Article:

Finch, A.P., Brazier, J. orcid.org/0000-0001-8645-4780 and Mukuria, C. orcid.org/0000-0003-4318-1481 (2021) Selecting bolt-on dimensions for the EQ-5D: testing the impact of hearing, sleep, cognition, energy, and relationships on preferences using pairwise choices. Medical Decision Making, 41 (1). pp. 89-99. ISSN 0272-989X

https://doi.org/10.1177/0272989x20969686

Original Article

MDM
Medical Decision Making

Medical Decision Making
2021, Vol. 41(1) 89–99

DOI: 10.1177/0272989X20969686
journals.sagepub.com/home/mdm
SAGE

Selecting Bolt-on Dimensions for the EQ-5D: Testing the Impact of Hearing, Sleep, Cognition, Energy, and Relationships on Preferences Using Pairwise Choices

Aureliano Paolo Finch, John Brazier, and Clara Mukuria

Background. Generic preference-based measures (GPBMs) such as the EQ-5D are valid across many conditions, but in some cases, "bolting on" additional dimensions may improve validity. The selection of "bolt-ons" has been based on the psychometric impact of individual dimensions, but preferences provide another important way to select them. This study aims to test the potential of using pairwise choices to inform the selection of bolt-ons for the EQ-5D-5L. **Methods.** General population preferences were collected using an online survey of 1040 UK residents. Three EQ-5D-5L health state pairs were selected based on pairs that had a 50:50 split in respondent preferences from a previous pairwise survey. Participants were presented with pairwise choices of EQ-5D-5L health states without and with bolt-ons of hearing, sleep, cognition, energy, and relationships, each added individually. Logistic models were used to assess the impact of bolt-ons, as well as bolt-ons at different severity levels, on the log odds of responders choosing between health states. **Results.** Preferences varied according to the bolt-ons and their severity level (only levels 1, 3, and 5 were used). Additions of bolt-ons at level 1 generally resulted in nonstatistically significant differences while additions of bolt-ons at level 3 and level 5 produced a negative and statistically significant impact on preferences for the health state with the bolt-on. At level 5, hearing had the largest impact, followed by cognition, relationships, energy, and sleep. At level 3, cognition produced the largest impact, followed by hearing and sleep with similar impacts, energy, and relationships. This ordering offers information for bolt-on selection, with hearing and cognition appearing as the most important. The weight placed on the different health problems is not constant across severity levels between bolt-ons. **Conclusions.** Pairwise choices provide a cost-effective approach of generating information on preferences to support bolt-on selection.

Keywords
bolt-on, cognition, EQ-5D, hearing, preferences

Date received: January 2, 2019; accepted: September 15, 2020

Quality-adjusted life years (QALYs) are the main outcome measure of economic evaluations comparing health interventions. QALYs provide an index of the length of life and the health-related quality of life (HRQoL) of patients, where quality of life is valued on a 0 (dead) to 1 (full health) scale and is usually estimated with generic preference-based measures (GPBMs). GPBMs include a health classification or descriptive system that is completed by patients and a tariff, typically derived using choice-based valuation methods such as time tradeoff (TTO) in a general population sample. A number of GPBMs exist, and they differ in the health dimensions they cover and the valuation methods they use.[1] Among

Corresponding Author
Aureliano Paolo Finch, Euroqol Research Foundation, Marten Meesweg 107, Rotterdam, 3068 AV, The Netherlands (finch@euroqol.org).

them, the EQ-5D EuroQol Research Foundation is the most widely employed internationally,[2,3] and it is the reference case for a number of health technology assessment bodies such as the National Institute of Health and Care Excellence.[4]

The EQ-5D is intended to be applicable across all conditions, and since its development, it has been validated in numerous disease areas.[5] However, evidence suggests that the measure might lack validity and responsiveness for some specific patient groups.[5,6] When this happens, HRQoL values can be obtained using a different GPBM or a preference-based disease-specific measure that might cover the missing dimensions.[1] GPBMs differ substantially in their content,[1] covering a large spectrum of items. While the EQ-5D mostly focuses on physical health, other measures such as the SF-6D, HUI 3, AQoL, and 15D include more domains related to mental constructs such as anger, energy, and intimacy, as well as social constructs such as relationships; the content of disease-specific preference-based measures varies even more.[1] In addition, these measures use different valuation methods.[1] This reduces the comparability across assessments and undermines consistency between decisions.[7–10] An alternative that might meet these concerns with less detrimental impact for comparability is the use of bolt-ons.

Bolt-ons are dimensions that can be added to the descriptive system of a GPBM,[11] resulting in an extended measure that retains the original dimensions plus the new bolt-ons. Bolt-on studies have investigated the effect of adding energy,[12] cognition,[13] sleep,[14] vision, hearing and tiredness,[15] and skin irritation[16] to the EQ-5D. These studies selected bolt-ons based on poor psychometric characteristics of EQ-5D in a relevant condition or simply argued that a relevant dimension was missing. The psychometric evidence used to identify missing dimensions is often weak, as the tests used are not always appropriate for GPBMs (for a detailed discussion, see Brazier and Deverill[17]). Moreover, many of these studies were not based on systematic reviews of the performance of the EQ-5D.

School of Health and Related Research, University of Sheffield, Sheffield, UK (APF, JB, CM); EuroQol Research Foundation, Rotterdam, the Netherlands (APF). The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support for this study was provided by a grant from the EuroQol Group. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. AF was funded through a scholarship from the School of Health and Related Research, University of Sheffield.

