5-level Bolt-Ons Have an Improved Psychometric Performance Compared to 3-level Bolt-ons in a Large, Multinational, Longitudinal General Population Sample

Sarah Dewilde, PhD [1], Mathieu F. Janssen, PhD [2], Brendan Mulhern, PhD[3], Fanni Rencz, PhD[4]

Affiliations

[1] Services in Health Economics (SHE), Brussels, Belgium

[2] Section Medical Psychology and Psychotherapy, Department of Psychiatry, Erasmus MC, Rotterdam, The Netherlands

[3] Centre for Health Economics Research and Evaluation, University of Technology Sydney, Sydney, Australia

[4] Department of Health Policy, Corvinus University of Budapest, Budapest, Hungary

Counts

Words: 5629

Abstract: 500 / 500

Number of tables/ Figures: 6 + 5 in Appendix

Pages: 25

Keywords: EQ-5D-5L; General population; bolt-on; EQ-5D-3L; psychometric properties

Acknowledgements: The authors would like to thank all the participants of the study for sharing their HRQoL data

Funding_statement: The study design, data collection, data analysis were funded by argenx BV, Ghent, Belgium. The study was supported by the EuroQol Research Foundation (1601-RA).

Disclosure of Conflicts of Interest: All authors are members of the EuroQol Group. Views expressed in the article are those of the authors and are not necessarily those of the EuroQol Research Foundation.

Ethical Publication Statement: We confirm that this report is consistent with the guidelines on ethical publication. All participants signed an informed consent form. The study was performed in accordance with the Declaration of Helsinki.

Data availability statement: Anonymized, aggregated study data is available upon reasonable request through the corresponding author.

ABSTRACT (500 / 500 WORDS)

Objective: To compare the psychometric properties of the 5-level versus 3-level bolt-ons for vision, breathing, tiredness, sleep, social relationships and self-confidence when added to the EQ-5D in a general population sample.

Methods: POPUP is an international, observational, longitudinal study collecting data from representative members of the general population (in the US, Canada, UK, Italy, Spain, Germany, Netherlands, Belgium). Respondents entered demographic data, provided information on pre-existing medical conditions and filled out a set of PROMs, including the EQ-5D-5L plus 5-level bolt-ons, and the EQ-5D-3L plus 3-level bolt-ons, Hospital Anxiety and Depression Scale and HUI-3. The following psychometric properties of the bolt-ons were assessed: redistribution properties, ceiling/floor profiles, informativity, divergent and convergent validity, agreement and known group validity.

Results: 9758 respondents filled in the survey in 2020, and 4839 respondents completed the same survey again in 2023. The overall agreement between the 3-level and 5-level bolt-on responses was excellent, as evidenced by the low percentage of inconsistent response pairs, ranging from 2.8% (tiredness) to 6.7% (vision). The proportion of inconsistent responses between the 3-level and 5-level increased with the severity of problems reported (<1% for level 1 vs. >12% for levels 2-3).
There was a ceiling effect for both the EQ-5D-3L and EQ-5D-5L (45.4% and 32.3%, respectively). All 3-level and 5-level bolt-ons reduced the ceiling of the respective EQ-5D, with the largest reduction achieved with vision, sleep and breathing. This reduction was moreover larger for all 5-level bolt-ons than for their 3-level pairs, with the largest difference noted for vision (3-level 60.5%, 5-level 41.1%) and the smallest for breathing (3-level 84.0%, 5-level 78.6%). Furthermore, adding a bolt-on increased the number of unique profiles from N=151 to N=279-306 for EQ-5D-3L and from N=696 to N=1242-1362 for EQ-5D-5L.
All six 5-level bolt-ons consistently showed higher absolute and relative informativity than their 3-level pairs, with the largest improvement being achieved for social relationships and the smallest for breathing.
Both for 3-level and 5-level bolt-ons, the social relationships, self-confidence and sleep bolt-ons showed strong correlation with the anxiety/depression item, while all other correlations with the core five dimensions were weak or moderate. Both the 3-level and 5-level breathing bolt-ons showed strong correlation with relevant items of other instruments (r=0.66-0.68), however, the correlations with other instruments were only weak for vision (r=0.12-0.37).
Intra-class correlation coefficients between rescaled Level Sum Scores of the EQ-5D-5L+bolt-ons and the EQ-5D-3L+bolt-ons showed strong agreement (ICC range 0.92-0.97, all p<0.001), which was also confirmed by Bland-Altman plots.
Adding 3-level or 5-level bolt-ons to the EQ-5D instruments did not notably improve the known group validity in groups defined by the number of chronic conditions, apart from breathing. Five-level bolt-ons showed improved sensitivity to change, with more Pareto changes (improvements, worsening, mixed changes) over time than the 3-level bolt-ons.

Conclusion: This is the first study in the literature to psychometrically validate a set of 3-level and 5-level bolt-ons in parallel. All 12 bolt-ons showed good measurement performance in assessing the health status of the general population, with the 5-level bolt-ons showing superior psychometric performance.

INTRODUCTION

The EQ-5D is the most commonly used preference-accompanied health-related quality of life (HRQoL) measure worldwide[1]. It is applied in observational studies, clinical trials, patient registries and importantly, it is used to estimate quality-adjusted life years (QALYs) for cost-utility analyses of health interventions. In nearly 30 countries, national health technology assessment bodies or pharmacoeconomic guidelines recommend the EQ-5D[2,3]. The EQ-5D has a five-dimensional descriptive system covering areas of mobility, self-care, usual activities, pain/discomfort and anxiety/depression. It uses sets of societal preference weights, specific to countries, to estimate utilities. Currently, the EQ-5D has two versions for adults: the EQ-5D-3L and EQ-5D-5L. The EQ-5D-3L was the original version with three response levels in each of the five dimensions[4]. After nearly two decades of use, the descriptive system was revised to standardise its wording and improve its sensitivity to health problems. This revision included expanding from three to five response levels in each dimension and changing some item wording (e.g. revising the middle level from 'some' to 'moderate' and 'confined to bed' to 'unable to walk about')[5].

An alternative approach to improving the sensitivity of the descriptive system, besides expanding the response levels, is the inclusion of additional dimensions not captured by the original five dimensions. These additional dimensions are typically called bolt-ons[6]. Dimensions not included in the EQ-5D descriptive system but known to impact HRQoL include sensory deprivation[7], aspects of physical functioning, mental health-related domains and social roles[8]. There is a growing interest within and outside the EuroQol Group in developing bolt-ons[6]. So far, approximately 50 different bolt-ons have been proposed for the adult EQ-5D instruments to measure constructs not included in the core five dimension, with the aim to improve content validity and sensitivity to change in specific populations or contexts[9]. Existing bolt-ons for the adult EQ-5D instruments include items for specific symptoms (e.g. vision, hearing, breathing, sleeping problems, fatigue, energy, itching, bleeding), physical functioning (e.g. bending knees, speaking), psychological symptoms (e.g. stress, self-confidence), social functioning (e.g. social relationships) and items related to overall quality of life (e.g. well-being, dignity)[9,10].

For most existing bolt-ons, there are both three-level and five-level versions[11]. While numerous studies have compared the measurement performance of the three-level and five-level versions of the descriptive system[12], both three- and five-level bolt-ons are rarely used in the same study. Little is known about the added value and assumed improvement in measurement performance of the five-level bolt-ons compared to the three-level bolt-ons. We are aware of only one study comparing the measurement properties of 3-level and 5-level bolt-ons, which measured vision impacts in cataract patients[13].

Most bolt-ons are used in the context of specific patient groups, such as the vision bolt-on for patients with cataract or visual impairment[13–15], the breathing bolt-on for patients with chronic obstructive respiratory

disease[16], the itching bolt-on for psoriasis, atopic dermatitis and burn patients[17–21], and the vitality, sleep, social relationships and social isolation bolt-ons for patients with multiple sclerosis[22]. However, most of these bolt-ons capture HRQoL aspects that are also relevant to general population samples with and without common or chronic conditions. Although several studies have reported on the measurement properties of various bolt-ons in general population samples, the wording of bolt-on items typically did not follow the EQ-5D format[23–26]. Furthermore, very few studies provided evidence on the sensitivity or responsiveness of bolt-ons[20,27].

The aim of this study is to assess and compare the psychometric performance of the EQ-5D-3L with 3-level and EQ-5D-5L with 5-level bolt-ons in a large, multi-country, longitudinal general population survey. We have focused on the following six bolt-ons: vision, breathing, tiredness, sleeping, social relationships and self-confidence, which have all been proved relevant in the context of the general population in a recent extensive study involving nine EQ-5D-5L bolt-ons[28]. We hypothesize an overall improvement in measurement performance with the 5-level bolt-ons for the EQ-5D-5L, compared to the 3-level bolt-ons for the EQ-5D-3L, as the more granular response levels allow for a more precise assessment of HRQoL. More specifically, we were expecting improved response distribution (lower ceiling and floor), improved discriminatory power, higher relative efficiency in known groups, and higher sensitivity to changes with the 5-level descriptive system. However, it was difficult to hypothesize whether or not there would be substantial variations in the performance of these bolt-ons.

METHODS

DATA COLLECTION

The Population Norms Study (POPUP) was a multinational digital study that aimed to collect HRQoL data from 9000 general population participants across eight countries (US, Canada, UK, Italy, Spain, Germany, The Netherlands, and Belgium) during its first wave in Q1 2021. A second wave of data collection aimed for 4500 repeat participants and was carried out in Q1 2023. In countries where this goal was not met (Canada and US), additional participants were recruited who only completed wave 2. Data from Wave 1 and wave 2 were combined for our analyses.

POPUP's main objective was to document international populations norms, which could then aid in the evaluation of disease-associated loss in HRQoL by serving as a baseline to contrast outcomes from populations with a particular disease. The samples from each country consisted of adult participants from representative panels from the market research company Bilendi, and sampling was based on previously agreed variables (age, gender, education, and region). Through the use of stepped random sampling methods and propensity weighting, it was ensured that participants with a variety of psychological, political and social profiles were enrolled in the study. Potential participants were contacted via email to complete the survey,

for which they would receive points that they could then convert into a selection of gifts. The details of the respondent sampling, methods of data collection, description of all endpoints and findings from the first wave are described in a separate manuscript[29].

