpretation of labels (e.g., stigma) to systems-based policy reasons for individuals that motivate preferences for a certain label over another due to external incentives (e.g., to qualify for parking spaces). For example, “unable to walk about” may be preferred to “severe problems walking about,” because of perceived benefits of motorized assistive devices and the possibility that a certain level of disability triggers health insurance benefits. Scales that agree with the latent domain and disagree with preferences may be clinically or prognostically informative (e.g., suicidal tendencies); however, such instruments fail to aggregate the burden of losses in HRQoL from the stakeholders' perspective for use in outcomes research.

The case of severity and extremum

The relatively high proportion of severe–extreme preference inversions indicates that the two terms are perceived as describing quite similar magnitudes along the latent domain or that the two terms are qualitatively different. The term “extreme” is a relatively value-neutral description of magnitude, and in that, it can be used to describe both good and bad phenomena. In contrast, “severe” indicates both magnitude and value—it tells us that we are dealing with something bad, something difficult, something that implies negative, and potentially grave consequences. This difference between “severe” and “extreme” could be more noticeable or salient in conjunction with specific health dimensions, such as anxiety/depression.

In complement to the paired comparisons, the survey included an open-text box and asked respondents, “Please enter any comments and/or suggestion you have regarding this survey.” In this open-text box, one respondent noted, “In terms of wording choice for intensity when measuring conditions, I felt that extreme was greater with pain and physical issues, but severe was greater for mental health-related issues.” Another stated, “The difference between ‘severe’ and ‘extreme’ as adjectives is difficult to differentiate for some of the questions. I could go either way on an answer.” Although this is anecdotal evidence, a handful of additional respondents made similar comments.

It is important to note that the small, perceived magnitude difference between “severe” and “extreme” is an issue likely to affect health valuation, as opposed to how people rate their own health. When respondents rate their health, their response is informed by the scale (i.e., the location of the label between the ends of the scale). When a respondent chooses between alternative health descriptions, the scale is not shown and respondents have nothing but the written descriptions to go by. Without a scale, their perception is limited to their own individual interpretation of how good and bad “severe” and “extreme” are. Therefore, labels along a scale may measure well, but describe poorly. In the case of severe–extreme preference inversions, there is good reason to believe that respondents rating their health problems as “extreme” are really in worse health than those rating their problems as “severe,” yet the difference in value between the two levels could be zero (i.e., indifference).

Optimally, “better” health along any EQ-5D item is preferred by all respondents. To mitigate preference inversion between Levels 4 and 5 in the EQ-5D-5L version, one option would be to revise the labels. For example, evidence from Table 3 suggests that replacing “unable to walk about” with “confined to bed” would decrease preference inversions in mobility. A more drastic solution to the problem would be to discard one level, rendering an EQ-5D-4L with better discrimination between all levels. A 4L version would have the added advantage of being less cognitively taxing, because it would essentially describe three levels of problems in addition to no problems—a setup easier to conceptualize and remember than the current EQ-5D-5L.

Perhaps, the most pragmatic solution would be to adjust the EQ-5D-5L valuation studies. Such studies may place the same value on Levels 4 and 5 (i.e., constraining the difference between “severe” and “extreme” to be nonnegative), essentially allowing a 4-level tariff. Another approach might be to add graphical or numeric labels. For example, which do you prefer? Level 4: severely anxious or depressed or Level 5: extremely anxious or depressed. Incorporating ordinal labels may prevent preference inversion; however, these labels may further confuse respondents who prefer “extreme” over “severe” outcomes.

The primary limitation of this study is that respondents completed the three versions of the EQ-5D prior to completing the paired comparisons. By showing them the items and labels in advance, the design has a downward bias against the prevalence of preference inversions. Furthermore, this study included only labels that are in the English versions of the EQ-5D instruments and did not explore preferences on alternative labels or translations, which limits the interpretation of the results. Given many EQ-5D translations and potential differences in interpretation, even among English-speaking populations and cultures, further evidence of the generalizability of these results is warranted.

In summary, we found support that the labels on the EQ-5D-3L and EQ-5D-Y versions were consistent with US preferences; but preference inversions between Levels 4 and 5 on the EQ-5D-5L were tested and confirmed, particularly for the anxiety/depression item. EQ-5D-5L valuation studies may find that Levels 4 and 5 have the same value; therefore, their QALYs may not account for differences between “severe” and “extreme” anxiety/depression. Options for mitigating the issue of EQ-5D-5L preference inversions include revising the wording for items and labels where inversion occurs or developing a 4-level version. The most important lesson is that future development of HRQoL measures should integrate psychometric and econometric methods in the assessment of items and labels in a manner that takes into account the perspective of the intended population.