An alternative source of potential bolt-ons is the dimensions available from other GPBMs, which have been shown to cover the majority of health and nonhealth domains of interest.[3] Eight candidate bolt-ons for the EQ-5D-5L were previously identified and tested using factor analysis of items from other GPBMs, including relationships, hearing, life satisfaction, speech, cognition, vision, energy/vitality, and sleep.[18,19] This candidate list is not likely to be complete, but it demonstrates the potential number of candidate dimensions.

Even if this list of 8 bolt-ons does not address all potential additional dimensions, it would still benefit from further reduction for 2 main reasons. First, adding all bolt-ons simultaneously would substantially reduce the feasibility for the next step of valuing the EQ-5D. Second, even if bolt-ons were added individually so that only one is used at a time, the process of developing and appending bolt-ons might be costly and complex. Evidence to date suggests that the impact of bolt-ons is not simply additive but also has an impact on the coefficients of the core 5 dimensions.[4] This implies that there would need to be a complete revaluation of the EQ-5D for each bolt-on combination.

Selecting bolt-ons is the product of 2 criteria. One is whether the bolt-on improves the psychometric performance of the EQ-5D. This can be examined in a number of ways, and 2 recent studies have looked at using structural equation modeling (SEM)[18] and impact on well-being using regression analyses.[20] These approaches are useful, but they do not consider the second important criterion, namely, whether bolt-ons have an impact on preferences for health states. This is a crucial aspect in GPBMs, as their ultimate goal is detecting health decrements that are considered relevant for influencing responders' choices to preference elicitation tasks.

The few existing studies investigating the impact of bolt-ons on utility values for the EQ-5D used TTO administered face to face[14,15] and show that some dimensions such as cognition[13] have a large impact on preferences while others such as sleep[14] do not. Each study separately assessed single bolt-ons rather than undertaking head-to-head comparisons of bolt-ons in the same population. One of the challenges of doing comparative work is the resource implication of undertaking TTO. TTO is also cognitively demanding and may make it more difficult for participants if they have to consider more than 1 additional dimension. Pairwise choices provide a flexible environment for examining the impact on preferences to more conventional methods such as the TTO. Pairwise comparisons have been widely used in a self-administered format, and they are arguably easier[21] and less cognitively demanding[22] than other techniques.

This study has 2 objectives: it examines whether the bolt-on dimensions (identified through previous psychometric analysis) change preferences for health states described in terms of the EQ-5D-5L, and it examines the degree of this change comparatively across the bolt-ons to see which of them has a larger impact on preferences as reflected in the latent scale produced by a discrete choice experiment (DCE). If a bolt-on has an impact on preferences, this implies it would likely alter the value set of the investigated GPBM.

Methods

The pairwise comparison tasks were developed and administered to a representative sample of the UK general population. Responders were asked to express their preference between pairs of health profiles. The next section describes some of the key aspects of the experiment.

Bolt-on Dimension Selection and Development

A selection had to be made of possible bolt-ons from the candidate list identified by Finch et al.,[5,19,20] which included life satisfaction, speech, vision, hearing, sleep, cognition, energy, and relationships, among others. The bolt-ons examined in this study are hearing, sleep, cognition, energy, and relationships. Relationships, energy, and hearing were selected as they had large, moderate, and small coefficients, respectively, when regressed over a proxy of HRQoL, the Health Visual Analogue Scale (VAS).[20] Cognition and sleep were selected as the former had a large impact on preferences for health states in one study,[13] while the latter did not have a significant impact on preferences for health states in another.[14]

Descriptors and labels of bolt-on dimensions were developed to closely resemble the format of the EQ-5D-5L. These were assessed by the research team in terms of their coherence with the EQ-5D-5L wording, their suitability for the lay public, and their consistency across dimensions and with the construct measured. Where there were inconsistencies, descriptors and labels were reworded and initial wordings replaced. If it was not possible to establish the best wording, then multiple wordings were examined in the face validity testing phase.

The face validity of bolt-on variants was tested in 2 focus groups. The first focus group recruited 5 members of the general public, and the second focus group had 6 patients affected by chronic health conditions (i.e., chronic obstructive pulmonary disease, type 1 diabetes, chronic fatigue syndrome, and endometriosis). Participants were asked to comment on the bolt-ons relevance, clarity (ease of understanding and responding to), and acceptability in different populations. A topic guide was used to aid the discussion. Focus group recordings were analyzed using content analysis, which involved a systematic identification of sections of the transcripts related to the aspects of interest such as clarity. Results were used to modify descriptors and labels if face validity problems were identified and also to select the best descriptors where multiple wording was presented. Final descriptors and labels for the 5 bolt-ons are presented in Table 1. Alternative descriptors and labels, as well as bolt-ons descriptors and labels for life satisfaction, speech, and vision, are presented in Supplemental Table S1.