All survey participants first completed demographics, indicated their living situation and completed a question on their current health status (indicating from a list of chronic conditions which ones they suffer from, or whether they don't suffer from chronic conditions at all). This was followed by the EQ-5D-5L and 5-level bolt-ons, then a range of other (generic) PROMs (including the HUI-3[30,31] and the HADS[32]), followed by the EQ-5D-3L and 3-level bolt-ons. Finally, respondents completed questions on medical care, sick leave, and needing help from a caregiver. This order was fixed and was identical for all participants. The survey was programmed in LimeSurvey and did not allow respondents to skip questions.

EQ-5D-3L & EQ-5D-5L

The EQ-5D-3L consists of a descriptive system assessing HRQoL on 5 dimensions[4]: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension is split into 3 severity levels (no problems, some / moderate problems, confined to bed / unable to / extreme problems). The questionnaire also includes the visual analogue scale (EQ VAS), a thermometer-like vertical scale ranging from 0 (the worst health you can imagine) to 100 (the best health you can imagine) on which respondents need to rate their overall health on the day of completion. Despite widespread use, the EQ-5D-3L had difficulty capturing small to moderate differences in health status and showed ceiling effects in some populations[33,34]. As a result, the EQ-5D-5L was developed[5]. Responses to the five dimensions can be combined into a utility value, which summarizes overall HRQoL and is anchored between 1 (full health) and 0 (being dead). Health conditions perceived worse than being dead are denoted by negative values[35]. In this present study however, these utility values are not used given the psychometric nature of the work, and the absence of value sets including the bolt-ons.

Bolt-on questions

Bolt-on questions serve to enhance the content validity and responsiveness of a descriptive system in certain settings and health conditions. Several previously developed and published bolt-on dimensions have been added to the standard EQ-5D dimensions in this survey before the EQ VAS, including vision[36], breathing[37], sleep[38], tiredness[36], social relationships[21], and self-confidence[21] with the same three and five response levels as the main descriptive system[39]. The bolt-ons breathing, social relationships and self-confidence were originally developed as 5-level versions and were converted into 3-level versions using levels "no problems", "moderate problems" and "extreme problems". The bolt-ons vision and tiredness,[40] and the bolt-on sleep[38] were originally developed as 3-level bolt-ons and had the same 5-level response levels as the main descriptive systems. Refer to **Supplemental Material Table S1** for detailed wording.

STATISTICAL ANALYSIS

For each of the analyses, we examined the core dimensions of the EQ-5D and the bolt-on dimensions individually, and also in combination with each other. Whilst the first aim was to compare the relative performance of the EQ-5D-3L + 3-level bolt-on versus EQ-5D-5L + 5-level bolt-on, and the analytical framework followed that of earlier 3-level versus 5-level comparisons[41,42]. We sometimes also examined the contrast with the main EQ-5D descriptive system without the bolt-on as a secondary endpoint of interest. The combinations of several bolt-ons added to the EQ-5D, however, were not examined.

For all analyses, all available simultaneous measurements were used, that is, data from respondents who only completed the first wave, data from respondents who completed the first and second wave, and data from new respondents in the second wave. This was done in order to make use of all available data, and to maximize the number of simultaneous 3-level and 5-level assessments. The following psychometric properties of the 3-level and 5-level bolt-ons were assessed in this analysis: redistribution properties, ceiling/floor profiles, discriminatory power, divergent and convergent validity, agreement and known group validity.

Distributional characteristics

Response distribution was computed for each EQ-5D and bolt-on dimension. To compare the distributions of the 3- and 5-level descriptive systems we displayed cross-tables of responses, flagging inconsistent pairs (which were at least 2 levels apart in the 3 and 5-level descriptive systems). We summed up the number of observed inconsistent pairs and divided this by the sample size to calculate the proportion of inconsistency. As a sensitivity analysis, we also attached weights representing the number of levels away from consistency (weights were equal to 2, 3 or 4, thereby penalizing responses that were further apart between 3-level and 5-level descriptive systems) to recalculate a weighted inconsistency proportion. We also calculated the size of the inconsistency using the formula: absolute (response3L_recoded – response5L)-1, with the 3-level responses recoded from 1, 2 and 3 into 1, 3 and 5.

Furthermore, we calculated the number of different profiles and presented these as a proportion of the number of all theoretically possible profiles; and we computed the % respondents attaining the ceiling (response of "1" on all dimensions) and the floor (worst response on all dimensions) of each measure. We expected the 5-level bolt-ons to have lower ceiling and floor effects, and have a lower % of all available profiles compared to the 3-level bolt-ons. We also calculated the ceiling effect in known groups defined by chronic conditions and by needing the help from a caregiver (see more detail below in Known Groups Validity).

Informativity

The informativity of bolt-ons was determined using Shannon's indices, with the Shannon index (H') for absolute and Shannon Evenness index (J') for relative informativity[43]. These indices were calculated for individual items as well as in combination with the EQ-5D descriptive system (using profiles) to assess

whether adding a bolt-on to the descriptive system would capture more information. We expected the 5-level bolt-ons to have higher discriminatory power, but did not have any hypotheses regarding specific bolt-ons.

Divergent validity

Spearman rank order correlations were calculated to assess the strength of association between the bolt-on dimensions and the core EQ-5D dimensions. Association were considered very weak ($r < 0.20$), weak ($r = 0.20 - 0.39$), moderate ($r = 0.40 - 0.59$) and strong ($r > 0.60$)[44]. We expected associations between tiredness, sleep and social relationships on the one hand with usual activities; between breathing and pain/discomfort; between social relationships and self-confidence with anxiety/depression[28]. We also expected stronger correlations with the 5-level bolt-ons that with the 3-level bolt-ons.

Convergent validity

We examined the convergent validity for the vision and breathing bolt-ons only, by calculating Spearman correlations for breathing with the breathing problems item from the MG-ADL, and for vision with the vision item from the HUI-3 (see well enough to read ordinary newsprint, see well enough to recognize a friend on the other side of the street). We hypothesized strong correlations between these items aiming to capture similar constructs.

Agreement

The Level Sum Score (LSS) was calculated for the EQ-5D plus each bolt-on separately, and rescaled between 0-100 to ensure comparability between the 3- and 5-level versions and between the descriptive systems with and without bolt-ons, with higher scores indicating more problems. Agreement between the 3- and 5-level EQ-5D plus bolt-ons was measured by calculating the Intra-Class Correlation Coefficients (ICC) using the SAS ICC macro[45] on the rescaled LSS. ICC values < 0.5 were indicative of poor reliability, 0.50 - 0.75 indicated moderate reliability, 0.75 - 0.90 indicated good reliability, and ICC > 0.90 indicate excellent reliability[46]. Furthermore, we inspected the agreement visually by plotting Kernel Density overlay plots of the rescaled LSS and Bland-Altman plots. We did not have any hypotheses regarding the agreement of specific bolt-ons, but we did expect the 3- and 5-level EQ-5D plus bolt-ons to have at least good reliability ($r > 0.75$).

Known Groups validity

Known-group validity was examined for respondent groups defined by number of chronic conditions (none / 1-2 / 3 or more) and by needing a regular help with daily activities, transport ect. from a caregiver (yes/no). We compared the difference between subgroups in the proportion of respondents reaching the ceiling, the mean rescaled LSS, and the relative efficiency. This latter was calculated as the ratio of two F-statistics: in the numerator the F-statistic of an ANOVA model with the rescaled LSS of the EQ-5D + bolt-on as dependent variable, and the subgroup as independent variable, and in the denominator the F-statistic of the same model

but with the rescaled LSS of the EQ-5D without a bolt-on as dependent variable. Values higher than 1.00 indicated that the addition of a bolt-on improved the efficiency in detecting differences between subgroups. We expected lower ceiling and higher LSS in subgroups with more chronic conditions and/or needing a caregiver. We also expected the addition of bolt-ons to the EQ-5d to improve the relative efficiency.

Longitudinal analysis

Using the data of respondents who completed both wave 1 and wave 2, we calculated level changes for the 3- and 5-level bolt-ons. Changes in responses could take values (-4, -3, -2, -1, 0, 1, 2, 3, 4) in the 5-level descriptive system, and values (-2, -1, 0, 1, 2) in the 3-level descriptive system. We examined the mean level change, the frequency of these changes, as well as changes in the rescaled LSS. Furthermore, correlations between changes in rescaled LSS were calculated. We expected to find fewer changes in the 3-level descriptive system, and more improvements or worsening in the 5-level system, which is in line with the hypotheses that a higher number of response levels improves the sensitivity of an item.

Pareto classification of changes

Longitudinal changes in the EQ-5D core dimensions in combination with a bolt-on were classified as "improved", "worsened", "no change", and "mixed changes"[47]. A health state was considered "better" than another if it was better on at least one dimension, and no worse on any other dimension. Conversely, a health state was "worse" than another if it was worse in at least one dimension, and no better on any other dimension. Mixed changes indicated combinations of some EQ-5D core dimensions and the bolt-on worsening, some improving, and/or some remaining unchanged. The category representing "no change" therefore comprised all health states in which the response was the same on all dimensions at both time points. In order to remove respondents which were at the best possible health state at the start of the data collection (in wave 1) and who therefore could not improve, the analysis of change was repeated among respondents who were worse than 11111(1). Similarly as above, we expected the 5-level descriptive system to have more improvements, worsening and mixed changes and fewer status quo.

Probability of superiority

For each core EQ-5D dimension and bolt-on separately, we calculated the probability to detect an improvement using a non-parametric method for measuring effect sizes in longitudinal datasets[48]. This probability was computed as the number of respondents with positive changes plus half the number of ties, divided by the number of respondents with data at both time-points. This can therefore be interpreted as the probability that within a randomly sampled pair of dependent scores, the score obtained at follow-up will be smaller than the score obtained at baseline. The probability ranges from 0 to 1 and is $p < 0.5$ if more respondents deteriorate than improve, $p = 0.5$ if the same number of respondents improve and deteriorate or do not change and $p > 0.5$ if more respondents improve than deteriorate. We expected the 5-level EQ-5D dimensions and bolt-ons to have a higher probability of superiority.

