



ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Patient-Reported Outcomes

Comparing Psychometric Properties of 6 of the 5-Level and 3-Level EQ-5D Bolt-Ons in a Large, Multinational, Longitudinal General Population Sample

Sarah Dewilde, PhD, Mathieu F. Janssen, PhD, Glenn Phillips, PhD, Brendan Mulhern, PhD, Fanni Rencz, PhD

ABSTRACT

Objectives: This study systematically compared the psychometric properties of 6 of the 5-level and 3-level EQ-5D bolt-ons (vision, breathing, tiredness, sleep, social relationships, and self-confidence).

Methods: Population Norms Study (POP-UP) is a longitudinal survey collecting general population data from United States, Canada, United Kingdom, Italy, Spain, Germany, the Netherlands, and Belgium. Respondents entered demographics and health conditions data and completed the EQ-5D-5L/3L with bolt-ons, Myasthenia Gravis Activities of Daily Living (MG-ADL), and Health Utility Index-3 (HUI-3). Psychometric properties included ceiling/floor, informativity, divergent and convergent validity, reliability, known-groups validity, and responsiveness.

Results: A total of 9,758 respondents completed the survey in 2021, and 4,839 respondents in 2023. The 5-level bolt-ons reduced ceiling effects by 35% and floor effects by 55% compared with the 3-level, with the largest reductions for vision and sleep (42%/57%) and the smallest for breathing (29%/44%). Informativity was higher for the 5-level bolt-ons (3%-11%) than 3-level (2%-9%), except for breathing (-2%). Most bolt-ons showed weak to moderate correlations with EQ-5D dimensions, except for social relationships, which correlated strongly with anxiety/depression ($r = 0.61$). Both measures showed strong convergent validity for breathing (5-level $r = 0.68$; 3-level $r = 0.66$) and lower for vision (5-level $r = 0.37$; 3-level $r = 0.32$). EQ-5D + 5-level and EQ-5D-5L + 3-level bolt-ons had excellent reliability (Intraclass Correlation Coefficients = 0.92-0.97). Adding bolt-ons did not significantly improve known-groups validity, except for breathing (relative efficiency EQ-5D + 5-level = 1.12; EQ-5D-5L + 3-level = 1.10). The 5-level bolt-ons improved responsiveness, detecting 47.1% more improvements and 55.7% more worsening over time.

Conclusions: All 6 of the 5-level EQ-5D bolt-ons outperformed the 3-level bolt-ons by reducing ceiling and floor, enhancing informativity, demonstrating divergent validity, and improving responsiveness.

Keywords: bolt-on, EQ-5D-3L, EQ-5D-5L, general population, psychometric properties.

VALUE HEALTH. 2025; ■(■):■-■

Highlights

- The EuroQol Group has prioritized developing bolt-on dimensions for the EQ-5D, although the bolt-ons 5-level versions' presumed superiority over 3-level versions is unclear. We evaluated 6 bolt-ons—vision, breathing, tiredness, sleep, social relationships, and self-confidence—comparing their 5-level and 3-level versions in the general population.
- The 6 5-level EQ-5D bolt-ons outperformed the 3-level by reducing ceiling and floor, enhancing informativity, demonstrating divergent validity, and improving responsiveness.
- Vision, sleep, and tiredness bolt-ons showed particular promise for expanding the EQ-5D-5L. These findings support ongoing bolt-on development and offer insights for broader instrument design. Further research is needed to test their performance in patient populations and examine country-specific differences in psychometric and economic outcomes.

Introduction

The EQ-5D is the most widely used preference-based measure of health-related quality of life (HRQoL) applied in observational studies, clinical trials, and patient registries to estimate quality-adjusted life-years (QALYs) for cost-utility analyses.¹ The EQ-5D covers 5 HRQoL dimensions.^{2,3} Interest is growing in adding “bolt-on” dimensions to improve the EQ-5D's sensitivity for specific populations or contexts.⁴ Approximately 50 different bolt-ons have been proposed for the adult EQ-5D instruments, addressing symptoms (eg, vision and breathing), physical and psychological functioning (eg, mobility and stress), social relationships, and overall quality of life (eg, well-being and dignity).⁵⁻⁸

Three-level and 5-level versions exist for most bolt-ons.⁹ Although numerous studies have compared the measurement performance of the EQ-5D's 3-level and 5-level versions,¹⁰ both bolt-on versions are rarely used in the same study. As a result, little is known about the assumed superior performance of the 5-level bolt-ons compared with the 3-level bolt-ons. Furthermore, very few studies provided evidence on the sensitivity or responsiveness of bolt-ons.^{11,12} In addition, although most bolt-ons are used in the context of specific patient populations,^{11,13-25} they may also capture HRQoL aspects relevant to the general population.

This study addresses key gaps in evidence on 5-level vs 3-level bolt-ons, including the incipient knowledge on their responsiveness,^{11,13,15} and psychometric properties in the general

population.^{26,27} We systematically compared 6 bolt-ons across both versions using a large multinational longitudinal survey. We hypothesize that the 5-level versions will outperform the 3-level ones by offering better distribution, informativity, group differentiation, and responsiveness.