Table 1 Descriptors and Labels of Bolt-on Dimensions Tested in the Pairwise Experiment[a]

Hearing	
I have no problems hearing	☐
I have slight problems hearing	☐
I have moderate problems hearing	☐
I have severe problems hearing	☐
I am unable to hear	☐
Sleep	
I have no problems sleeping	☐
I have slight problems sleeping	☐
I have moderate problems sleeping	☐
I have severe problems sleeping	☐
I have extreme problems sleeping	☐
Relationships	
I have no problems with my social relationships	☐
I have slight problems with my social relationships	☐
I have moderate problems with my social relationships	☐
I have severe problems with my social relationships	☐
I am unable to have social relationships	☐
Energy	
I have no problems with my energy levels	☐
I have slight problems with my energy levels	☐
I have moderate problems with my energy levels	☐
I have severe problems with my energy levels	☐
I have extreme problems with my energy levels	☐
Cognition	
I have no problems with remembering things	☐
I have slight problems with remembering things	☐
I have moderate problems with remembering things	☐
I have severe problems with remembering things	☐
I am unable to remember things	☐

[a]Reproduced by permission of EuroQol Research Foundation. Reproduction of this version is not allowed. For reproduction, use or modification of the EQ-5D (any version), please register your study by using the online EQ registration page: www.euroqol.org.

Selection of EQ-5D-5L Health States Pairs and Bolt-on Levels

Given the large number of bolt-on dimensions, a decision had to be made between selecting numerous health states with fewer respondents per health state and selecting a smaller number of health states but eliciting preferences from a larger sample of respondents. As this was the first study using pairwise choices to support selection of bolt-ons, the latter was chosen to increase the confidence in the results obtained, which in turn can better inform future research.

To develop the health states, EQ-5D-5L pairs were needed on to which the selected bolt-on dimensions could be added. The ideal pairs of EQ-5D-5L health states would be those in which there was equal preference (50:50) across each pairwise choice to maximize the ability to assess the impact on preferences when bolt-ons were added. Three pairs of health states were selected from the valuation study used to develop the EQ-5D-5L tariff for England,[22,23] which employed both TTO and discrete-choice experiments based on pairwise choices. The selected pairs of health states are health state pair 1 (state A 11122 v. state B 23111), health state pair 2 (state A 52211 v. state B 11325), and health state pair 3 (state A 33142 v. state B 34333). (EQ-5D-5L has 5 dimensions and the numbers represent the severity levels: 1, no problem; 2, slight problems; 3, moderate problems; 4, severe problems; and 5, extreme problems/unable.)

There are 25 possible combinations of bolt-on levels for each bolt-on for each pairwise choice (i.e., 1 v. 1, 1 v. 2, 1 v. 3, etc.). Due to resource limitations, it was not feasible to test all these possible combinations. Hence, 3 levels per bolt-on were chosen for this study: levels 1, 3, and 5. The first level was included since it allows an assessment of whether the simple presence of a bolt-on dimension changes preferences for the pairs of health states presented. The third and fifth levels were selected to allow investigation of the impact of severity on the relative importance of bolt-ons. The selected health states and bolt-on levels used in this study are presented in Table 2.

There were therefore 3 EQ-5D health state pairs and 5 bolt-ons with 3 severity levels each investigated in this study. Bolt-ons at severity 3 and 5 were always added to health state A in each pair (i.e., 11122, 52211, and 33142). In total, 48 pairwise questions were included in the survey, 3 of which did not include bolt-ons.

Table 2 Choice Questions by Bolt-ons

Bolt-on	Health State A	Bolt-on Level	Health State B	Bolt-on Level
No bolt-on	11122	—	23111	—
No bolt-on	52211	—	11325	—
No bolt-on	33142	—	34333	—
Hearing	11122	1	23111	1
Hearing	11122	3	23111	1
Hearing	11122	5	23111	1
Hearing	52211	1	11325	1
Hearing	52211	3	11325	1
Hearing	52211	5	11325	1
Hearing	33142	1	34333	1
Hearing	33142	3	34333	1
Hearing	33142	5	34333	1
Sleep	11122	1	23111	1
Sleep	11122	3	23111	1
Sleep	11122	5	23111	1
Sleep	52211	1	11325	1
Sleep	52211	3	11325	1
Sleep	52211	5	11325	1
Sleep	33142	1	34333	1
Sleep	33142	3	34333	1
Sleep	33142	5	34333	1
Relationships	11122	1	23111	1
Relationships	11122	3	23111	1
Relationships	11122	5	23111	1
Relationships	52211	1	11325	1
Relationships	52211	3	11325	1
Relationships	52211	5	11325	1
Relationships	33142	1	34333	1
Relationships	33142	3	34333	1
Relationships	33142	5	34333	1
Energy	11122	1	23111	1
Energy	11122	3	23111	1
Energy	11122	5	23111	1
Energy	52211	1	11325	1
Energy	52211	3	11325	1
Energy	52211	5	11325	1
Energy	33142	1	34333	1
Energy	33142	3	34333	1
Energy	33142	5	34333	1
Cognition	11122	1	23111	1
Cognition	11122	3	23111	1
Cognition	11122	5	23111	1
Cognition	33142	1	34333	1
Cognition	33142	3	34333	1
Cognition	33142	5	34333	1
Cognition	52211	1	11325	1
Cognition	52211	3	11325	1
Cognition	52211	5	11325	1

Survey Administration and Design

The pairwise choices were administered in an online survey. The survey had 4 components presented in the following order: 1) background and sociodemographic questions, 2) self-reported health assessed through the EQ-5D-5L + bolt-ons, 3) familiarization session, and 4) 8 pairwise choice questions. Each pairwise comparison asked respondents to select the profile they preferred (an example of the pairwise question for a bolt-on at level 3 is presented in Figure 1). No indifference option was provided, in line with previous research,[22] which implied that respondents had to choose option A or option B.