Acknowledgments

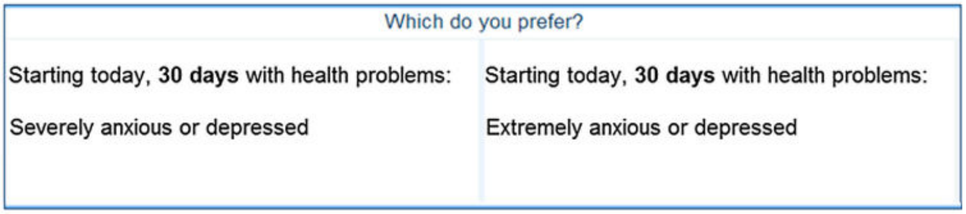
The authors thank Michelle Owens, Shannon Runge, and Carol Templeton at Moffitt Cancer Center for their contributions to the research and creation of this paper. Financial support for this study was provided by a grant from the National Institutes of Health, Department of Health and Human Services, through the National Cancer Institute (1R01CA160104), the EuroQol Group (EQ Project 2013100) and Dr. Craig's support account at Moffitt Cancer Center. The funding agreements ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

References

1. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu ZL, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Medical Care. 2002; 40(2):113–128. [PubMed: 11802084]
2. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics. 2002; 21(2):271–292. [PubMed: 11939242]
3. Craig BM, Pickard AS, Stolk E, Brazier JE. US valuation of the SF-6D. Medical Decision Making. 2013; 33(6):793–803. [PubMed: 23629865]
4. Craig BM, Pickard AS, Lubetkin EI. Health problems are more common, but less severe when measured using newer EQ-5D versions. Journal of clinical epidemiology. 2014; 67(1):93–99. [PubMed: 24075597]
5. Craig BM, Busschbach JJV. Toward a more universal approach in health valuation. Health Economics. 2011; 20(7):864–875. [PubMed: 20677328]
6. Xie, F.; Pullenayegum, E.; Bansback, N.; Bryan, S.; Ohinmaa, A.; Poissant, L.; Johnson, JA. The Canadian EQ-5D-5L valuation study: An exploratory analysis, Table 1.4, 30th Scientific Plenary Meeting of the EuroQol Group. Montreal: EuroQol Group; 2013.
7. Craig B, Reeve B. Patient-reported outcomes and preference research: Igniting the candle at both ends and the middle. ISPOR Connections Uniting Research and Practice. 2012; 18(5):24.
8. Lipscomb J, Drummond M, Fryback D, Gold M, Revicki D. Retaining, and enhancing, the QALY. Value in Health. 2009; 12:S18–S26.
9. Pickard AS, Kohlmann T, Janssen MF, Bonsel G, Rosenbloom S, Cella D. Evaluating equivalency between response systems: Application of the Rasch model to a 3-level and 5-level EQ-5D. Medical Care. 2007; 45(9):812–819. [PubMed: 17712251]
10. The EuroQol Group. EuroQol—A new facility for the measurement of health-related quality of life. Health Policy. 1990; 16(3):199–208. [PubMed: 10109801]
11. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Quality of Life Research. 2011; 20(10):1727–1736. [PubMed: 21479777]
12. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: A multi-country study. Quality of Life Research. 2013; 22(7):1717–1727. [PubMed: 23184421]
13. Ravens-Sieberer U, Wille N, Badia X, Bonsel G, Burstrom K, Cavrini G, et al. Feasibility, reliability, and validity of the EQ-5D-Y: Results from a multinational study. Quality of Life Research. 2010; 19(6):887–897. [PubMed: 20401552]
14. Wille N, Badia X, Bonsel G, Burstrom K, Cavrini G, Devlin N, et al. Development of the EQ-5D-Y: A child-friendly version of the EQ-5D. Quality of Life Research. 2010; 19(6):875–886. [PubMed: 20405245]
15. Craig, BM.; Schell, MJ.; Brown, PM.; Reeve, BB.; Cella, D.; Hays, RD.; Lipscomb, J.; Pickard, AS.; Revicki, DA. HRQoL values for cancer survivors: Enhancing PROMIS measures for CER. Vol. $3,978,708. H. Lee Moffitt Cancer Center: NIH; 2011. p. 89

16. Craig, B.; Reeve, B. Moffitt Cancer Center; 2012. Methods report on the PROMIS valuation study: Year 1. http://labpages.moffitt.org/craigb/Publications/Report120928.pdf

17. Craig BM, Pickard AS, Lubetkin EI. Health problems are more common, but less severe when measured using newer EQ-5D versions. Journal of Clinical Epidemiology. 2014; 67(1):93–99. [PubMed: 24075597]



Which do you prefer?

Starting today, **30 days** with health problems:	Starting today, **30 days** with health problems:
Severely anxious or depressed	Extremely anxious or depressed

Fig. 1. Example of preference inversion paired comparison

Table 1

Proportion of respondents who prefer the better statement: EQ-5D-3L and EQ-5D-Y

Better statement (*A*)	**Worse statement (*B*)**	***P* (*A* > *B*)**	***N***	***P* > 0.98**[a]
EQ-5D-3L, Level 2 to Level 3				
Some problems in walking about	Confined to bed	1.00	56	1.00
Some problems performing usual activities	Unable to perform usual activities	0.98	63	0.56
Some problems washing or dressing self	Unable to wash or dress self	0.98	58	0.52
Moderate pain or discomfort	Extreme pain or discomfort	0.98	55	0.50
Moderately anxious or depressed	Extremely anxious or depressed	0.98	58	0.52
EQ-5D-Y, Level 2 to Level 3				
Some problems walking about	A lot of problems walking about	1.00	64	1.00
Some problems doing usual activities	A lot of problems doing usual activities	1.00	58	1.00
Some problems washing or dressing self	A lot of problems washing or dressing self	1.00	60	1.00
Some pain or discomfort	A lot of pain or discomfort	1.00	52	1.00
A bit worried, sad, or unhappy	Very worried, sad, or unhappy	0.98	54	0.49