Methods

Data Collection

This secondary analysis uses data from the Population Norms Study (POP-UP), a multinational digital survey assessing HRQoL norms in the general population across 8 countries (United States, Canada, United Kingdom, Italy, Spain, Germany, the Netherlands, and Belgium). These norms serve as a baseline for comparing health outcomes in specific conditions, such as Myasthenia Gravis.^{28,29} The first survey wave took place in Q1 2021 and the second in Q1 2023, recontacting previous participants and adding new ones as needed. The target sample was 13,500 (1,000 per country), with quotas based on age, gender, education, and region. Low-quality responses were excluded. Participants were invited via email and completed demographic questions, medical history, and a fixed sequence of questionnaires: the EQ-5D-5L and 5-level bolt-ons, the Myasthenia Gravis Activities of Daily Living (MG-ADL),³⁰ the Health Utility index-3 (HUI-3),³¹ the Hospital Anxiety and Depression Scale,³² and the EQ-5D-3L and 3-level bolt-ons. The survey, built in LimeSurvey, required responses for all items. Ethical approval was obtained in all 8 countries: Veritas Institutional Review Board (Canada, #2021-2434-5740-1), Ghent University Hospital (Belgium, #BC-07857), and Salus Institutional Review Board (other countries, #PN8450). Full details are available elsewhere.^{28,29}

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

The EQ-5D-3L assesses 5 HRQoL dimensions—mobility, self-care, usual activities, pain/discomfort, and anxiety/depression—each with 3 severity levels. It also includes the EQ VAS, in which respondents rate their health from 0 (worst) to 100 (best) on the day of the survey.² The EQ-5D-5L, with 5 response levels, was developed to address the EQ-5D-3L's limited sensitivity and ceiling effects.³ Response levels describe a health state that is converted into a utility value based on population preferences.³³ Because no value set exists for bolt-ons, Level Sum Scores (LSSs) were calculated for the EQ-5D-3L and 5L, with and without each bolt-on, and rescaled to 0 to 100. Higher scores indicate more problems.^{34,35}

Bolt-On Dimensions

The bolt-ons were added to the EQ-5D dimensions before the EQ VAS and included vision,^{36,37} breathing,³⁸ tiredness,²⁰ sleep,³⁹ social relationships,²⁰ and self-confidence,²⁰ with 3 and 5 response levels as the EQ-5D descriptive system (see Appendix Table 1 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2025.09.009>).⁴⁰ Bolt-ons were selected based on their relevance to general HRQoL, broad applicability across populations, and to provide baseline norms for comparing health outcomes in populations with Myasthenia Gravis.⁴⁰ Bolt-ons were developed within the EQ-5D framework, translated, published, and approved by EuroQoL.

Statistical Analysis

Our primary aim was to compare the performance of 6 EQ-5D-5L + 5-level bolt-ons with EQ-5D-3L + 3-level bolt-ons, using an

established framework for 3-level vs 5-level comparisons.¹⁰ Comparisons with the core EQ-5D-3L and 5L (without bolt-ons) were examined secondarily. To maximize paired assessments, data from respondents in either or both waves were included. We assessed distributional characteristics, divergent and convergent validity, reliability, known-groups validity, and responsiveness. These psychometric properties were selected based on prior literature comparing the 3-level and 5-level EQ-5D versions, the availability of longitudinal data, and the expertise of the authors.^{10,41} Formal hypothesis tests were not applied because the study was descriptive in nature.

Distributional characteristics

Distributional characteristics included analyzing the distribution of the EQ-5D and EQ-5D + bolt-ons' responses, the proportion of different health profiles, ceiling and floor effects, and informativity.³⁴ The distribution of responses was cross-tabulated between the 2 EQ-5D descriptive systems. Subsequently, the proportion of inconsistent response pairs was calculated. Inconsistent response pairs were defined as responses differing by at least 2 levels between the 3-level and 5-level EQ-5D versions. We additionally calculated a weighted proportion of inconsistent pairs by multiplying the proportion of inconsistent responses by a weight (weights of 2, 3, or 4), depending on the number of levels they differ, assigning greater importance to larger inconsistencies.⁴² The size of the inconsistency was calculated using the formula: absolute (response3L_recoded—response5L), in which the 3-level responses were recoded from 1, 2, and 3 to 1, 3, and 5, respectively.⁴² The size of inconsistency reflects variability in responses within each dimension, with higher values indicating greater discordance. We hypothesize that the added response options in the 5-level bolt-ons will increase response distribution and sensitivity compared with the 3-level versions.

Proportion of different profiles, ceiling, and floor

The proportion of different health profiles was calculated as the number of observed profiles divided by the number of possible profiles.³⁴ Ceiling and floor were defined as the proportion of respondents with the best (11111) and worst (33333 or 55555) health state, respectively.³⁴ The 5-level versions (EQ-5D-5L and bolt-ons) provide more response options, reducing clustering at “no problems” or “extreme problems.” We expected this to result in a relative reduction in ceiling and floor compared with the 3-level versions.

Informativity

The informativity of bolt-on was determined using Shannon's indices, with the Shannon index (H') for absolute and Shannon Evenness index (J') for relative informativity. H' measures response spread, with higher values indicating a more even distribution and more detailed health descriptions.⁴³ J' measures response distribution, regardless of response levels. If most respondents choose the same level, informativity is lower, indicating limited new information from the item or instrument.⁴³ H' and J' indices were calculated at item and instrument levels and assessed descriptively.⁴³ We expected the 5-level bolt-ons to have higher informativity than the 3-level versions at the item level and an overall improvement in informativity at the instrument level after adding the bolt-ons to their respective EQ-5D versions.

Divergent validity

Divergent validity was assessed using Spearman's rank order correlations between bolt-ons and EQ-5D dimensions.^{44,45}

Correlations were considered none ($r < 0.10$), weak ($r = 0.10$ – 0.29), moderate ($r = 0.30$ – 0.49), and strong ($r \geq 0.50$).⁴⁶ Weak or no correlations indicate that the bolt-ons and EQ-5D dimensions are unrelated constructs, thus avoiding overlap. Low correlations between vision and EQ-5D dimensions were expected. Moderate correlations were anticipated between the other bolt-ons (tiredness, sleep, and social relationships) and usual activities, between breathing and pain/discomfort, as well as between social relationships and self-confidence and anxiety/depression.^{24–27} We also hypothesized that the correlations between the EQ-5D-5L and the 5-level bolt-ons would be stronger than those between the EQ-5D-3L and the 3-level bolt-ons.