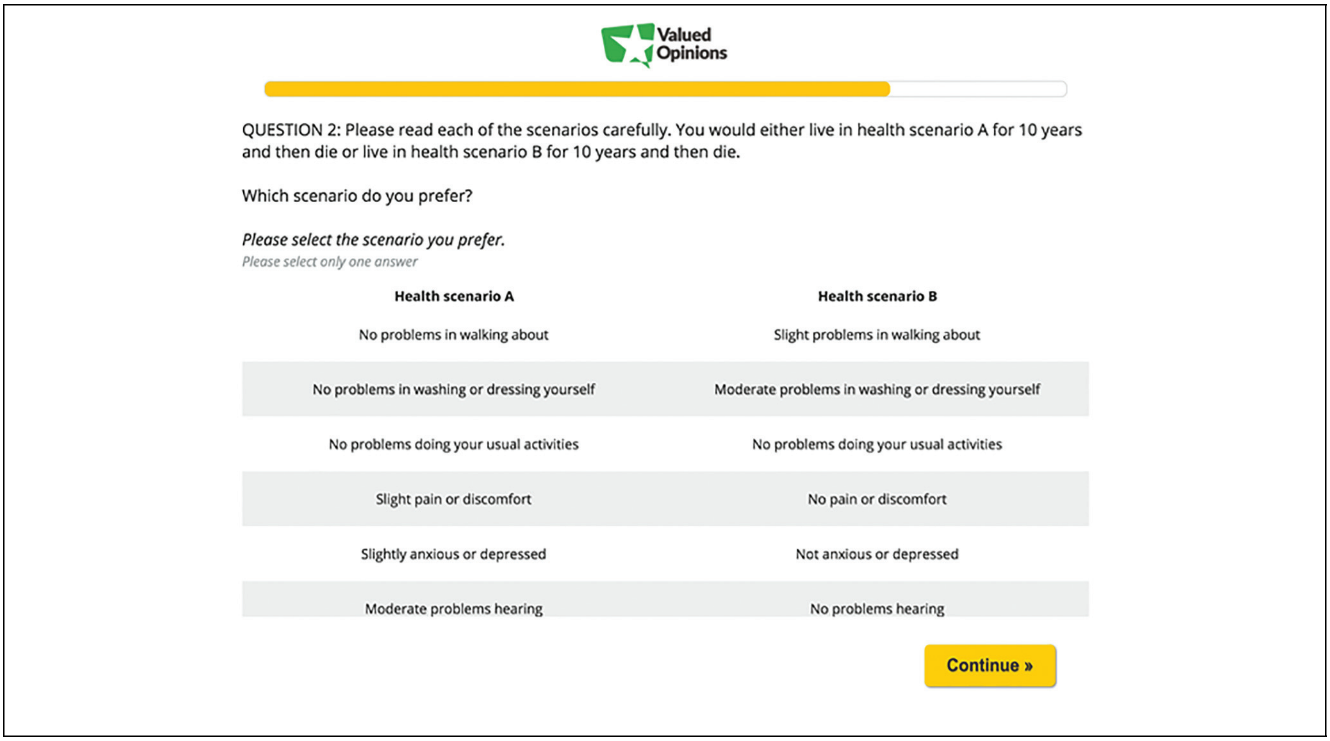

Figure 1 Example of pairwise comparison as presented to responders.

A design was employed for the survey, in which each pairwise question was assigned to a block. Each block included 8 pairwise questions. This was considered a feasible number of tasks per participant based on previous research.[24] Participants allocated to blocks 1, 2, and 3 completed 1 task comparing pairs of EQ-5D-5L states without bolt-ons in each block and 7 tasks comparing pairs of EQ-5D-5L states with bolt-ons. Participants allocated to blocks 4, 5, and 6 completed 8 tasks comparing pairs of EQ-5D-5L states all with bolt-ons. To avoid focusing effects, each block included all 3 EQ-5D-5L health state pairs at least once and different combinations of bolt-ons.

The survey presented 2 levels of randomization. First, participants were randomized to 1 of the 6 blocks (although the order within each block was not randomized). Subsequently, a randomization of the side in terms of which options participants saw as option A and option B was performed to avoid any position bias.

Sample

It was estimated that to detect a 10% difference between responses with and without a bolt-on dimension, using a 2-sided test with a power of 0.8 and significance level of 1%, 170 responders per health state pair were required.[25] As the 48 health state pairs were presented in 6 blocks, the target sample for the study was of 1020 participants (i.e., 170 * 6).

Participants were recruited using an existing UK online panel administered by Research Now, a market research company, using quotas for sex, age, education, whether they had children, religion, and marital status to achieve a representative sample. The panel is made up of individuals who have previously signed up to answer surveys in return for points that can be exchanged for goods. Each responder used a weblink to access the survey and was for this reason able to self-complete it at his or her own convenience after providing informed consent. The survey was administered in May 2017. The University of Sheffield provided ethical approval.

Analyses

The background characteristics of the participants allocated to the different blocks were compared. Fisher exact test and χ^2 tests were used to identify the presence of statistically significant differences in age, sex, and social and economic status across the 6 blocks.

To test whether bolt-ons had a statistically significant impact on preferences, we used logistic regressions with clustered sandwich estimators. Use of cluster sandwich estimators accounted for the possible intraindividual correlation generated by the panel structure of the data. First, a logistic model was estimated by regressing respondents' choice over dummy variables signaling the presence or absence of each bolt-on (i.e., hearing, sleep, cognition, energy, relationships, and a dummy identifying the health state pairs). To ensure the generalizability of our findings, we tested confounding by adjusting for background sociodemographic characteristics. Marginal effects (i.e., log odds ratios for the bolt-ons β coefficients) are presented.