For example, all 56 respondents prefer “some problems in walking about” over “confined to bed” ($P(A > B) = 1.00$)

[a] One-sided *P* values for a *t* test on whether the proportion is >0.98. Sample size varied (52–64) due to sampling. Each respondent completed only 1 pair (no missingness)

Table 2

Proportion of respondents who prefer the better statement: EQ-5D-5L

Better statement (*A*)	**Worse statement (*B*)**	**P ($A > B$)**	**N**	**$P > 0.98$[a]**
Level 2 to Level 3				
Slight problems in walking about	Moderate problems in walking about	0.98	59	0.53
Slight problems doing usual activities	Moderate problems doing usual activities	0.92	53	0.06
Slight problems washing or dressing self	Moderate problems washing or dressing self	0.97	60	0.26
Slight pain or discomfort	Moderate pain or discomfort	0.97	61	0.26
Slightly anxious or depressed	Moderately anxious or depressed	1.00	58	1.00
Level 2 to Level 4				
Slight problems in walking about	Severe problems in walking about	0.95	63	0.14
Slight problems doing usual activities	Severe problems doing usual activities	0.98	54	0.49
Slight problems washing or dressing self	Severe problems washing or dressing self	1.00	55	1.00
Slight pain or discomfort	Severe pain or discomfort	0.98	63	0.56
Slightly anxious or depressed	Severely anxious or depressed	1.00	59	1.00
Level 3 to Level 4				
Moderate problems in walking about	Severe problems in walking about	0.98	59	0.53
Moderate problems doing usual activities	Severe problems doing usual activities	0.98	56	0.51
Moderate problems washing or dressing self	Severe problems washing or dressing self	0.97	62	0.27
Moderate pain or discomfort	Severe pain or discomfort	0.98	54	0.49
Moderately anxious or depressed	Severely anxious or depressed	1.00	53	1.00
Level 3 to Level 5				
Moderate problems in walking about	Unable to walk about	1.00	54	1.00
Moderate problems doing usual activities	Unable to do usual activities	0.97	60	0.26
Moderate problems washing or dressing self	Unable to wash or dress self	1.00	59	1.00
Moderate pain or discomfort	Extreme pain or discomfort	0.98	55	0.50
Moderately anxious or depressed	Extremely anxious or depressed	0.98	58	0.52
Level 4 to Level 5				
Severe problems in walking about	Unable to walk about	0.88	57	0.01
Severe problems doing usual activities	Unable to do usual activities	0.89	57	0.02
Severe problems washing or dressing self	Unable to wash or dress self	0.91	56	0.03
Severe pain or discomfort	Extreme pain or discomfort	0.84	55	<0.01
Severely anxious or depressed	Extremely anxious or depressed	0.49	59	<0.01

For example, 29 out of 59 respondents prefer “severely anxious or depressed” over “extremely anxious or depressed” (P ($A > B$) = 0.49)

[a] One-sided P values for a t test on whether the proportion is >0.98. Sample size varied (53–63) due to sampling. Each respondent completed only 1 pair (no missingness)

Table 3

Proportion of respondents who prefer the better statement: a comparison of levels across EQ-5D-3L, EQ-5D-Y, and EQ-5D-5L

Better statement (*A*)	**Worse statement (*B*)**	***P* (*A* > *B*)**	***N***	***P* > 0.98**[a]
Severe problems in walking about (EQ-5D-5L Mobility, Level 4)	Confined to bed (EQ-5D-3L Mobility, Level 3)	0.96	55	0.24
A lot of pain or discomfort (EQ-5D-Y Pain/Discomfort, Level 3)	Severe pain or discomfort (EQ-5D-5L Pain/Discomfort, Level 4)	0.96	56	0.24
Very worried, sad, or unhappy (EQ-5D-Y Anxiety/Depression, Level 3)	Severely anxious or depressed (EQ-5D-5L Anxiety/Depression, Level 4)	0.90	60	0.02
Unable to walk about (EQ-5D-5L Mobility, Level 5)	Confined to bed (EQ-5D-3L Mobility, Level 3)	0.92	59	0.03
A lot of pain or discomfort (EQ-5D-Y Pain/Discomfort, Level 3)	Extreme pain or discomfort (EQ-5D-5L Pain/Discomfort, Level 5)	1.00	62	1.00

For example, 53 out of 55 respondents prefer "severe problems in walking about" over "confined to bed" (*P* (*A* > *B*) = 0.96)

[a] One-sided *P* values for a *t* test on whether the proportion is >0.98. Sample size varied (55–62) due to sampling. Each respondent completed only 1 pair (no missingness)