Ethical approval

Ethical approval was obtained in all eight countries. The study was authorized by Veritas IRB in Canada (reference number 2021-2434-5740-1), the ethics committee at Ghent University Hospital in Belgium (reference number BC-07857) and Salus IRB reviewed and approved the other countries (reference number PN8450).

RESULTS

Study participants

A total of 9,758 members of the general population from eight different countries located in North America (US and Canada) and Europe (Belgium, Germany, Italy, the Netherlands, Spain, and the UK) filled in the survey in 2021 (wave 1), and 4,839 respondents completed the survey a second time in 2023 (wave 2), of which 3,571 who also participated in the first wave and 1,268 new respondents. In each country, the wave 1 sample matched the national general adult population in terms of age, gender, education, and region (**Supplementary Table S2**). Around 60% of the respondents from the general population suffered from chronic health conditions. Anxiety (14.7%) and depression (11.5%) were the most prevalent, followed by thyroid problems/disorder (9.7%), diabetes (9.2%), and respiratory disease (8.8%).

Available data on the EQ-5D

The combined sample of all simultaneous measurements of the EQ-5D-3L+bolt-ons and EQ-5D-5L+bolt-ons in wave 1 and 2 was N=14,597, whereas N=3,571 respondents had complete data on the 3-level and 5-level descriptive systems and bolt-ons, at both time points.

Redistribution properties of 3-level versus 5-level

The redistribution of the 3-level responses onto the 5-level responses is displayed in **Table 1**. Of all EQ-5D dimensions and bolt-ons, the social relationships bolt-on has the highest reporting of extreme problems, followed by tiredness and sleep. In contrast, mobility, self-care, usual activities and breathing were the dimensions with the highest reporting of no problems.

The consistency in responses between the 3-level and 5-level pairs was excellent, and was moreover similar for the bolt-ons and for the EQ-5D dimensions. The consistency was highest for self-care, breathing and tiredness and lowest for vision and anxiety/depression. The average proportion of inconsistent pairs was 3.5% for the EQ-5D dimensions and 4.0% for the bolt-ons (sum of red shaded cells in Table 1). When the weighted inconsistency proportion was computed, then the average proportion of inconsistencies was 7.8% for the EQ-

5D dimensions and the bolt-ons, and 8.7% for vision (**Supplemental Material Table S3**). Furthermore, it was noted that the inconsistency increased with the severity of problems reported, and that this effect was stronger in the EQ-5D dimensions than in the bolt-ons: on average 1.2% for level 1, 9.6% for level 2 and 32.4% for level 3 for the EQ-5D dimensions, and 2.6%, 4.8% and 21.3% for levels 1, 2 and 3 in the bolt-ons.

Comparison of profiles, ceiling and floor effects

Adding a bolt-on nearly doubled the number of unique profiles in each descriptive system, however it reduced the proportion of observed profiles among all possible profiles from 63% to, on average, 40.6% for the 3-level system and from 22.3% to 8.4% in the 5-level descriptive system. When comparing the performance of the 5-level versus the 3-level descriptive systems, a 3.4-fold increase in the number of different profiles was noted when using 5 response levels (**Table 2**). The lowest number of different profiles was found when adding the breathing bolt-on to the EQ-5D, whereas the highest number of profiles were observed when adding tiredness or sleep to the EQ-5D-5L or when adding tiredness of self-confidence to the EQ-5D-3L.

The ceilings of the EQ-5D-3L and EQ-5D-5L were 45.4% and 32.3%. On average, the inclusion of a bolt-on reduced the ceiling by 22% in the 5-level descriptive system and by 14% in the 3-level descriptive system, with the largest reduction achieved with vision, sleep and tiredness. When comparing the 3-level and 5-level descriptive systems with each other, it is noted that this relative reduction in ceiling was on average 35% larger for 5-level bolt-ons than for their 3-level pairs.

Comparison of Divergent and convergent validity between 3-level and 5-level bolt-ons and their EQ-5D descriptive system

Divergent validity: Both for 3-level and 5-level bolt-ons, the social relationships, self-confidence and sleep bolt-ons showed strong correlation with the anxiety/depression item, while all other correlations with the core five dimensions were weak or moderate (**Table 3**). Furthermore, tiredness and sleep were strongly correlated, and social relationships with self-confidence. These findings were similar for the 3-level and 5-level bolt-ons, but in general correlations were, on average, 5% stronger when the 5-level descriptive system was used.

Convergent validity: Both the 3-level and 5-level breathing bolt-ons showed strong correlation with the breathing problems item from the MG-ADL (r=0.66-0.68), however, the correlations of the vision bolt-on with the vision item from the HUI-3 (see well enough to read ordinary newsprint, see well enough to recognize a friend on the other side of the street), was only weak (r=0.37), which was not as hypothesized (detailed results not shown).

Table 1. Distribution of the respondents across the 5 domains of the EQ-5D-5-level

EQ-5D-3L	EQ-5D-5L						
	Mobility	1	2	3	4	5	*Total*
	1	75.2%	6.1%	0.9%	0.2%	0.2%	*12,040*
	2	1.4%	8.4%	5.0%	1.8%	0.4%	*2,470*
	3	0.2%	0.1%	0.1%	0.1%	0.1%	*86*
	Total	*11,206*	*2,118*	*867*	*308*	*96*	*N=14,597*
	Self-Care	1	2	3	4	5	*Total*
	1	90.0%	2.3%	0.4%	0.1%	0.1%	*13,547*
	2	1.1%	3.4%	1.5%	0.3%	0.1%	*950*
	3	0.1%	0.1%	0.1%	0.2%	0.2%	*101*
	Total	*13,318*	*845*	*295*	*88*	*53*	*N=14,597*
	Usual Activities	1	2	3	4	5	*Total*
	1	75.6%	6.5%	0.8%	0.1%	0.1%	*12,136*
	2	2.2%	7.5%	4.6%	1.2%	0.1%	*2,293*
	3	0.1%	0.1%	0.2%	0.4%	0.3%	*169*
	Total	*11,377*	*2,073*	*819*	*258*	*72*	*N=14,597*
	Pain/ Discomfort	1	2	3	4	5	*Total*
	1	44.3%	12.5%	0.8%	0.1%	0.1%	*8,430*
	2	2.1%	23.1%	12.4%	2.1%	0.2%	*5,811*
	3	0.1%	0.1%	0.2%	1.4%	0.6%	*355*
	Total	*6,780*	*5,210*	*1,956*	*531*	*118*	*N=14,597*
	Anxiety / Depression	1	2	3	4	5	*Total*
	1	57.7%	10.2%	1.3%	0.2%	0.1%	*10,139*
	2	2.0%	14.3%	9.0%	1.7%	0.3%	*3,982*
	3	0.2%	0.2%	0.5%	1.4%	1.1%	*476*
	Total	*8,733*	*3,588*	*1,574*	*483*	*219*	*N=14,597*

EQ-5D-3L	EQ-5D-5L						
	Vision	1	2	3	4	5	*Total*
	1	39.7%	16.1%	4.2%	0.5%	0.1%	*8,834*
	2	1.4%	19.4%	15.5%	1.8%	0.1%	*5,575*
	3	0.1%	0.2%	0.1%	0.5%	0.3%	*188*
	Total	*6,002*	*5,201*	*2,906*	*419*	*69*	*N=14,597*
	Breathing	1	2	3	4	5	*Total*
	1	77.2%	6.0%	0.6%	0.1%	0.1%	*12,261*
	2	1.3%	9.0%	4.1%	0.6%	0.1%	*2,185*
	3	0.1%	0.1%	0.1%	0.4%	0.3%	*150*
	Total	*11,472*	*2,204*	*705*	*158*	*58*	*N=14,597*
	Tiredness	1	2	3	4	5	*Total*
	1	47.6%	9.8%	1.0%	0.2%	0.1%	*8,558*
	2	0.7%	20.5%	13.4%	2.6%	0.1%	*5,448*
	3	0.1%	0.1%	0.5%	2.1%	1.3%	*591*
	Total	*7,066*	*4,433*	*2,165*	*718*	*215*	*N=14,597*
	Sleep	1	2	3	4	5	*Total*
	1	40.7%	13.9%	1.1%	0.2%	0.1%	*8,185*
	2	1.7%	22.4%	13.0%	2.6%	0.2%	*5,821*
	3	0.0%	0.2%	0.6%	1.9%	1.3%	*590*
	Total	*6,201*	*5,326*	*2,150*	*683*	*235*	*N=14,597*
	Social relationships	1	2	3	4	5	*Total*
	1	55.5%	11.0%	1.5%	0.2%	0.1%	*9,962*
	2	1.2%	13.7%	9.8%	2.6%	0.2%	*4,023*
	3	0.1%	0.1%	0.4%	1.8%	1.8%	*610*
	Total	*8,287*	*3,617*	*1,705*	*674*	*312*	*N=14,597*
	Self-confidence	1	2	3	4	5	*Total*
	1	66.6%	8.7%	1.1%	0.2%	0.1%	*11,183*
	2	1.7%	10.2%	6.6%	1.7%	0.2%	*2,970*
	3	0.2%	0.1%	0.5%	1.2%	1.1%	*445*
	Total	*9,987*	*2,772*	*1,198*	*445*	*196*	*N=14,597*

Table 2. Changes in floor and ceiling effects, and absolute (H') and relative informativity (J')