Second, we tested the null hypothesis that additions of bolt-on level dummies to the main effect bolt-on model resulted in a significant improvement in model fit. The Wald test (i.e., pseudo-score test) was used for this purpose. Following the Wald test results, a main effect model using dummies for each bolt-on level and for the health state pairs was estimated. Marginal effects (i.e., log odds ratios for the β coefficients of the bolt-on level dummies) are presented, once again adjusting for background sociodemographics.

Third, to further assess bolt-ons' impact on different health state pairs, models were estimated separately for each of the bolt-on options at level 1, level 3, or level 5, for each of the 3 pairs investigated. Marginal effects (i.e., log odds ratios for the β coefficients for each bolt-on for each level and health state) are once again presented.

For all models, marginal effects were used to compare the impact of different bolt-ons. For example, the marginal effect for health state 11122 with hearing at level 5 was compared with the marginal effect for the health state 11122 with relationships at level 5 and so on. Analyses were conducted using STATA/MP 14.1 (StataCorp, Cary, NC).

Results

In total, 1581 individuals entered the survey, but 342 were excluded as they did not select all options in the consent form. A further 169 were excluded as they did not complete the survey, 5 as they "speeded" through the survey (threshold for speeding is calculated as the median completion time divided by 3), and 25 as their quota was already full. The final analysis set comprised 1040 participants. Each pairwise choice was completed by a minimum number of 167 respondents and a maximum of 180, depending on the block. The mean time taken to completion was 9.19 minutes (range, 2.12–245.33 minutes), and the median time was 7.05 minutes. Participants in block 2 took the shortest mean time (7.23 minutes), while participants in block 4 took the longest mean time (10.52 minutes). Background characteristics and health of the sample are presented in Table 3. No statistically significant differences were found between participants allocated to the 6 blocks in terms of sex, marital status, profession, and highest education achieved. Differences in age were seen between blocks 1 and 4 ($P < 0.001$), with block 1 appearing more normally distributed with the majority of respondents being 45 to 54 years old, and block 4 presenting a relatively uniform number of responders in each age category. Self-reported health status was generally similar across blocks except for responders in block 2 who reported more problems in self-care than responders in block 6, and this difference was statistically significant ($P < 0.05$).

In the main effect bolt-on model, hearing had the largest impact (–0.16), followed by cognition (–0.15), relationships (–0.12), sleep (–0.09), and energy (–0.08). None of the sociodemographic confounding variables had a statistically significant impact on the results, as well as health state pairs.

When testing extensions of the bolt-on main effect model using the Wald test, the null hypothesis that dummies for bolt-ons at levels 3 and 5 were simultaneously equal to zero was rejected, showing that their inclusion resulted in a statistically significant improvement in model fit. Additions of bolt-ons at level 1 did not result in a significant improvement in model fit.

When estimating main effects for the bolt-on levels, cognition had the largest marginal effect at level 3 (–0.16). For the rest of the domains, at level 3, relationships had a small marginal effect (–0.04) and energy the second smallest (–0.05). Sleep and hearing reported the same marginal effects (–0.12). At level 5, hearing had the largest impact (–0.43), followed by cognition (–0.31), relationships (–0.29), energy (–0.27), and sleep (–0.19). Table 4 reports the marginal effects and clustered sandwich estimator standard errors for the bolt-on main effects, as well as the bolt-on level main effects for all health state pairs together, and provides a ranking of bolt-on marginal effects for severity levels 3 and 5. Once again, none of the sociodemographic confounding variables had a statistically significant impact on the results, as well as health state pairs.

Table 5 presents the marginal effects of the β coefficients from the logistic models with clustered sandwich estimators' standard errors and ranking of bolt-ons for severity levels 3 and 5 in terms of marginal effect size. The marginal effects for all bolt-ons at level 1 were

Table 3 Background Characteristics and Health of the Sample

Characteristic	No. (%)
Sex	
Female	520 (50)
Male	520 (50)
Age, y	
18−24	99 (10)
25−34	216 (21)
35−44	220 (21)
45−54	227 (22)
55−64	160 (15)
65+	118 (11)
Status	
Single	374 (36)
Married	597 (57)
Separated	42 (4)
Widowed	20 (2)
Prefer not to say	7 (1)
Education	
O-level/GCSE	246 (24)
A-level	198 (19)
Diploma	105 (10)
First degree	286 (28)
Postgraduate degree	163 (16)
Other	42 (4)
Employment	
In employment	650 (63)
Retired	149 (14)
Homemaker	71 (7)
Student	71 (7)
Seeking work	17 (2)
Unemployed	34 (3)
Long-term sick	42 (4)
Other	6 (1)
Children	
Yes	545 (52)
No	495 (48)
Religion	
Yes	300 (29)
No	705 (68)
Prefer not to say	35 (3)
Self-reported health status dimension	
Mobility	
Level 1	802 (77)
Level 2	159 (15)
Level 3	52 (5)
Level 4	26 (3)
Level 5	1 (0)
Self-care	
Level 1	940 (90)
Level 2	67 (6)
Level 3	29 (3)
Level 4	4 (0)
Level 5	0 (0)
Usual activities	
Level 1	795 (76)
Level 2	151 (15)
Level 3	65 (6)
Level 4	22 (2)
Level 5	7 (1)
Pain/discomfort	
Level 1	533 (50)
Level 2	362 (35)
Level 3	111 (11)
Level 4	28 (3)
Level 5	6 (1)
Anxiety/depression	
Level 1	565 (56)
Level 2	266 (23)
Level 3	134 (13)
Level 4	54 (5)
Level 5	21 (2)
Hearing	
Level 1	850 (82)
Level 2	141 (14)
Level 3	41 (4)
Level 4	8 (1)
Level 5	0 (0)
Sleep	
Level 1	487 (47)
Level 2	338 (33)
Level 3	137 (13)
Level 4	65 (6)
Level 5	13 (1)
Cognition	
Level 1	658 (63)
Level 2	311 (30)
Level 3	63 (6)
Level 4	7 (1)
Level 5	1 (0)
Energy	
Level 1	447 (43)
Level 2	365 (35)
Level 3	163 (16)
Level 4	52 (5)
Level 5	13 (1)
Relationships	
Level 1	729 (70)
Level 2	204 (20)
Level 3	66 (6)
Level 4	30 (3)
Level 5	11 (1)