Convergent validity

Convergent validity was assessed using Spearman's rank order correlations between bolt-ons and available Patient-Reported Outcomes-Measures (PROMs) in the POP-UP study that evaluate related constructs.^{44,45} We hypothesized strong correlations between the vision bolt-on and HUI-3's vision item, as well as between the breathing bolt-on and the MG-ADL breathing item, given their conceptual similarity. Convergent validity of tiredness, sleep, social relationships, and self-confidence was not assessed, as specific PROMs were not available.

Reliability

Reliability refers to how consistently the same construct is measured over time, and it was assessed using the Intraclass Correlation Coefficients (ICC)⁴⁷ using the SAS ICC macro.⁴⁸ A general linear mixed model was fitted, including the rescaled LSS for the EQ-5D-3L and EQ-5D-5L, alone and with each bolt-on. ICC values indicated poor (<0.5), moderate (0.50 – 0.75), good (0.75 – 0.90), or excellent (>0.90) reliability.⁴⁹ The 5- and 3-level EQ-5D with bolt-ons were expected to show good reliability ($r > 0.75$), as both versions have demonstrated strong agreement in general populations, especially for overall health status.¹⁰ Kernel density plots were used to visualize LSS distributions for each bolt-on, EQ-5D-3L, and EQ-5D-5L. Similar distributions indicate good agreement between versions, which was further illustrated using Bland-Altman plots.

Known-groups validity

Known-groups validity was assessed based on the number of chronic health conditions (none/1-2/3+) and the need for caregiver assistance (yes/no), as these may reflect overall health and function.^{50–52} Distinguishing these groups with bolt-ons could inform policymakers on population-level interventions. We compared subgroups based on the proportion reporting the ceiling on EQ-5D + bolt-on, mean LSSs, and relative efficiency. Relative efficiency was calculated as the ratio of 2 F-statistics: 1 from an Analysis of Variance (ANOVA) model with EQ-5D + bolt-on LSSs and subgroup as the independent variable, and the other with EQ-5D LSSs. Values > 1.00 indicated improved efficiency with bolt-ons. We expected lower ceiling effects and higher LSS in subgroups with more chronic conditions or caregiver needs, and improved relative efficiency with bolt-ons added to the EQ-5D.

Responsiveness

Responsiveness refers to the ability of the EQ-5D + bolt-ons to detect changes over time.⁴⁷ Responsiveness was measured using level change in responses, Paretian Classification of Health Change, and probability of superiority.³⁴ Level change in responses for the 3- and 5-level bolt-ons were calculated using data exclusively from respondents who completed both waves.

Changes in responses may range between -4 and 4 for the 5-level and -2 and 2 for the 3-level bolt-ons. We examined the mean level change, the frequency of these changes, and the mean change in LSSs. Furthermore, Pearson's correlations between changes in mean 5-level and 3-level EQ-5D + bolt-on LSSs were calculated. We expected to find a higher level change in responses in the 5-level compared with the 3-level descriptive system.

Paretian classification health change

Paretian classification health (PCHC) was used to assess changes in EQ-5D + bolt-on health states over time, once again using data exclusively from respondents who participated in both waves. These changes were categorized as improved, worsened, mixed, or no change.³⁴ A state was "better" if at least 1 dimension improved without any worsening and "worse" if 1 worsened without any improvement. Mixed indicated both improvement and worsening; no change meant identical responses across all dimensions. Analyses were repeated, excluding respondents with perfect health (11111) at wave 1. We expected the 5-level bolt-ons to show more change and less stability than the 3-level.

Probability of superiority

Based on the PCHC classification, for each core EQ-5D dimension and bolt-on separately, we calculated the probability to detect a superior health state between waves 1 and 2 using a non-parametric method for measuring effect sizes.⁵³ The probability (P) was calculated as the number of respondents who improved, plus half the number of stables, divided by the number of respondents at both waves. A $P < .5$ indicates more respondents deteriorated than improved, whereas $P > .5$ indicates more improved than deteriorated. We expected the 5-level bolt-ons to show a higher probability of superiority.

Results

Study Participants

In total, 9,758 general population participants from the 8 countries completed the survey in 2021 (wave 1) and 4,839 respondents completed it in 2023 (wave 2), comprised 3,571 first-wave participants and 1,268 new respondents, totaling 14,597 observations. Around 60% of respondents reported 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%) (Table 1).

Available Data on the EQ-5D

The combined sample completing the EQ-5D-3L + bolt-ons and EQ-5D-5L + bolt-ons in waves 1 and 2 was 14,597 respondents, of which 3,571 completed data in both waves (Table 1).

Distributional Characteristics

Social relationships had the highest rate of extreme problems (4.2%), followed by tiredness (4.1%) and sleep (4.0%), whereas mobility, self-care, usual activities, and breathing had the highest reports of no problems (Table 2). On average, 3.5% of EQ-5D responses were inconsistent, ranging from 2.5% (breathing) to 6.7% (vision). After weighting, inconsistency rose to 7.8% for EQ-5D responses, with 5.5% for breathing and 14.5% for vision (see Appendix Table 2 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2025.09.009>). This suggests inconsistency

Table 1. Description of respondent characteristics in waves 1 and 2 of the POP-UP study.