	EQ-5D-3L		EQ-5D-5L		EQ-5D-5L	EQ-5D-5L + bolt-on	EQ-5D-3L + bolt-on
EQ-5D	**%**	**N**	**%**	**N**	**vs. EQ-5D-3L**	**vs. EQ-5D-5L alone**	**vs. EQ-5D-3L alone**
% / N of possible profiles	63.0%	153	22.3%	696	3.5	ref	ref
Ceiling: 11111	45.4%	14597	32.6%	14596	-28%	ref	ref
Floor: 55555 or 33333	0.13%	14597	0.06%	14597	-54%	ref	ref
Shannon H'	3.24		4.89		51%	ref	ref
Shannon J'	0.41		0.42		3%	ref	ref
EQ-5D + vision	**%**	**N**	**%**	**N**			
% / N of possible profiles	40.1%	292	8.3%	1300	3.5	87%	91%
Ceiling: 111111	33.8%	14597	19.7%	14596	-42%	-40%	-25%
Floor: 555555 or 333333	0.11%	14597	0.05%	14597	-55%	-17%	-15%
Shannon H'	4.22		6.51		54%	33%	30%
Shannon J'	0.44		0.47		5%	11%	9%
EQ-5D + breathing	**%**	**N**	**%**	**N**			
% / N of possible profiles	38.3%	279	7.9%	1242	3.5	78%	82%
Ceiling: 111111	43.7%	14597	30.9%	14596	-29%	-5%	-4%
Floor: 555555 or 333333	0.09%	14597	0.05%	14597	-44%	-17%	-31%
Shannon H'	3.79		5.74		51%	17%	17%
Shannon J'	0.40		0.41		3%	-2%	-2%
EQ-5D + tiredness	**%**	**N**	**%**	**N**			
% / N of possible profiles	41.6%	303	8.7%	1362	3.5	96%	98%
Ceiling: 111111	36.8%	14597	24.1%	14596	-34%	-26%	-19%
Floor: 555555 or 333333	0.10%	14597	0.05%	14597	-50%	-17%	-23%
Shannon H'	4.20		6.41		53%	31%	30%
Shannon J'	0.44		0.46		4%	9%	8%
EQ-5D + sleep	**%**	**N**	**%**	**N**			
% / N of possible profiles	40.6%	296	8.6%	1343	3.5	93%	93%
Ceiling: 111111	37.7%	14597	23.6%	14596	-38%	-28%	-17%
Floor: 555555 or 333333	0.12%	14597	0.04%	14597	-67%	-33%	-8%
Shannon H'	4.11		6.34		54%	30%	27%
Shannon J'	0.43		0.45		5%	8%	6%
EQ-5D + social relationships	**%**	**N**	**%**	**N**			
% / N of possible profiles	40.9%	298	8.4%	1318	3.4	89%	95%
Ceiling: 111111	40.3%	14597	26.5%	14596	-34%	-19%	-11%
Floor: 555555 or 333333	0.12%	14597	0.05%	14597	-58%	-17%	-8%
Shannon H'	4.03		6.22		54%	27%	25%
Shannon J'	0.42		0.45		5%	6%	4%
EQ-5D + self-confidence	**%**	**N**	**%**	**N**			
% / N of possible profiles	42.0%	306	8.3%	1299	3.2	87%	100%
Ceiling: 111111	41.8%	14597	28.6%	14596	-32%	-12%	-8%
Floor: 555555 or 333333	0.12%	14597	0.05%	14597	-58%	-17%	-8%
Shannon H'	3.96		6.05		53%	24%	22%
Shannon J'	0.42		0.43		4%	3%	2%
EQ-5D + bolt-on averages	**%**		**%**				
% / N of possible profiles	40.6%		8.4%		3.4	88%	93%
Ceiling: 111111	39.0%		25.6%		-35%	-22%	-14%
Floor: 555555 or 333333	0.11%		0.05%		-55%	-19%	-15%
Absolute Informativity Shannon H	4.05		6.21		53%	27%	25%
Relative Informativity Shannon J'	0.43		0.45		5%	6%	4%

Table 3. Spearman rank order correlations between bolt-ons and EQ5D dimensions.

3 Level Bolt-ons:	EQ-5D-3L dimensions					3 Level Bolt-ons:					
	Mobility	Self-care	Usual Activities	Pain / Discomfort	Anxiety / Depression	Vision	Breathing	Tiredness	Sleep	Social relationships	Self-confidence
Vision	0.190	0.198	0.213	0.257	0.243	1.000					
Breathing	0.353	0.325	0.385	0.348	0.278	0.247	1.000				
Tiredness	0.259	0.240	0.329	0.391	0.434	0.231	0.267	1.000			
Sleep	0.314	0.291	0.404	0.458	0.513	0.249	0.328	0.520	1.000		
Social relationships	0.199	0.240	0.313	0.290	0.610	0.214	0.248	0.415	0.472	1.000	
Self-confidence	0.226	0.281	0.335	0.274	0.510	0.205	0.259	0.359	0.412	0.604	1.000
5 Level Bolt-ons:	**EQ-5D-5L dimensions**					**5 Level Bolt-ons:**					
	Mobility	Self-care	Usual Activities	Pain / Discomfort	Anxiety / Depression	Vision	Breathing	Tiredness	Sleep	Social relationships	Self-confidence
Vision	0.229	0.203	0.232	0.298	0.259	1.000					
Breathing	0.381	0.357	0.419	0.375	0.289	0.294	1.000				
Tiredness	0.289	0.273	0.368	0.433	0.471	0.246	0.302	1.000			
Sleep	0.350	0.318	0.453	0.479	0.526	0.285	0.355	0.556	1.000		
Social relationships	0.190	0.222	0.304	0.279	0.642	0.229	0.237	0.435	0.495	1.000	
Self-confidence	0.206	0.256	0.318	0.260	0.549	0.217	0.253	0.378	0.422	0.658	1.000

***** Association considered very weak (<0.20), weak (0.20-0.39), moderate (0.40-0.59) and strong (0.60<)

Agreement between 3-level and 5-level pairs

Intra-class correlation coefficients between rescaled LSSs of the EQ-5D-5L+bolt-ons and the EQ-5D-3L+bolt-ons showed strong agreement (ICC range 0.92-0.97, all p<0.001, **Table 4**), which was also confirmed by Kernel distributions and Bland-Altman plots (**Supplementary material Table S4**).

Table 4. Measure of agreement between 3-level and 5-level EQ-5D+bolt-ons

Bolt-on	ICC between rescaled LSS of the EQ-5D-5L+ bolt on and rescaled LSS of the EQ-5D-3L + bolt-on (range 0-100)	95% lower confidence bound	95% upper confidence bound
Vision	0.920	0.918	0.922
Breathing	0.963	0.962	0.964
Tiredness	0.970	0.969	0.971
Sleep	0.968	0.967	0.969
Social relationships	0.946	0.945	0.948
Self-confidence	0.956	0.955	0.957

Changes in Known Groups validity

The known groups validity of the bolt-ons in subgroups of the general population is shown in **Supplemental Material Table S5**. Across the subgroups defined by number of chronic conditions and by needing a caregiver (yes/no), and for all bolt-ons, the ceiling was reduced for respondents who have worse health (more co-morbidities, or who need a caregiver). The mean rescaled LSS also behaved as expected in all bolt-ons, with higher LSS noted in respondents with worse health status, confirming the known groups validity of the bolt-ons in these respondent groups. However, adding 3-level or 5-level bolt-ons to the EQ-5D did not notably improve the relative efficiency in comparison to the EQ-5D alone, apart from breathing for comorbidity groups (relative efficacy 1.10-1.12).

Longitudinal analysis and comparison of captured changes over time of 3-level and 5-level bolt-ons

There were 3,571 respondents who provided EQ-5D-5L and EQ-5D-3L data in Wave 1 and Wave 2. Across the board, about 72% of respondents did not experience a change in the bolt-on dimensions, whereas 14% improved and 14% worsened (**Table 5**). The bolt-ons in which most improvements as well as most worsening over time were observed were vision, tiredness and sleep, whereas considerably fewer changes were noted in the breathing bolt-on. There was a remarkable symmetry in level changes up and down, for all bolt-ons and in both descriptive systems.

The 5-level bolt-ons were more sensitive than the 3-level bolt-ons to pick up changes in response over time, as evidenced by the lower proportion of respondents reporting no change (range 8.2% to 18.1% lower for the 5-level bolt-ons), by the higher proportion of improvements (range 29.2% to 58.5% higher for 5-level that

for 3-level) and by the higher proportion of people worsening (range 51.0% to 64.6% higher for the 5-level bolt-ons compared to the 3-level bolt-ons). The 5-level descriptive system also displayed more heterogeneity in changes between the bolt-ons than when using the 3-level descriptive system .

There was no pattern to be detected in the mean changes in rescaled LSS, which was sometimes higher and sometimes lower in the 5-level bolt-ons. The correlations between a change in the rescaled LSS for the 3-level bolt-ons and the rescaled LSS for the 5-level bolt-ons were moderate (between 0.519 and 0.588).

Pareto Class changes with the 3-level and 5-level bolt-ons

In **Table 6** the Pareto class changes are displayed for the EQ-5D in combination with a bolt-on (as opposed to **Table 5** which looked at the bolt-ons individually). Comparing the 3-level with 5-level descriptive systems, we found that the 5-level descriptive system was more sensitive to change. This was evidenced by an average reduction of 14.3% fewer status quo (range 8.2% to 18.1%), balanced out by an average 47.1% more mixed changes, 2% more improvements and 3% more worsenings than when measured with the 3-level descriptive system. When the analysis was repeated excluding respondents scoring 11111(1), similar observations were made showing mostly more mixed changes at the expense of reductions in the status quo. An exception to this latter observation was the EQ-5D + vision, where the proportion of respondents with mixed changes or no changes was almost the same between the 3-level and 5-level descriptive systems.

Probability of superiority contrasted between 3-level and 5-level bolt-ons

Based on the Pareto changes in Table 6, the probability that a health state in wave 2 was found to be superior (improved health) compared to in wave 1 was calculated. We found that this probability was on average 0.501 for the 3-level and 0.498 for the 5-level systems. Furthermore, we did not find this probability to vary much across dimensions and bolt-ons: the minimum was 0.476 for the sleep bolt-on and 0.513 for the anxiety/depression dimension (detailed results not shown).