GCSE, General Certificate of Secondary Education.

generally small and not statistically significant. For example, in health state pair 2, energy at level 1 had a nonstatistically significant marginal effect of −0.01, while hearing at level 1 had a nonstatistically significant marginal effect of –0.00 in health state pair 3. Of all bolt-ons tested at level 1, only relationships reported a statistically significant marginal effect and only for health state pair 3 (i.e., −0.18).

Table 4 Marginal Effects and Clustered Sandwich Estimator Standard Errors for the Bolt-on Main Effect and Bolt-on Level Main Effects for All Health State Pairs and Ranking of Bolt-ons for Severity Levels 3 and 5 in Terms of Marginal Effect Size, Adjusting for Sociodemographics

Bolt-on	Bolt-on Level	Marginal Effect (SE)	Ranking by		
			Overall	Level 3	Level 5
Hearing	All levels	−0.16 (0.01)[a]	1	–	–
Sleep	All levels	−0.09 (0.01)[a]	5	–	–
Cognition	All levels	−0.15 (0.01)[a]	2	–	–
Energy	All levels	−0.08 (0.01)[a]	4	–	–
Relationships	All levels	−0.12 (0.01)[a]	3	–	–
Hearing	Level 1	0.02 (0.03)	–	2	1
	Level 3	−0.12 (0.02)[a]			
	Level 5	−0.43 (0.02)[a]			
Sleep	Level 1	−0.03 (0.02)	–	2	5
	Level 3	−0.12 (0.02)[a]			
	Level 5	−0.19 (0.02)[a]			
Cognition	Level 1	−0.02 (0.02)	–	1	2
	Level 3	−0.16 (0.02)[a]			
	Level 5	−0.31 (0.02)[a]			
Energy	Level 1	0.04 (0.02)	–	4	4
	Level 3	−0.05 (0.02)[a]			
	Level 5	−0.27 (0.02)[a]			
Relationships	Level 1	−0.03 (0.02)[a]	–	5	3
	Level 3	−0.04 (0.02)[b]			
	Level 5	−0.29 (0.02)[a]			

–, Not applicable.
[a]Statistically significant at $P < 0.01$.
[b]Statistically significant at $P < 0.05$.

Table 5 Marginal Effects of the β Coefficients from the Logistic Models with Clustered Sandwich Estimators' Standard Errors and Ranking of Bolt-ons for Severity Levels 3 and 5 in Terms of Marginal Effect Size, Adjusting for Sociodemographics[a]

Bolt-on	Bolt-on Level	Health State Pair 1			Health State Pair 2			Health State Pair 3		
		ME (SE)	Ranking by		ME (SE)	Ranking by		ME (SE)	Ranking by	
			Level 3	Level 5		Level 3	Level 5		Level 3	Level 5
Hearing	Level 1	−0.09 (0.05)			−0.06 (0.05)			−0.00 (0.05)		
	Level 3	−0.22 (0.05)[b]	3	1	−0.03 (0.03)	4	1	−0.26 (0.05)[b]	2	1
	Level 5	−0.43 (0.04)[b]			−0.29 (0.05)[b]			−0.43 (0.05)[b]		
Sleep	Level 1	−0.07 (0.05)			0.04 (0.05)			−0.07 (0.04)		
	Level 3	−0.23 (0.05)[b]	2	5	−0.11 (0.05)[c]	2	3	−0.20 (0.05)[b]	3	5
	Level 5	−0.29 (0.06)[b]			−0.21 (0.05)[b]			−0.24 (0.05)[b]		
Cognition	Level 1	−0.09 (0.05)			−0.07 (0.05)			−0.07 (0.05)		
	Level 3	−0.27 (0.05)[b]	1	2	−0.11 (0.05)[c]	2	4	−0.31 (0.05)[b]	1	2
	Level 5	−0.40 (0.04)[b]			−0.18 (0.05)[b]			−0.39 (0.05)[b]		
Energy	Level 1	0.01 (0.05)			−0.01 (0.05)			−0.06 (0.05)		
	Level 3	−0.16 (0.05)[b]	5	3	0.01 (0.03)	5	4	−0.20 (0.05)[b]	4	4
	Level 5	−0.37 (0.05)[b]			−0.18 (0.05)[b]			−0.33 (0.05)[b]		
Relationships	Level 1	−0.04 (0.05)			−0.07 (0.05)			−0.18 (0.05)[b]		
	Level 3	−0.18 (0.05)[b]	4	4	−0.14 (0.05)[c]	1	2	−0.19 (0.05)[b]	5	3
	Level 5	−0.36 (0.05)[b]			−0.22 (0.05)[b]			−0.38 (0.05)[b]		

ME, marginal effect; SE, robust standard error.
[a]Health state pair 1: state A 11122 v. state B 23111; health state pair 2: state A 52211 v. state B 11325; health state pair 3: state A 33142 v. state B 34333. Bolt-ons at severity 3 and 5 were always added on health states A.
[b]Statistically significant at $P < 0.01$.
[c]Statistically significant at $P < 0.05$.