Proportion of respondents	Wave 1	Wave 2
	n = 9758	n = 4839
Age (years)		
18-34	27.5%	27.6%
35-54	37.0%	36.9%
55+	35.5%	35.5%
Gender		
Female	51.3%	51.4%
Male	48.8%	48.6%
Education		
Primary education	4.7%	4.4%
Secondary education	56.2%	55.3%
Higher education	39.1%	40.4%
EQ VAS		
Mean (SD)	75.7 (17.4)	74.1 (17.7)
Health condition		
None	41.0%	41.5%
Types of health conditions		
Anxiety	14.4%	15.7%
Depression	11.4%	12.0%
Diabetes with/without complications	10.0%	10.4%
Thyroid problems/thyroid disorder	7.9%	8.8%
Respiratory disease (COPD and asthma)	8.7%	8.3%
Gastrointestinal problems	6.7%	6.2%
Rheumatoid arthritis/psoriatic arthritis	4.7%	4.4%
Cardiovascular disease	4.5%	4.0%
Osteoporosis	3.2%	4.2%
Cancer	2.0%	1.8%
Kidney disease	1.0%	1.2%
Congestive heart failure	0.9%	1.0%
Crohn's/ulcerative colitis	0.7%	0.5%
Liver disease	0.8%	0.6%
Multiple sclerosis	0.5%	0.6%
Peptic ulcer	0.5%	0.5%
HIV	0.4%	0.4%
Systemic lupus erythematosus	0.3%	0.2%
Hemiplegia/paraplegia	0.3%	0.4%
Lupus nephritis	0.2%	0.1%
Dementia	0.1%	0.0%

COPD indicates chronic obstructive pulmonary disease.

increases with problem severity, especially in core EQ-5D dimensions.

Vision (0.52) and tiredness (0.50) showed the highest inconsistency across bolt-ons, whereas breathing (0.22) and self-confidence (0.30) showed the lowest. Vision's inconsistency may stem from its 3-level label "some problems" differing from the "moderate problems" used in other bolt-ons and the 5-level system. Other inconsistencies may reflect how respondents interpret severity across systems with different response options.

Proportion of Different Profiles, Ceiling, and Floor

Adding any bolt-on increased the number of unique profiles but reduced the proportion of observed profiles: from 63.0% to ~40.6% for EQ-5D-3L and from 22.3% to ~8.4% for EQ-5D-5L (Table 3). The 5-level system captured 3.2 to 3.5 times more profiles than the 3-level. The fewest profiles were generated by the EQ-5D + breathing, and the most by EQ-5D-5L + tiredness or sleep, or EQ-5D-3L + tiredness or self-confidence.

Ceilings were 45.4% for EQ-5D-3L and 32.6% for EQ-5D-5L (Table 3). The 5-level bolt-ons reduced ceilings 32% to 42% more than 3-level versions, with highest relative reductions observed for vision (42%), sleep (38%), and tiredness (34%). Floor effects were 50% to 67% lower with 5-level bolt-ons, with the biggest reductions from sleep (67%), social relationships (58%), and self-confidence (58%).

Informativity

The 5-level bolt-ons captured more information than the 3 level, as shown by higher absolute and relative item informativity (Table 4). Sleep, tiredness, and vision performed the best, whereas breathing and self-confidence captured the least. At the instrument level, adding vision, tiredness, or sleep improved relative informativity by 8% to 11%, whereas breathing reduced it by 2%.

Divergent Validity

Correlations between bolt-ons and EQ-5D dimensions were mostly weak to moderate, except for a strong correlation between social relationships and anxiety/depression (see Appendix Table 3 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2025.09.009>). Social relationships also strongly correlated with self-confidence. As expected, correlations were about 5% stronger with the 5-level bolt-ons, reflecting its greater sensitivity.

Convergent Validity

In both systems, breathing showed strong correlations with the MG-ADL breathing item (5-level $r = 0.68$; 3-level $r = 0.66$). In contrast, vision showed only moderate correlation with the HUI-3 vision item (5-level $r = 0.37$; 3-level $r = 0.32$), likely due to differences in how vision is assessed—EQ-5D focuses on general vision status, whereas HUI-3 measures functional ability.

Reliability

Excellent reliability was observed between the LSSs of EQ-5D-5L + bolt-ons and EQ-5D-3L + bolt-ons, indicating that adding bolt-ons did not introduce excessive variability in LSS over time (ICC range 0.92-0.97, all $P < .001$, Appendix Fig. 1 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2025.09.009>). Good agreement was confirmed by similar distributions Kernel density plots and by the scattered points in the Blant-Altman plots (see Appendix Fig. 2 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2025.09.009>).

Known-Groups Validity

For all bolt-ons, the ceiling was reduced in respondents with multiple chronic health conditions or those needing a caregiver (see Appendix Table 4 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2025.09.009>). Higher mean LSS scores were observed for most bolt-ons in respondents with multiple chronic conditions, except for the breathing bolt-on, which had lower LSS scores than the EQ-5D-3L and EQ-5D-5L alone. Higher mean LSS scores were observed only for the tiredness and sleep bolt-ons in respondents needing a caregiver. However, adding either the 3-level or 5-level bolt-ons to the EQ-5D did not improve relative efficiency compared with the EQ-5D alone, except for breathing in respondents with multiple chronic health conditions (relative efficiency EQ-5D + 3-level = 1.10; EQ-5D-5L + 5-level = 1.12).

Table 2. Distribution of the EQ-5D-3-level responses across the EQ-5D-5-level responses (n = 14,597).