Table 5. Comparison of longitudinal changes between 3-level and 5-level bolt-ons

Change Wave 2 - Wave 1	Vision		Breathing		Tiredness		Sleep		Social Relationships		Self-confidence	
	3-Level	5-Level	3-Level	5-Level	3-Level	5-Level	3-Level	5-Level	3-Level	5-Level	3-Level	5-Level
-4		0.1%		0.0%		0.1%		0.1%		0.1%		0.1%
-3		0.4%		0.1%		0.3%		0.2%		0.4%		0.5%
-2	0.1%	2.5%	0.1%	1.0%	0.2%	2.7%	0.3%	2.1%	0.4%	2.2%	0.4%	2.0%
-1	13.4%	17.4%	6.3%	8.9%	13.0%	16.6%	12.4%	15.4%	10.3%	14.1%	10.1%	10.9%
0	74.4%	61.4%	87.1%	80.0%	75.7%	62.1%	72.7%	59.6%	79.6%	68.5%	80.8%	72.7%
1	11.9%	15.7%	6.4%	8.7%	10.8%	15.9%	14.4%	19.4%	9.3%	12.4%	8.3%	11.1%
2	0.2%	2.3%	0.1%	1.3%	0.3%	2.0%	0.3%	2.8%	0.4%	1.9%	0.5%	2.3%
3		0.2%		0.0%		0.4%		0.3%		0.4%		0.4%
4		0.0%		0.0%		0.1%		0.1%		0.0%		0.1%
Mean change in rescaled LSS	0.10	0.04	0.22	0.17	0.05	0.07	0.39	0.40	0.14	0.05	0.09	0.18
Correlation (Δ 3L rescaled LSS ; Δ 5L rescaled LSS)	0.519		0.565		0.586		0.584		0.588		0.583	
5-level versus 3-level: change in % no change	-17.4%		-8.2%		-18.0%		-18.1%		-14.0%		-10.1%	
5-level versus 3-level: change in % improvement	50.3%		55.1%		49.1%		40.7%		58.5%		29.2%	
5-level versus 3-level: change in % worsening	51.0%		55.5%		64.6%		54.5%		50.9%		57.6%	

Table 6. Comparison of Pareto class changes between the EQ-5D-3L+3-level and EQ-5D-5L+5-level bolt-ons

For all respondents	EQ-5D		EQ-5D + Vision		EQ-5D + Breathing		EQ-5D + Tiredness		EQ-5D + Sleep		EQ-5D + Social Relationship		EQ-5D + Self-Confidence	
	3L	5L	3L	5L	3L	5L	3L	5L	3L	5L	3L	5L	3L	5L
Improved	22%	26%	26%	28%	24%	26%	26%	28%	26%	27%	24%	27%	24%	26%
Mixed	5%	12%	10%	21%	7%	16%	9%	19%	8%	19%	8%	19%	8%	18%
No change	49%	35%	38%	24%	46%	32%	41%	26%	39%	25%	43%	25%	43%	30%
Worsened	23%	27%	26%	27%	24%	26%	25%	27%	27%	29%	25%	29%	24%	27%
Difference Improved	4%		2%		2%		2%		1%		3%		2%	
Difference Mixed	7%		11%		9%		11%		11%		11%		9%	
Difference No Change	-14%		-14%		-14%		-15%		-14%		-18%		-14%	
Difference Worsened	3%		1%		2%		2%		2%		4%		2%	
For all respondents who are worse than 11111(1)	**EQ-5D**		**EQ-5D + Vision**		**EQ-5D + Breathing**		**EQ-5D + Tiredness**		**EQ-5D + Sleep**		**EQ-5D + Social Relationship**		**EQ-5D + Self-Confidence**	
	3L	5L	3L	5L	3L	5L	3L	5L	3L	5L	3L	5L	3L	5L
Improved	40%	38%	38%	31%	41%	37%	40%	36%	41%	35%	41%	36%	42%	37%
Mixed	9%	18%	15%	15%	12%	23%	14%	25%	13%	25%	13%	25%	14%	25%
No change	30%	19%	26%	28%	25%	16%	24%	13%	22%	13%	23%	14%	23%	15%
Worsened	21%	24%	22%	26%	22%	24%	22%	25%	25%	26%	23%	24%	22%	24%
Difference Improved	-2%		-7%		-4%		-4%		-5%		-4%		-5%	
Difference Mixed	9%		0%		11%		12%		12%		12%		11%	
Difference No Change	-10%		2%		-9%		-11%		-9%		-9%		-8%	
Difference Worsened	4%		4%		2%		3%		2%		1%		2%	

DISCUSSION

This study is the first in the literature to systematically compare the measurement performance of the EQ-5D-3L with three-level bolt-ons and the EQ-5D-5L with five-level bolt-ons in any population. It extends the existing literature, which includes a wide range of studies in both general and patient populations[12], comparing the performance of the EQ-5D-3L and EQ-5D-5L over the past 15 years.

Comparison with the results from the 3L-5L core EQ-5D analysis

Consistent with previous studies[41], the EQ-5D-5L with five-level bolt-ons generally demonstrated superior performance to the EQ-5D-3L with three-level bolt-ons across most measurement properties, including reduction in ceiling, improved informativity and responsiveness. However, some variation was observed across the six bolt-ons. Overall, there was good agreement between the three-level and five-level bolt-ons, with a low inconsistency rate and excellent agreement in terms of ICC values for level sum scores. Vision had the highest overall inconsistency rate, whereas breathing had the lowest. The five-level bolt-ons were able to capture substantially more information compared to the three-level ones, as indicated by the higher number of observed health state profiles and improvements in absolute and relative informativity. The improvement in the number of profiles and informativity did not differ prominently across the six bolt-on areas. Both three-level and five-level bolt-ons reduced the ceiling effect of their respective instruments, with vision, sleep, and tiredness showing the most significant improvements. The five-level versions, particularly for vision and tiredness, exhibited the largest reduction in ceiling effect compared to their three-level counterparts, while breathing showed the smallest difference. Regarding divergent validity, the five-level bolt-ons showed stronger correlations with the core dimensions in most cases. This is in line with previous experiences with correlations between EQ-5D-3L and EQ-5D-5L dimensions and other instruments, where the latter typically shows stronger correlations due to its more precise measurement, given the higher number of response levels[49–55]. Regarding known-groups validity, although both the EQ-5D-3L and EQ-5D-5L plus bolt-ons were able to differentiate across groups in terms of mean LSS, the difference in relative efficiency was negligible both between the EQ-5D and EQ-5D+bolt-on, and between the 3L and 5L versions across the six bolt-on areas. Responsiveness was also improved with the five-level bolt-ons at both the item level and the instrument level (EQ-5D+bolt-on). At the item level, particularly for vision, sleep and tiredness, there was significant improvement with the five-level bolt-on. At the instrument level, all bolt-ons except for vision demonstrated similar improvements. Interestingly, for vision, there was minimal change between the EQ-5D-3L+bolt-on and the EQ-5D-5L+bolt-on in this respect.

The five-level versions of the six bolt-ons tested in this study have recently been included in a large general population survey in Hungary[28]. A key common finding is the strong psychometric performance of the vision, tiredness and sleep bolt-ons in general population samples. In this study, sleep, social relationships and tiredness all demonstrated a strong correlation with the core dimension of anxiety/depression. However, in

the Hungarian study, this correlation was only observed for self-confidence. These findings suggest that experiencing anxiety and depression may be associated with problems sleeping, difficulties in social relationships and low self-confidence, potentially diminishing the value of these bolt-ons. However, qualitative evidence indicates that respondents appreciate the self-confidence item, find it highly relevant, and do not consider it to overlap with anxiety/depression as a construct[18,19]. Further qualitative research would be beneficial to explore the added value of sleep and social relationships as bolt-ons.

Strengths and limitations

This study collected data from a large representative sample from the general population, in two waves and with no missing data, allowing us to perform a wide range of analyses, including longitudinal analysis. The multinational aspect of this study implied that the 3-level and the 5-level bolt-ons were tested across 8 countries. This study is among the largest samples so far collecting data on 3-level bolt-ons[56,57], and the largest study with 5-level bolt-ons. The longitudinal nature of the data also allowed us to assess sensitivity to changes in HRQoL. Previous work assessed the responsiveness of the vision and self-confidence bolt-ons in various patient samples[13,15,20]; however, to our knowledge this is the first study to report on the responsiveness of the breathing, tiredness, sleep and social relationships bolt-ons.

This study has a few limitations to acknowledge. Firstly, the positioning of the two descriptive systems and the bolt-ons within the survey was not randomised, therefore the presence of some ordering effects cannot be ruled out. Secondly, follow-up data was not available for all respondents, potentially introducing bias into the follow-up results. Furthermore, some respondents contributed more than one set of data in some of the analyses, and if the follow-up sample differs in a consistent way from the baseline sample then this could introduce bias in the results. In all European countries, the majority of the follow-up respondents were a subset of the respondents from the first wave, who were representative for the general population. For the US and Canada, the second wave of data collection was not successful in retaining respondents who also participated in the first wave (many people had left the panels two years later, and some did not want to participate again), therefore the majority of respondents from these two countries were new (but also recruited in a manner that was representative for the general population). Thirdly, data collection occurred during the first wave of the COVID-19 pandemic, which may have influenced responses and, consequently, the measurement performance of the instruments and bolt-ons. Since these results were obtained from a general population sample, they cannot be generalised to specific patient populations. Fourthly, we were unable to assess test-retest reliability due to limitations in the dataset, which also lacks substantial evidence regarding any bolt-ons[27]. Addressing this gap should be a priority for future research projects. Fifthly, there were limited variables in the dataset suitable for known-groups validity analyses, and our findings in this respect are rather explorative. Future research is needed with clinically defined known groups to test these bolt-ons. Lastly, we examined the psychometric performance of the EQ-5D and its respective bolt-ons, all using an identical number of response levels. However, it presents an interesting research direction to investigate whether the

use of five-level versions is warranted across all HRQoL domains, or if, for specific constructs and contexts, three or four-level bolt-ons might be more suitable for the EQ-5D-5L.