Bolt-ons at level 3 generally resulted in statistically significant negative marginal effects in health state pairs 1 and 3. In health state pair 2, marginal effects for hearing and energy were small and not statistically significant. Marginal effects ranged between −0.27 (cognition) and −0.16 (energy) in health state pair 1 and between −0.31 (cognition) and −0.19 (relationships) in health state pair 3, while they were smaller when a bolt-on at level 3 was added to health state pair 2, ranging between −0.11 (relationships) and 0.01 (energy). Cognition had the largest marginal effects in all 3 pairs, showing that this bolt-on substantially reduced the probability of responders choosing the health state to which it was added. By contrast, energy at level 3 had the smallest marginal effect for health states 1 and 2 and the second smallest marginal effect for health state pair 3. There was no consistent ordering for sleep, hearing, and relationships.

As expected, bolt-ons at level 5 always resulted in statistically significant negative marginal effects, which were smaller than marginal effects for bolt-ons at level 3. Marginal effects ranged between −0.43 (hearing) and −0.29 (sleep) in health state pair 1, between −0.29 (hearing) and −0.18 (energy and cognition) in health state pair 2, and between −0.43 (hearing) and −0.24 (sleep) in health state pair 3. Hearing consistently reported the largest marginal effects for all 3 health state pairs. This was followed by cognition, which reported the second largest marginal effect in health state pairs 1 and 3. Sleep had the lowest impact in 2 of the 3 health state pairs (1 and 3).

Discussion

This study investigated the potential of using pairwise comparisons to determine whether bolt-on dimensions previously identified through factor analysis change preferences for EQ-5D-5L health states. The aim was to test the use of a simple low-cost pairwise comparison as a method for informing the selection of bolt-on dimensions.

The study showed that each of the individual bolt-ons had a significant impact on preferences for the EQ-5D-5L. The extent of this impact varied according to the bolt-ons and their severity level, as well as the health states to which they were added. Additions of bolt-ons at level 1 generally resulted in small and not statistically significant marginal effects. Additions of bolt-ons at level 3 generally produced negative and statistically significant marginal effects, showing a reduction in the log odds ratio of individuals choosing the health state to which the moderate level was added. Additions of bolt-ons at level 5 generally resulted in even larger negative marginal effects compared to bolt-ons at level 3. The dimensions that had the largest impact were hearing and cognition, while sleep and energy had less impact.

These findings agree with those of previous research in that they show that hearing and cognition make a significant impact on the judgments people place on the EQ-5D health states.[13,15] Our findings also show that sleep has less impact on preferences for EQ-5D health states, although in contrast to a previous study, its impact was significant.[14]

This study found that at severity level 5, hearing had the largest marginal effect, followed by cognition, relationships, and energy with relatively similar marginal effects and sleep with the smallest marginal effect. By contrast, at severity level 3, cognition reported the largest marginal effect, followed by sleep, hearing, and relationships, with energy registering the smallest marginal effect. This suggests that the relative weight responders place on different health problems is not constant across levels of severity between bolt-ons. This finding is relevant for selecting bolt-on dimensions, as it highlights the need for a judgment on what decision rule needs to be followed. One possibility might be choosing the bolt-ons that have the greatest impact on preferences, in terms of log odds, compared to the same health state without bolt-ons, based on the worst severity level. Alternatively, bolt-ons might be selected based on the overall log odds (i.e., main effect bolt-on model). Either way, other considerations remain fundamental for the final selection, such as what other dimensions are already present in the descriptive system of the examined measure and the psychometric evidence for the impact of the bolt-on for the performance of the measure in terms of validity and responsiveness in the population of interest.

Another important issue is that the inclusion of bolt-ons in the parent measure requires a complete revaluation of the extended measure's tariff. This study provides some evidence showing that the addition of a bolt-on at level 1 might have little impact on the value of the remaining core items/dimensions of the GPBM, although it does not clarify what impact the bolt-ons at levels 3 and 5 would have on the core dimensions of the parent measure. It still needs to be clarified whether interactions are present to inform which modeling approach should be preferred in bolt-on studies.

This study has a number of limitations. Only 3 pairs of EQ-5D-5L health states were selected. This design appropriately responded to the methodological questions investigated in this study, but it has the disadvantage of not being powered for advanced econometric investigations (e.g., interactions). Moreover, previous research has shown that preferences for bolt-ons might vary depending on the severity of the health states to which they are added,[15] and for this reason, using other pairs from the 3 selected might have generated different results. The results reported in this study show that the relative weight responders place on different health problems is not constant but rather depends on the severity level of the additional dimension, and further testing with levels 2 and 4 for the same bolt-ons is required to confirm these findings. In addition, although each block included different bolt-ons and health state pairs to minimize focusing effects, the order in which pairwise questions were presented within each block was not randomized, which may have an impact on preferences. Finally, the study used pairwise choices while other methods could have been used, including a ranking exercise or another choice-based task that would allow comparability with more widely used valuation methods such as TTO.