5Q-5D-3L responses	EQ-5D-5L responses					Total
	1	2	3	4	5	
Mobility						
1	75.2%	6.1%	0.9%	0.2%	0.2%	82.5%
2	1.4%	8.4%	5.0%	1.8%	0.4%	16.9%
3	0.2%	0.1%	0.1%	0.1%	0.1%	0.6%
Total	76.8%	14.5%	5.9%	2.1%	0.7%	
Self-care						
1	90.0%	2.3%	0.4%	0.1%	0.1%	92.8%
2	1.1%	3.4%	1.5%	0.3%	0.1%	6.5%
3	0.1%	0.1%	0.1%	0.2%	0.2%	0.7%
Total	91.2%	5.8%	2.0%	0.6%	0.4%	
Usual activities						
1	75.6%	6.5%	0.8%	0.1%	0.1%	83.1%
2	2.2%	7.5%	4.6%	1.2%	0.1%	15.7%
3	0.1%	0.1%	0.2%	0.4%	0.3%	1.2%
Total	77.9%	14.2%	5.6%	1.8%	0.5%	
Pain/discomfort						
1	44.3%	12.5%	0.8%	0.1%	0.1%	57.8%
2	2.1%	23.1%	12.4%	2.1%	0.2%	39.8%
3	0.1%	0.1%	0.2%	1.4%	0.6%	2.4%
Total	46.5%	35.7%	13.4%	3.6%	0.8%	
Anxiety/Depression						
1	57.7%	10.2%	1.3%	0.2%	0.1%	69.5%
2	2.0%	14.3%	9.0%	1.7%	0.3%	27.3%
3	0.2%	0.2%	0.5%	1.4%	1.1%	3.3%
Total	59.8%	24.6%	10.8%	3.3%	1.5%	
Vision						
1	39.7%	16.1%	4.2%	0.5%	0.1%	60.5%
2	1.4%	19.4%	15.5%	1.8%	0.1%	38.2%
3	0.1%	0.2%	0.1%	0.5%	0.3%	1.3%
Total	41.1%	35.6%	19.9%	2.9%	0.5%	
Breathing						
1	77.2%	6.0%	0.6%	0.1%	0.1%	84.0%
2	1.3%	9.0%	4.1%	0.6%	0.1%	15.0%
3	0.1%	0.1%	0.1%	0.4%	0.3%	1.0%
Total	78.6%	15.1%	4.8%	1.1%	0.4%	
Tiredness						
1	47.6%	9.8%	1.0%	0.2%	0.1%	58.6%
2	0.7%	20.5%	13.4%	2.6%	0.1%	37.3%
3	0.1%	0.1%	0.5%	2.1%	1.3%	4.1%
Total	48.4%	30.4%	14.8%	4.9%	1.5%	
Sleep						
1	40.7%	13.9%	1.1%	0.2%	0.1%	56.1%
2	1.7%	22.4%	13.0%	2.6%	0.2%	39.9%
3	0.0%	0.2%	0.6%	1.9%	1.3%	4.0%
Total	42.5%	36.5%	14.7%	4.7%	1.6%	
Social relationships						
1	55.5%	11.0%	1.5%	0.2%	0.1%	68.3%
2	1.2%	13.7%	9.8%	2.6%	0.2%	27.6%
3	0.1%	0.1%	0.4%	1.8%	1.8%	4.2%
Total	56.8%	24.8%	11.7%	4.6%	2.1%	
Self-confidence						
1	66.6%	8.7%	1.1%	0.2%	0.1%	76.6%
2	1.7%	10.2%	6.6%	1.7%	0.2%	20.4%
3	0.2%	0.1%	0.5%	1.2%	1.1%	3.1%
Total	68.4%	19.0%	8.2%	3.1%	1.3%	

Red shaded cells indicate inconsistency between the 3- and the 5-level response

Table 3. Relative reductions in the number of unique profiles, ceiling, and floor.

	EQ-5D-3L		EQ-5D-5L		EQ-5D-5L vs EQ-5D-3L	EQ-5D-3L 1 bolt-on vs EQ5D-3L alone	EQ-5D-5L 1 bolt-on vs EQ- % 5D-5L alone
	%	n	%	n			
EQ-5D							
Number of possible profiles	63	153	22.3	696	3.5	Ref	Ref
Ceiling: 11111	45.4	14 597	32.6	14 596	-28%	Ref	Ref
Floor: 55555 or 33333	0.13	14 597	0.06	14 597	-54%	Ref	Ref
EQ-5D + vision							
Number of possible profiles	40.1	292	8.3	1300	3.5	87%	91%
Ceiling: 111111	33.8	14 597	19.7	14 596	-42%	-40%	-25%
Floor: 555555 or 333333	0.11	14 597	0.05	14 597	-55%	-17%	-15%
EQ-5D + breathing							
Number of possible profiles	38.3	279	7.9	1242	3.5	78%	82%
Ceiling: 111111	43.7	14 597	30.9	14 596	-29%	-5%	-4%
Floor: 555555 or 333333	0.09	14 597	0.05	14 597	-44%	-17%	-31%
EQ-5D + tiredness							
Number of possible profiles	41.6	303	8.7	1362	3.5	96%	98%
Ceiling: 111111	36.8	14 597	24.1	14 596	-34%	-26%	-19%
Floor: 555555 or 333333	0.1	14 597	0.05	14 597	-50%	-17%	-23%
EQ-5D + sleep							
Number of possible profiles	40.6	296	8.6	1343	3.5	93%	93%
Ceiling: 111111	37.7	14 597	23.6	14 596	-38%	-28%	-17%
Floor: 555555 or 333333	0.12	14 597	0.04	14 597	-67%	-33%	-8%
EQ-5D + social relationships							
Number of possible profiles	40.9	298	8.4	1318	3.4	89%	95%
Ceiling: 111111	40.3	14 597	26.5	14 596	-34%	-19%	-11%
Floor: 555555 or 333333	0.12	14 597	0.05	14 597	-58%	-17%	-8%
EQ-5D + self-confidence							
Number of possible profiles	42	306	8.3	1299	3.2	87%	100%
Ceiling: 111111	41.8	14 597	28.6	14 596	-32%	-12%	-8%
Floor: 555555 or 333333	0.12	14 597	0.05	14 597	-58%	-17%	-8%
Average (%)							
Number of possible profiles	40.6		8.4		3.4	93%	88%
Ceiling: 111111	39		25.6		-35%	-14%	-22%
Floor: 555555 or 333333	0.11		0.05		-55%	-15%	-19%

Responsiveness

Of the 3,571 respondents who provided EQ-5D + bolt-ons data in both waves, about 72.9% did not experience a change in the bolt-on dimensions, whereas 14.0% improved and 13.5% worsened (Table 5). The vision, tiredness, and sleep bolt-ons captured the most improvements and the most worsening over time, whereas breathing detected considerably fewer changes.