Future research should focus on the psychometric performance of the bolt-ons in different countries and/or different languages. Furthermore, different patterns in the performance of the bolt-ons and variations between the three-level and five-level bolt-ons in patient populations could also be the focus of future research endeavours. *As a first step towards this latter research objective, we are presenting the performance of the 5-level bolt-ons in a rare disease population (myasthenia gravis) in the poster session of this conference.*

CONCLUSION

This is the first study in the literature to psychometrically validate a set of 3-level and 5-level bolt-ons in parallel. All 12 bolt-ons showed good measurement performance in assessing the HRQoL of the general population, with the 5-level bolt-ons showing superior psychometric performance. This finding is in line with current trends in research, given the more prominent role that the 5-level EQ-5D descriptive system is now playing, at the expense of a decline in use of the 3-level descriptive system. Furthermore, this study provides useful information in understanding where and how bolt ons could be used, and helps informing the development of the bolt-on framework, which is a key part of the new Euroqol strategic plan. Our findings are useful for informing future bolt-on development, selection and testing, as well as instrument development beyond the EQ-5D, as they provide information about six additional constructs alongside the EQ-5D.

REFERENCES

1. Richardson J, Khan MA, 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*. 2015;35(3):276-291. doi:10.1177/0272989X14543107
2. Rencz F, Gulácsi L, Drummond M, et al. EQ-5D in Central and Eastern Europe: 2000-2015. *Qual Life Res*. 2016;25(11):2693-2710. doi:10.1007/S11136-016-1375-6
3. Kennedy-Martin M, Slaap B, Herdman M, et al. Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *Eur J Health Econ*. 2020;21(8):1245-1257. doi:10.1007/S10198-020-01195-8
4. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199-208. doi:10.1016/0168-8510(90)90421-9
5. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727-1736. doi:10.1007/S11136-011-9903-X
6. Mulhern BJ, Sampson C, Haywood P, et al. Criteria for developing, assessing and selecting candidate EQ-5D bolt-ons. *Qual Life Res*. 2022;31(10). doi:10.1007/S11136-022-03138-7
7. Shah KK, Mulhern B, Longworth L, Janssen MF. Views of the UK General Public on Important Aspects of Health Not Captured by EQ-5D. *Patient*. 2017;10(6):701-709. doi:10.1007/S40271-017-0240-1

8. Efthymiadou O, Mossman J, Kanavos P. Health related quality of life aspects not captured by EQ-5D-5L: Results from an international survey of patients. *Health Policy*. 2019;123(2):159-165. doi:10.1016/J.HEALTHPOL.2018.12.003

9. Geraerds AJLM, Bonsel GJ, Janssen MF, Finch AP, Polinder S, Haagsma JA. Methods Used to Identify, Test, and Assess Impact on Preferences of Bolt-Ons: A Systematic Review. *Value Health*. 2021;24(6):901-916. doi:10.1016/J.JVAL.2020.12.011

10. Finch AP, Mulhern B. Where do measures of health, social care and wellbeing fit within a wider measurement framework? Implications for the measurement of quality of life and the identification of bolt-ons. *Soc Sci Med*. 2022;313. doi:10.1016/J.SOCSCIMED.2022.115370

11. Rencz F, Pangestu S, Mulhern B, Paolo Finch A, Janssen MF. *PROSPERO International Prospective Register of Systematic Reviews Cognition Bolt-Ons for the EQ-5D-5L and EQ-5D-3L: A Systematic Review Citation Cognition Bolt-Ons for the EQ-5D-5L and EQ-5D-3L: A Systematic Review Review Question*.

12. Buchholz I, Janssen MF, Kohlmann T, Feng YS. A Systematic Review of Studies Comparing the Measurement Properties of the Three-Level and Five-Level Versions of the EQ-5D. *Pharmacoeconomics*. 2018;36(6):645-661. doi:10.1007/S40273-018-0642-5/TABLES/7

13. Gandhi M, Ang M, Teo K, et al. A vision "bolt-on" increases the responsiveness of EQ-5D: preliminary evidence from a study of cataract surgery. *Eur J Health Econ*. 2020;21(4):501-511. doi:10.1007/S10198-019-01156-W

14. Luo N, Wang X, Ang M, et al. A Vision "Bolt-On" Item Could Increase the Discriminatory Power of the EQ-5D Index Score. *Value Health*. 2015;18(8):1037-1042. doi:10.1016/J.JVAL.2015.08.002

15. Breheny K, Hollingworth W, Kandiyali R, et al. Assessing the construct validity and responsiveness of Preference-Based Measures (PBMs) in cataract surgery patients. *Qual Life Res*. 2020;29(7):1935-1946. doi:10.1007/S11136-020-02443-3

16. Hoogendoorn M, Jowett S, Dickens AP, et al. Performance of the EQ-5D-5L Plus Respiratory Bolt-On in the Birmingham Chronic Obstructive Pulmonary Disease Cohort Study. *Value Health*. 2021;24(11):1667-1675. doi:10.1016/J.JVAL.2021.05.006

17. Dijkshoorn JN, Haagsma JA, Vlies CH van der, Hop MJ, Baar ME van, Spronk I. Assessing Health-Related Quality of Life of Adult Patients with Intermediate Burns: The Added Value of an Itching and Cognition Item for the EQ-5D: A Retrospective Observational Study. *European Burn Journal 2022, Vol 3, Pages 264-277*. 2022;3(2):264-277. doi:10.3390/EBJ3020023

18. Szlávicz E, Szabó Á, Kinyó Á, et al. Content validity of the EQ-5D-5L with skin irritation and self-confidence bolt-ons in patients with atopic dermatitis: a qualitative think-aloud study. *Qual Life Res*. 2024;33(1):101-111. doi:10.1007/S11136-023-03519-6

19. Rencz F, Mukuria C, Bató A, Poór AK, Finch AP. A qualitative investigation of the relevance of skin irritation and self-confidence bolt-ons and their conceptual overlap with the EQ-5D in patients with psoriasis. *Qual Life Res*. 2022;31(10):3049-3060. doi:10.1007/S11136-022-03141-Y

20. Pickard AS, Gooderham M, Hartz S, Nicolay C. EQ-5D health utilities: exploring ways to improve upon responsiveness in psoriasis. *J Med Econ*. 2017;20(1):19-27. doi:10.1080/13696998.2016.1219359

21. 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):1156-1162. doi:10.1016/J.JVAL.2013.10.003

22. Campbell JA, Ahmad H, Chen G, et al. Validation of the EQ-5D-5L and psychosocial bolt-ons in a large cohort of people living with multiple sclerosis in Australia. *Qual Life Res*. 2023;32(2):553-568. doi:10.1007/S11136-022-03214-Y

23. Kim SH, Jo MW, Ock M, Lee S Il. Exploratory Study of Dimensions of Health-related Quality of Life in the General Population of South Korea. *J Prev Med Public Health*. 2017;50(6):361-368. doi:10.3961/JPMPH.16.076

24. Perneger T V., Courvoisier DS. Exploration of health dimensions to be included in multi-attribute health-utility assessment. *Int J Qual Health Care*. 2011;23(1):52-59. doi:10.1093/INTQHC/MZQ068

25. Chen G, Olsen JA. Filling the psycho-social gap in the EQ-5D: the empirical support for four bolt-on dimensions. *Qual Life Res*. 2020;29(11):3119-3129. doi:10.1007/S11136-020-02576-5

26. Jelsma J, Maart S. Should additional domains be added to the EQ-5D health-related quality of life instrument for community-based studies? An analytical descriptive study. *Popul Health Metr*. 2015;13(1). doi:10.1186/S12963-015-0046-0

27. Igarashi A, Sakata Y, Azuma-Kasai M, et al. Linguistic and Psychometric Validation of the Cognition Bolt-On Version of the Japanese EQ-5D-5L for the Elderly. *J Alzheimers Dis*. 2023;91(4):1447-1458. doi:10.3233/JAD-221080

28. Rencz F, Janssen MF. Testing the Psychometric Properties of 9 Bolt-Ons for the EQ-5D-5L in a General Population Sample. *Value Health*. Published online 2024. doi:10.1016/J.JVAL.2024.03.2195

29. Dewilde S, Tollenaar NH, Phillips G, Paci S, Janssen MF. POPUP: an observational digital study reporting general population norms for the EQ-5D-5L and HUI-3 in 8 countries. *Submitted to Archives of Public Health - under review*.

30. Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems. Health Utilities Index. *Pharmacoeconomics*. 1995;7(6):490-502. doi:10.2165/00019053-199507060-00004

31. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health Qual Life Outcomes*. 2003;1. doi:10.1186/1477-7525-1-54

32. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983;67(6):361-370. doi:10.1111/J.1600-0447.1983.TB09716.X

33. Cheng LJ, Pan T, Chen LA, et al. The Ceiling Effects of EQ-5D-3L and 5L in General Population Health Surveys: A Systematic Review and Meta-analysis. *Value in Health*. 2024;0(0). doi:10.1016/j.jval.2024.02.018

34. Longworth L, Singh J, Brazier J. An Evaluation of the Performance of Eq-5d: A Review of Reviews of Psychometric Properties. *Value Health*. 2014;17(7):A570. doi:10.1016/J.JVAL.2014.08.1907

35. Brazier J, Deverill M, Green C. A review of the use of health status measures in economic evaluation. *J Health Serv Res Policy*. 1999;4(3):174-184. doi:10.1177/135581969900400310

36. 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(1):52-60. doi:10.1016/J.JVAL.2014.09.004

37. Hoogendoorn M, Oppe M, Boland MRS, Goossens LMA, Stolk EA, Rutten–van Mölken MPMH. Exploring the Impact of Adding a Respiratory Dimension to the EQ-5D-5L. *Med Decis Making*. 2019;39(4):393-404. doi:10.1177/0272989X19847983

38. Yang Y, Brazier J, Tsuchiya A. Effect of adding a sleep dimension to the EQ-5D descriptive system: a "bolt-on" experiment. *Med Decis Making*. 2014;34(1):42-53. doi:10.1177/0272989X13480428

39. 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. doi:10.1016/J.JVAL.2018.07.001

40. 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: a systematic review, statistical modelling and survey. *Health Technol Assess*. 2014;18(9):1-224. doi:10.3310/HTA18090

41. Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L Better Than EQ-5D-3L? A Head-to-Head Comparison of Descriptive Systems and Value Sets from Seven Countries. *Pharmacoeconomics*. 2018;36(6):675-697. doi:10.1007/S40273-018-0623-8

42. Janssen MF, Buchholz I, Golicki D, Bonsel GJ. Is EQ-5D-5L Better Than EQ-5D-3L Over Time? A Head-to-Head Comparison of Responsiveness of Descriptive Systems and Value Sets from Nine Countries. *Pharmacoeconomics*. 2022;40(11):1081. doi:10.1007/S40273-022-01172-4

43. Shannon CE, Weaver W. *The Mathematical Theory of Communication.* University of Illinois Press; 1949.

44. Evans JD. *Straightforward Statistics for the Behavioral Sciences*. Thomson Brooks/Cole Publishing Co; 1996.

45. Lu L, Shara N. Reliability Analysis: Calculate and Compare Intraclass Correlation Coefficients (ICC) in SAS®. Published online 2007.

46. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. Prentice Hall; 2000.

47. Devlin NJ, Parkin D, Browne J. Patient-reported outcome measures in the NHS: new methods for analysing and reporting EQ-5D data. *Health Econ*. 2010;19(8):886-905. doi:10.1002/HEC.1608

48. Buchholz I, Thielker K, Feng YS, Kupatz P, Kohlmann T. Measuring changes in health over time using the EQ-5D 3L and 5L: a head-to-head comparison of measurement properties and sensitivity to change in a German inpatient rehabilitation sample. *Qual Life Res*. 2015;24(4):829-835. doi:10.1007/S11136-014-0838-X

49. Koszorú K, Hajdu K, Brodszky V, et al. Comparing the psychometric properties of the EQ-5D-3L and EQ-5D-5L descriptive systems and utilities in atopic dermatitis. *Eur J Health Econ*. 2023;24(1):139-152. doi:10.1007/S10198-022-01460-Y

50. Golicki D, Niewada M, Buczek J, et al. Validity of EQ-5D-5L in stroke. *Qual Life Res*. 2015;24(4):845-850. doi:10.1007/S11136-014-0834-1

51. Michalowsky B, Hoffmann W, Mohr W, Rädke A, Xie F. Comparing the psychometric properties of EQ-5D-3L and EQ-5D-5L proxy ratings by informal caregivers and a health professional for people with dementia. *Health Qual Life Outcomes*. 2022;20(1). doi:10.1186/S12955-022-02049-Y

52. Jin X, Al Sayah F, Ohinmaa A, Marshall DA, Smith C, Johnson JA. The EQ-5D-5L Is Superior to the -3L Version in Measuring Health-related Quality of Life in Patients Awaiting THA or TKA. *Clin Orthop Relat Res*. 2019;477(7):1632-1644. doi:10.1097/CORR.0000000000000662

53. Bató A, Brodszky V, Gergely LH, et al. The measurement performance of the EQ-5D-5L versus EQ-5D-3L in patients with hidradenitis suppurativa. *Qual Life Res*. 2021;30(5):1477-1490. doi:10.1007/S11136-020-02732-X

54. Rencz F, Lakatos PL, Gulácsi L, et al. Validity of the EQ-5D-5L and EQ-5D-3L in patients with Crohn's disease. *Qual Life Res*. 2019;28(1):141-152. doi:10.1007/S11136-018-2003-4

55. Poór AK, Rencz F, Brodszky V, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L in psoriasis patients. *Qual Life Res*. 2017;26(12):3409-3419. doi:10.1007/S11136-017-1699-X

56. Lutomski JE, Hoekstra T, Akker M van den, et al. Multimorbidity patterns in older persons and their association with self-reported quality of life and limitations in activities of daily living. *Arch Gerontol Geriatr*. 2023;115. doi:10.1016/J.ARCHGER.2023.105134

57. Franse CB, Van Grieken A, Qin L, Melis RJF, Rietjens JAC, Raat H. Ethnic differences in frailty: a cross-sectional study of pooled data from community-dwelling older persons in the Netherlands. *BMJ Open*. 2018;8(8). doi:10.1136/BMJOPEN-2018-022241

SUPPLEMENTARY MATERIAL

Table S1. Descriptive system of the 3-level and 5-level bolt-ons

Descriptive system	
Bolt-ons to EQ-5D-5L	**Bolt-ons to EQ-5D-3L**
Vision (e.g. using glasses or contact lenses if needed)	**Vision (e.g. using glasses or contact lenses if needed)**
I have no problems seeing	I have no problems seeing
I have slight problems seeing	I have some problems seeing
I have moderate problems seeing	I have extreme problems seeing
I have severe problems seeing	
I have extreme problems seeing	
Breathing problems (e.g. shortness of breath, wheezing, coughing, sputum,…)	**Breathing problems (e.g. shortness of breath, wheezing, coughing, sputum,…)**
I have no problems breathing	I have no breathing problems
I have slight problems breathing	I have moderate breathing problems
I have moderate problems breathing	I have extreme breathing problems
I have severe problems breathing	
I have extreme problems breathing	
Tiredness	**Tiredness**
I am not tired	I am not tired
I am slightly tired	I am moderately tired
I am moderately tired	I am extremely tired
I am severely tired	
I am extremely tired	
Sleep	**Sleep**
I have no problems sleeping	I have no problems sleeping
I have slight problems sleeping	I have moderate problems sleeping
I have moderate problems sleeping	I have extreme problems sleeping
I have severe problems sleeping	
I have extreme problems sleeping	
Self-confidence	**Self-confidence**
I have no problems with self-confidence	I have no problems with self-confidence
I have slight problems with self-confidence	I have moderate problems with self-confidence
I have moderate problems with self-confidence	I have extreme problems with self-confidence
I have severe problems with self-confidence	
I have extreme problems with self-confidence	
Social relationships	**Social relationships**
I have no problems with social relationships	I have no problems with social relationships
I have slight problems with social relationships	I have moderate problems with social relationships
I have moderate problems with social relationships	I have extreme problems with social relationships
I have severe problems with social relationships	
I have extreme problems with social relationships	

Table S2. Description of respondent characteristics in wave 1 and 2 of the POPUP study

Proportion of respondents in Wave 1	Belgium	Canada	Germany	Italy	Netherlands	Spain	UK	US
	N=1000	N=1000	N=1000	N=1000	N=1000	N=1000	N=1000	N=2000
Age								
18-34	28.3%	29.4%	24.1%	22.9%	26.3%	27.0%	28.9%	30.6%
35-54	35.7%	39.1%	36.2%	37.0%	37.4%	38.4%	35.3%	36.7%
55 +	36.0%	31.6%	39.7%	40.2%	36.4%	34.6%	35.8%	32.7%
Gender								
Female	50.4%	50.5%	51.7%	52.3%	51.0%	51.2%	51.5%	51.4%
Male	49.7%	49.6%	48.3%	47.7%	49.1%	48.9%	48.6%	48.6%
Education								
Primary education	3.4%	2.0%	6.4%	6.7%	1.7%	11.2%	1.8%	4.4%
Secondary education	56.5%	43.7%	67.4%	42.9%	60.0%	61.9%	62.4%	55.6%
Higher education	40.1%	54.3%	26.3%	50.4%	38.3%	27.0%	35.8%	39.9%
EQ VAS								
Mean	76.2	73.9	76.8	76.1	76.9	76.9	72.3	79
Comorbidities								
No comorbidities	41.1%	38.1%	37.5%	37.9%	48.7%	42.0%	44.5%	39.6%
Types of comorbidities								
Anxiety	12.8%	24.3%	9.2%	14.9%	5.9%	17.1%	20.6%	12.4%
Depression	9.1%	15.8%	14.4%	6.7%	7.9%	10.4%	17.0%	10.7%
Thyroid problems / Thyroid disorder	9.3%	9.1%	18.7%	12.3%	5.2%	8.8%	5.8%	8.3%
Diabetes with / without complications	6.7%	9.6%	9.0%	6.1%	9.0%	5.9%	7.8%	19.3%
Respiratory Disease (COPD, asthma,..)	8.5%	7.9%	9.7%	5.0%	12.4%	6.7%	12.1%	7.9%
Rheumatoid arthritis/Psoriatic Arthritis	10.7%	7.6%	12.7%	8.0%	8.0%	7.9%	8.2%	6.5%
Gastro-intestinal problems	9.7%	5.9%	3.4%	10.2%	6.3%	5.6%	8.1%	5.7%
CVD	5.4%	2.8%	6.7%	4.1%	7.5%	4.4%	1.8%	4.0%
Osteoporosis	3.6%	2.9%	1.6%	6.7%	2.3%	2.0%	2.7%	3.4%
Cancer	2.8%	1.9%	3.1%	1.1%	2.2%	1.3%	1.1%	2.2%