Nevertheless, this study provides important evidence in that it proposes a flexible and easy to use method for selecting bolt-on dimensions. Further research is recommended on testing other bolt-ons, levels, and potentially including duration in the comparisons to see how this compares to the previous studies that used TTO.

ORCID iD

Aureliano Paolo Finch https://orcid.org/0000-0003-3950-9210

Supplemental Material

Supplementary material for this article is available on the *Medical Decision Making* Web site at http://journals.sagepub.com/home/mdm.

References

1. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluations*. 2nd ed. Oxford, UK: Oxford University Press; 2017.
2. Brazier J, Longworth L.NICE DSU Technical Support Document 8: an introduction to the measurement and evaluation of health for NICE submissions. 2011. Available from: http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD8-Introduction-to-MVH_final.pdf
3. Richardson J, Khan M, Iezzi A, Maxwell A. Comparing and explaining differences in the magnitude, content, and sensitivity of utilities predicted by the EQ-5D, SF-6D, HUI 3, 15D, QWB, and AQoL-8D multiattribute utility instruments. *Med Decis Making*. 2014;35(3):276–91.
4. Brazier JE, Rowen D, Mavranezouli I, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess*. 2012;16(32):1–114.
5. Finch AP, Brazier J, Mukuria C. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. *Eur J Health Econ*. 2018;19(4):557–70.
6. Longworth L, Yang Y, Young T, et al. Use of generic and condition specific measures of health related quality of life in NICE decision making: systematic review, statistical modelling and survey. *Health Technology Assessment*. 2014;18(9):1–224.
7. Brazier J, Tsuchiya A. Preference-based condition specific measures of health: what happens to cross programme comparability. *Health Econ*. 2010;19(2):125–9.
8. Dowie J. Decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions. *Health Econ*. 2002;11(1):1–8.
9. Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal cardiovascular or psychosomatic disorders. *Qual Life Res*. 2008;17(3):485–95.
10. Brazier J, Rowen D, Tsuchiya A, Yang Y, Young TY. The impact of adding an extra dimension to a preference-based measure. *Soc Sci Med*. 2011;73:245–53.
11. EuroQol Group Symposium. *New Research for Future Approaches to Measuring and Valuing Health*. Berlin Germany: International Society for Pharmacoeconomics and Outcomes Research (ISPOR); 2012.
12. Gudex C. Are we lacking a dimension of energy in the EuroQol instrument? Paper presented at: 8th Plenary Meeting of the EuroQol Group; Lund, Sweden; September 1991.
13. Krabbe PFM, Stouthard MEA, Essink-Bot ML, Bonsel GJ. The effect of adding a cognitive dimension to the EuroQol multi-attribute health status classification system. *J Clin Epidemiol*. 1999;52(4):293–301.
14. Yang Y, Brazier JE, Tsuchiya A. Effect of adding a sleep dimension to the EQ5D descriptive system: a "bolt-on" experiment. *Med Decis Making*. 2014;34:42–53.
15. Yang Y, Rowen D, Brazier J, Tsuchiya A, Young T, Longworth L. An exploratory study to test the impact on three "bolt-on" items to the EQ-5D. *Value Health*. 2015;18:52–60.
16. Swinburn P, Lloyd A, Boye KS, Edson-Heredia E, Bowman L, Janssen B. Development of a disease-specific version of the EQ-5D-5L for use in patients suffering from psoriasis: lessons learned from a feasibility study in the UK. *Value Health*. 2013;16(8):1156–62.
17. Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. *Health Econ*. 1999;8:41–51.
18. Finch AP, Brazier JE, Mukuria C. An exploratory study on using principal component analysis and confirmatory factor analysis to identify bolt-on dimensions: the EQ-5D case study. *Value Health*. 2017;20(10):1362–75.
19. Finch AP, Brazier JE, Mukuria C. An exploratory study on the possibility of identifying bolt-ons using principal component analysis, confirmatory factor analysis and structural equation modeling: the EQ-5D case study (2016).

EuroQol Group, 33rd Scientific Plenary, Berlin, Germany, September 2016.

20. Finch AP, Brazier JE, Mukuria C. Selecting bolt-on dimensions for the EQ-5D: examining their contribution to health related quality of life. *Value Health*. 2019;22(1):50–61.
21. Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. *Med Decis Making*. 2015;35(3):328–40.
22. Mulhern B, Bansback N, Brazier J, et al. Preparatory study for the re-valuation of the EQ-5D tariff: methodology report. *Health Technol Assess*. 2014;18:12.
23. Devlin N, Shah K, Feng Y, Mulhern B, van Hout B. Valuing health related quality of life: a value set for England. Office of Health Economics Research Paper 16/01. 2016. Available from: file:///Users/Aureliano/Downloads/25%20Jan%20%20OHE%20research%20paper_value%20set%20paper%20CORRECTED.pdf
24. Bansback N, Hole AR, Mulhern B, Tsuchiya A. Testing a discrete choice experiment including duration to value health states for large descriptive systems: addressing design and sampling issues. *Soc Sci Med*. 2014;114:38–48.
25. Casagrande JT, Pike MT, Smith PJ. An improved approximate formula for calculating sample sizes for comparing two binomial distributions. *Biometrics*. 1978;34:3.