The 5-level bolt-ons picked up more changes in responses over time. It identified a lower proportion of respondents reporting being stable (on average 14.3% fewer for the 5-level bolt-ons), a higher proportion of improvements (on average 47.1% increase in improvements) and a higher proportion of worsening responses (55.7% more worsening). We detected no pattern in the mean changes in LSSs, which was sometimes higher and sometimes lower in the 5-level bolt-ons. The correlations between the changes in EQ-5D-3L + bolt-on vs EQ-5D-5L + bolt-on LSSs were strong (0.519-0.588).

Paretian Classification Health Change

Table 6 shows the PCHC health state changes for the EQ-5D + bolt-ons (as opposed to Table 5 which shows responsiveness in the bolt-ons individually). The 5-level descriptive system was more sensitive to change than the 3-level, detecting relative

reduction of 34.7% fewer status quo balanced out by 128.7% more mixed changes, 9.4% more improvements, and 10.5% more worsening. Excluding respondents scoring 11111 yielded similar results, with mostly more mixed changes at the expense of reductions in the status quo. However, for EQ-5D + vision, the proportion of respondents with mixed changes or no changes was almost the same in both systems.

Probability of Superiority

The probability that health states were rated better in wave 2 than wave 1 averaged 0.50 for both 3-level and 5-level systems, with little variation across dimensions. The lowest was 0.48 (sleep), and the highest 0.51 (anxiety/depression, vision, tiredness, and social relationships). Overall, HRQoL remained stable, suggesting that 2021 population norms still applied in 2023.

Discussion

Main Findings

The EQ-5D-5L + 5-level bolt-ons generally outperformed the EQ-5D-3L + 3-level bolt-ons in most psychometric properties. On average, the 5-level bolt-ons reduced ceiling by 35% and floor by

Table 4. Comparison of informativity at the item and instrument level.

	Item Shannon H	Item Shannon J	EQ-5D + bolt-on Shannon H	EQ-5D + bolt-on Shannon J	Change in EQ-5D + bolt-on Shannon H	Change in EQ-5D + bolt-on Shannon J
EQ-5D-5L + 5-level bolt-on						
EQ-5D-5L	-	-	4.89	0.42	<i>Ref</i>	<i>Ref</i>
Vision	1.70	0.73	6.51	0.47	33%	11%
Breathing	1.00	0.43	5.74	0.41	17%	-2%
Tiredness	1.74	0.75	6.41	0.46	31%	9%
Sleep	1.77	0.76	6.34	0.45	30%	8%
Social relationships	1.65	0.71	6.22	0.45	27%	6%
Self-confidence	1.36	0.59	6.05	0.43	24%	3%
EQ-5D-3L + 3-level bolt-on						
EQ-5D	-	-	3.24	0.41	<i>Ref</i>	<i>Ref</i>
Vision	1.05	0.66	4.22	0.44	30%	9%
Breathing	0.69	0.44	3.79	0.40	17%	-2%
Tiredness	1.17	0.74	4.20	0.44	30%	8%
Sleep	1.18	0.75	4.11	0.43	27%	6%
Social relationships	1.08	0.68	4.03	0.42	25%	4%
Self-confidence	0.92	0.58	3.96	0.42	22%	2%

Shannon H indicates absolute informativity; Shannon J, relative informativity.

55% vs the 3-level versions, with vision, sleep, and tiredness lowering the ceiling most, and sleep, social relationships, and self-confidence reducing the floor most. Reductions in the ceiling and floor may increase inconsistency between descriptive systems¹⁰; for vision, inconsistency may additionally stem from differences in response wording. The 5-level bolt-ons enhanced informativity, with the vision, tiredness, and sleep bolt-ons capturing more information. Informativity improved with the vision, tiredness, and sleep bolt-ons, likely because these aspects are not directly covered in the EQ-5D dimensions. In contrast, breathing overlapped with EQ-5D dimensions, adding redundancy and slightly reducing informativity.

The EQ-5D-5L and 5-level bolt-ons showed slightly higher divergent validity compared with the 3-level, likely due to more precise measurement.^{51,54-59} However, social relationships and anxiety/depression had a strong overlap, consistently with findings of previous studies.^{22,24} Convergent validity was limited by the unavailability of PROMs for some bolt-ons, but breathing had strong correlations with the MG-ADL's breathing problems. Vision showed only a moderate correlation with HUI-3's vision, likely due to differences in constructs but also possibly due to measurement error, imprecision in external anchors, or translation effects. As anticipated, there was an excellent reliability between the LSSs of EQ-5D-5L + bolt-ons and

Table 5. Comparison of responsiveness.