Proportion of respondents in Wave 2	Belgium	Canada	Germany	Italy	Netherlands	Spain	UK	US
	N=500	N=500	N=500	N=500	N=500	N=500	N=500	N=1000
Age								
18-34	28.3%	29.4%	24.1%	22.9%	26.4%	27.1%	28.9%	30.6%
35-54	35.8%	39.1%	36.1%	37.0%	37.6%	38.4%	35.3%	36.7%
55 +	35.9%	31.6%	39.8%	40.2%	36.0%	34.6%	35.8%	32.7%
Gender								
Female	50.7%	50.5%	51.7%	52.3%	51.3%	51.2%	52.5%	51.4%
Male	49.3%	49.5%	48.3%	47.7%	48.7%	48.9%	47.5%	48.6%
Education								
Primary education	4.1%	2.7%	5.4%	7.1%	1.9%	11.5%	0.2%	3.4%
Secondary education	55.7%	44.0%	64.7%	42.0%	57.3%	57.6%	62.5%	56.8%
Higher education	40.2%	53.4%	29.9%	51.0%	40.9%	31.0%	37.3%	39.8%
EQ VAS								
Mean	73.7	74.1	73.8	73.8	75.0	72.5	71.4	76.3
Comorbidities								
No comorbidities	40.9%	40.4%	37.4%	38.0%	50.3%	42.7%	47.0%	38.6%
Types of comorbidities								
Anxiety	14.8%	20.2%	10.0%	18.5%	8.2%	20.6%	23.1%	13.2%
Depression	8.9%	14.1%	16.5%	8.4%	7.1%	12.3%	21.6%	9.5%
Thyroid problems / Thyroid disorder	9.6%	8.9%	16.1%	10.5%	3.5%	8.5%	6.9%	7.8%
Diabetes with / without complications	7.2%	11.5%	10.4%	5.9%	8.8%	5.8%	5.5%	19.0%
Respiratory Disease (COPD, asthma,..)	7.3%	8.3%	9.2%	4.0%	9.0%	7.9%	10.4%	9.1%
Rheumatoid arthritis/Psoriatic Arthritis	4.3%	5.1%	5.0%	2.9%	5.0%	4.2%	4.9%	4.0%
Gastro-intestinal problems	11.3%	7.9%	2.5%	8.6%	4.5%	5.7%	7.5%	3.7%
CVD	4.5%	2.7%	7.1%	4.0%	7.5%	2.8%	1.3%	3.1%
Osteoporosis	3.3%	5.9%	2.3%	8.4%	2.6%	2.8%	3.2%	4.6%
Cancer	1.6%	1.8%	1.5%	1.2%	1.7%	1.4%	1.3%	2.8%

Table S3. Proportion of inconsistent 3-level and 5-level pairs for the EQ-5D descriptive system and the bolt-ons

	% Inconsistent pairs	% Inconsistent pairs, weighted by number of level separations	Size of the inconsistency
Mobility	3.4%	7.7%	-0.73
Self-Care	2.2%	5.0%	-0.90
Usual Activities	3.8%	8.3%	-0.74
Pain / Discomfort	3.6%	7.8%	-0.35
Anxiety / Depression	4.7%	10.2%	-0.52
Vision	6.7%	14.5%	-0.48
Breathing	2.5%	5.5%	-0.78
Tiredness	2.8%	6.2%	-0.50
Sleep	4.2%	9.0%	-0.59
Social relationships	3.9%	8.4%	-0.62
Self-confidence	4.0%	8.8%	-0.70

* Weights: 2 = two level apart from an accepted mapping level; 3 = 3 levels apart; 4 = 4 levels apart. E.g. Response level 2 in 3L and response level 1 in 5L are 2 levels apart; Response level 3 in 3L and response level 1 in 5L are 4 levels apart

[$] Inconsistency calculated as follows: abs (Response3L_recoded – Response5L) – 1; with the 3-level responses recoded into 1, 3 and 5 instead of 1, 2 and 3.

Table S4. Visual depiction of agreement between 3-level and 5-level pairs

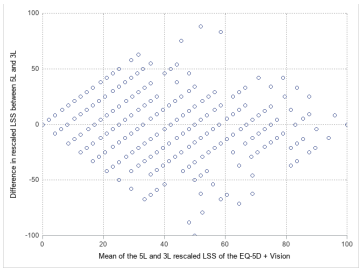
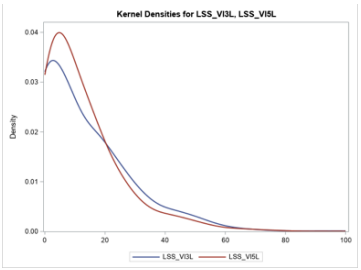
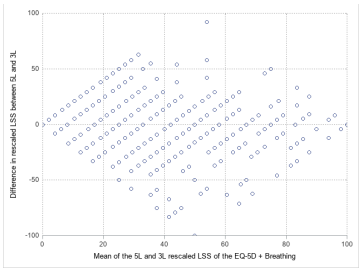
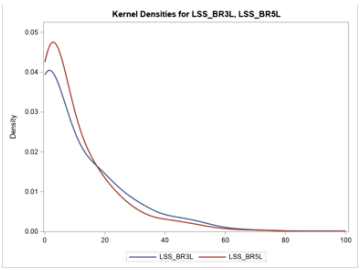
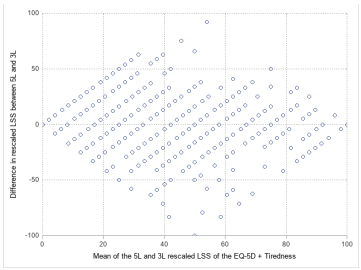
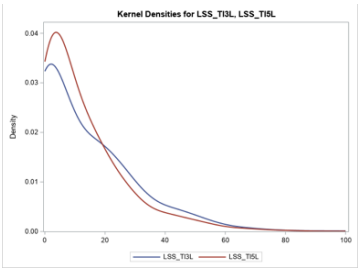
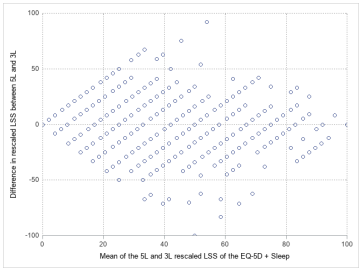
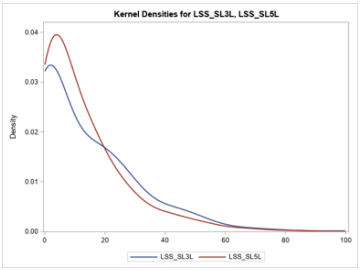
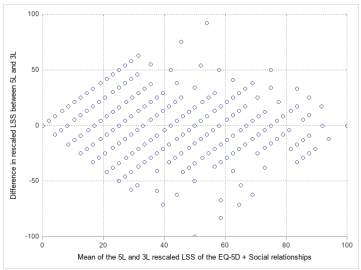
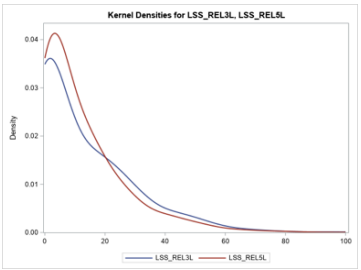
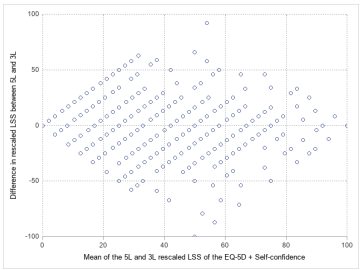
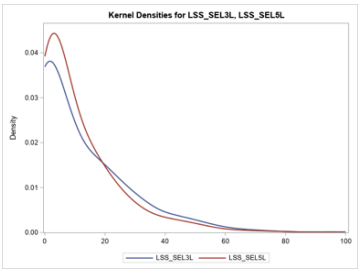
Bolt-on	Bland-Altman plots	Kernel densities of the rescaled LSS
Vision		
Breathing		
Tiredness		
Sleep		
Social Relationships		
Self-confidence		

Table S5. Comparison of known groups validity between 3-level and 5-level bolt-ons

Number of chronic conditions	Ceiling			Mean LSS (transformed to 0-100 scale)			Relative efficiency
	None	1-2	>3	None	1-2	>3	vs no bolt-on
EQ-5D-5L	48.9%	18.4%	2.7%	6	14	28	1.00
EQ-5D-5L + vision	30.3%	10.1%	2.0%	8	16	29	0.98
EQ-5D-5L + breathing	47.5%	16.0%	1.8%	5	13	28	1.12
EQ-5D-5L + tiredness	37.2%	12.2%	2.0%	7	16	31	0.98
EQ-5D-5L + sleep	36.0%	12.4%	1.9%	7	16	31	0.98
EQ-5D-5L + social relationships	39.8%	14.8%	2.0%	7	15	30	1.00
EQ-5D-5L + self-confidence	42.8%	16.2%	2.2%	6	14	28	0.96
EQ-5D-3L	64.0%	29.9%	6.9%	6	16	31	1.00
EQ-5D-3L + vision	48.3%	21.5%	5.1%	8	17	31	0.94
EQ-5D-3L + breathing	63.0%	27.4%	5.2%	6	15	30	1.10
EQ-5D-3L + tiredness	53.4%	22.8%	4.4%	8	18	33	1.05
EQ-5D-3L + sleep	54.2%	23.7%	5.0%	8	18	34	1.06
EQ-5D-3L + social relationships	57.3%	26.0%	6.0%	7	17	33	1.04
EQ-5D-3L + self-confidence	59.4%	27.3%	5.9%	6	16	31	1.01
Caregivers	**Ceiling**			**Mean LSS (transformed to 0-100 scale)**			**Relative efficiency**
	No caregiver	**With caregiver**		**No caregiver**	**With caregiver**		**vs no bolt-on**
EQ-5D-5L	34.2%	10.4%		6.3	35.7		1.00
EQ-5D-5L + vision	20.4%	9.4%		11.2	35.0		0.86
EQ-5D-5L + breathing	32.4%	9.7%		8.8	33.7		1.00
EQ-5D-5L + tiredness	25.2%	8.8%		10.9	36.1		0.86
EQ-5D-5L + sleep	24.6%	8.4%		11.1	37.0		0.89
EQ-5D-5L + social relationships	27.7%	9.0%		10.5	35.3		0.84
EQ-5D-5L + self-confidence	30.0%	8.8%		9.6	34.6		0.92
EQ-5D-3L	47.6%	13.0%		10.4	38.5		1.00
EQ-5D-3L + vision	35.4%	11.8%		11.9	37.4		0.88
EQ-5D-3L + breathing	45.9%	12.6%		9.8	36.7		1.01
EQ-5D-3L + tiredness	38.6%	11.8%		12.2	38.8		0.84
EQ-5D-3L + sleep	39.5%	11.9%		12.4	39.9		0.87
EQ-5D-3L + social relationships	42.3%	12.2%		11.4	38.2		0.86
EQ-5D-3L + self-confidence	43.9%	12.0%		10.6	37.7		0.94

* Relative efficiency is the ratio of the F-test statistic from a model with the rescaled LSS of the EQ-5D descriptive system versus the rescaled LSS of the EQ-5D+bolt-on descriptive system, in an ANOVA model with the known groups as dependent variable