Change wave 2-wave 1	Vision		Breathing		Tiredness		Sleep		Social relationships		Self-confidence		Average	
	3-level	5-level	3-level	5-level	3-level	5-level	3-level	5-level	3-level	5-level	3-level	5-level		
Mean change in rescaled LSS	0.1	0.04	0.22	0.17	0.05	0.07	0.39	0.4	0.14	0.05	0.09	0.18	-	
Improvement	-4	n/a	0.1%	n/a	0.0%	n/a	0.1%	n/a	0.1%	n/a	0.1%	n/a	0.1%	14.00%
	-3	n/a	0.4%	n/a	0.1%	n/a	0.3%	n/a	0.2%	n/a	0.4%	n/a	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%	
Stable	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%	72.90%
Worsening	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%	13.50%
	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	n/a	0.2%	n/a	0.0%	n/a	0.4%	n/a	0.3%	n/a	0.4%	n/a	0.4%	
	4	n/a	0.0%	n/a	0.0%	n/a	0.1%	n/a	0.1%	n/a	0.0%	n/a	0.1%	
Change in %														
Stable*		-17.4%		-8.2%		-18.0%		-18.1%		-14.0%		-10.1%		-14.3%
Improvement*		50.3%		55.1%		49.1%		40.7%		58.5%		29.2%		47.1%
Worsening*		51.0%		55.5%		64.6%		54.5%		50.9%		57.6%		55.7%
Correlation in change*		0.519		0.565		0.586		0.584		0.588		0.583		-

LSS indicates level sum scores.

*Change in the 5-level vs the 3-level rescaled LSS.

Table 6. Comparison of Paretian classification health change.

	EQ-5D		EQ-5D + vision		EQ-5D + breathing		EQ-5D + tiredness		EQ-5D + sleep		EQ-5D + social relationships		EQ-5D + self-confidence	
	3L	5L	3L	5L	3L	5L	3L	5L	3L	5L	3L	5L	3L	5L
For all respondents														
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%
Stable	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%
<i>Comparison of 5-level vs 3-level :</i>														
Difference in % improved	16%		9%		10%		8%		5%		11%		7%	
Difference in % mixed	145%		109%		129%		121%		140%		143%		114%	
Difference in % stable	-28%		-37%		-30%		-37%		-37%		-42%		-31%	
Difference in % worsened	14%		5%		9%		10%		8%		18%		10%	
For all respondents worse than 11111 (1)														
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%
Stable	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%
<i>Comparison of 5-level vs 3-level :</i>														
Difference in % improved	-6%		-17%		-10%		-10%		-13%		-11%		-12%	
Difference in % mixed	100%		2%		87%		85%		97%		94%		76%	
Difference in % stable	-35%		8%		-35%		-45%		-39%		-38%		-35%	
Difference in % worsened	18%		20%		10%		14%		7%		5%		11%	

EQ-5D-3L + bolt-ons, indicating that adding bolt-ons seem to have minimal impact to the difference between 3-level and 5-level.⁶⁰

Ceiling was reduced, and mean LSS scores were higher among respondents with multiple chronic conditions or those needing a caregiver. However, adding 3- or 5-level bolt-ons to the EQ-5D did not substantially improve discrimination between known groups, except for a modest gain with breathing. The limited efficiency gain may reflect the heterogeneity of health conditions in the general population and varying relevance of bolt-ons. Disease-specific comparisons were beyond the scope of this study, which aimed to evaluate the psychometric performance of 3- vs 5-level bolt-ons in a general population context.

The 5-level bolt-ons captured more change over time than the 3-level versions, with more respondents reporting improvement (notably in social relationships, breathing, and vision) and worsening (mainly in tiredness, self-confidence, and breathing). PCHC analysis confirmed greater responsiveness of the 5-level system, with fewer cases of no change. Improvement probabilities were similar (~0.5) across systems, with minimal change across dimensions despite the 2-year interval spanning the start of the COVID-19 pandemic. Key results are summarized in [Appendix Table 5 in Supplemental Materials](https://doi.org/10.1016/j.jval.2025.09.009) found at <https://doi.org/10.1016/j.jval.2025.09.009>.

Comparison With the Literature

Comparisons with existing literature are limited because most studies assess either the EQ-5D-3L or 5L separately in disease-specific populations. To our knowledge, no prior studies have compared ceiling effects between 3- and 5-level bolt-ons. Previous research comparing the EQ-5D-5L and 3L reported a 27% relative ceiling reduction in a student population,⁴¹ similar to our 28%. In our study, adding bolt-ons exceeded this benchmark, with reductions of 32% to 42%, supporting their psychometric meaningfulness. Few have examined these bolt-ons in the general population, and none have compared their 3- and 5-level versions.^{26,27} The first study assessed ceiling effects, divergent/convergent validity, and known-groups validity for the 5-level versions of the 6 bolt-ons using general population data from

Hungary.²⁶ A key shared finding was that all bolt-ons reduced ceiling. Both studies found vision as one of the lower-scoring dimensions, with Hungary showing weaker correlations across EQ-5D-5L dimensions. Differences in PROMs and known-groups analyses limit direct comparisons. The second study, conducted in Australia, found a 6% ceiling reduction in breathing—comparable to our 5%—and weak to moderate correlations with EQ-5D-5L dimensions ($r = 0.39-0.48$ vs ours: $0.29-0.42$).²⁷ Known-groups comparisons were limited because of differing subgroup definitions.

Strengths and Limitations

To our knowledge, this is the first study to systematically compare EQ-5D-3L + 3-level and EQ-5D-5L + 5-level bolt-ons. Using a large, representative, and complete general population sample across 2 waves, we conducted extensive analyses, including responsiveness. To our knowledge, this is the largest study using 5-level bolt-ons and one of few with 3-level bolt-on data. Unlike previous research focused on vision and self-confidence in patient populations with cataract and psoriasis^{11,13,15}, we also assessed responsiveness for breathing, tiredness, sleep, and social relationships bolt-ons for the first time.

This study has a few limitations. First, the order of the descriptive systems was not randomized, and neither was the order of the bolt-ons; therefore, ordering effects cannot be excluded. Second, not all wave 1 participants responded in wave 2. Of 4,500 recontacted, 3,571 completed both waves with no missing EQ-5D data (79.3% response rate). In the United States and Canada, most wave 2 participants were new due to low retention in online panels. Still, wave 2 remained nationally representative by age, gender, education, and region. No systematic differences were found between wave 1 and 2 panels; therefore, bias was not expected. Third, data collection occurred during the first COVID-19 wave, which may have influenced health perceptions and EQ-5D responsiveness. However, minimal change between waves suggests limited impact. Future research should examine long-term COVID-19 pandemic effects on HRQoL. Fourth, this study assumed similar psychometric properties across EQ-5D

translations; however, linguistic and cultural differences may affect responses. Appendix Table 6 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2025.09.009> shows variation in distributional characteristics across countries, but the patterns were consistent in direction. Therefore, pooling the data appeared appropriate and increased statistical power for comparing the 3-L and 5-L versions. In addition, we used the LSS, which offers a value set-independent summary score, advantageous for multicountry comparisons.³⁵ Future research will examine country-specific psychometric performance in greater detail, particularly for the 5-level version, which demonstrated superior properties overall. Fifth, because the POP-UP study population is not a diseased population, defining distinct known-groups was challenging. The number of chronic health conditions and the need for caregiver assistance may not serve as ideal known-groups in the classical psychometric sense, because they are heterogeneous and not defined conditions or clinical diagnoses. Nonetheless, we observed that the ceiling decreased after adding 3-level or 5-level bolt-ons for both known-groups and the LSS increased with more chronic conditions, suggesting some level of validity. Future research is needed with more appropriate known-groups to test these bolt-ons in general populations. Lastly, we assessed the psychometric performance of the EQ-5D and its bolt-ons, all with 5 response levels. Future research could explore whether fewer levels (eg, 4 levels) might be more appropriate for certain dimensions or contexts.⁶¹

Implications for Practice

Implementing bolt-ons in data collection is relatively straightforward, and the 5-level versions are recommended for their stronger psychometric performance and precision, particularly vision, sleep, tiredness, social relationships, and self-confidence in the general population. Importantly, bolt-ons may affect EQ-5D value sets and utility scores, raising concerns about comparability in economic evaluations.^{4,7} Further research is needed to evaluate how bolt-ons affect valuations and to develop guidelines for their use in economic evaluations of interventions targeting general populations.

Conclusions

The 6 EQ-5D-5L versions with 5-level bolt-ons outperformed the 3-L versions by reducing ceiling and floor effects, increasing informativity, showing divergent validity, and improving responsiveness. Further country- or language-specific analyses, particularly for the 5-level bolt-ons, which demonstrated superior properties overall, are needed to explore variation in psychometric performance and potential impacts on utility values and cost-utility results.

Author Disclosures

Author disclosure forms can be accessed below in the Supplemental Material section. Views expressed in the article are those of the authors and are not necessarily those of the EuroQol Research Foundation. Dr Mulhern is an editor for *Value in Health* and had no role in the peer-review process of this article.

Supplemental Material

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.jval.2025.09.009>.

Article and Author Information

Accepted for Publication: September 6, 2025

Published Online: xxxx

doi: <https://doi.org/10.1016/j.jval.2025.09.009>

Author Affiliations: Services in Health Economics (SHE), Brussels, Belgium (Dewilde); Section Medical Psychology and Psychotherapy, Department of Psychiatry, Erasmus MC, Rotterdam, The Netherlands (Janssen); argenX, Ghent, Belgium (Phillips); Centre for Health Economics Research and Evaluation, University of Technology Sydney, Sydney, Australia (Mulhern); Department of Health Policy, Corvinus University of Budapest, Budapest, Hungary (Rencz).

Correspondence: Sarah Dewilde, PhD, Services in Health Economics (SHE), Boulevard Lambertmont 418, 1030 Brussels, Belgium. Email: sd@she-consulting.be

Authorship Confirmation: All authors certify that they meet the ICMJE criteria for authorship.

Funding/Support: 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).

Role of the Funder/Sponsor: S.D., M.F.J., B.M., and F.R. are members of the EuroQol Group. The funder had no role in the design and conduct of the study; collection, management, analysis; preparation, review, of the manuscript; or decision to submit the manuscript for publication. G.P. had a role in the interpretation of the data and approval of the manuscript.

Acknowledgment: The authors thank all the participants of the study for sharing their HRQoL data.

REFERENCES

- 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 Mak*. 2015;35(3):276–291.
- EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199–208.
- 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.
- 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):3041–3048.
- 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.
- 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.
- 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.
- Finch AP, Mulhern B. Where do measures of health, social care and well-being fit within a wider measurement framework? Implications for the measurement of quality of life and the identification of bolt-ons. *Soc Sci Med*. 2022;313:115370.
- PROSPERO. Cognition bolt-ons for the EQ-5D-5L and EQ-5D-3L: a systematic review. National Institute for Health and Care Research. <https://www.crd.york.ac.uk/PROSPERO/view/CRD42023445567>. Accessed March 13, 2025.
- 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.
- 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.
- 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.
- 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.
- 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.

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.
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.
17. Dijkshoorn JN, Haagsma JA, van der Vlies CH, Hop MJ, van Baar ME, 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. *Eur Burn J.* 2022;3(2):264–277.
18. Szlávics 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.
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.
20. 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.
21. 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.
22. Xu RH, Rencz F, Sun R, Dong D, Zhang S. Development and testing of the psychometric properties of 20 bolt-on items for the EQ-5D-5L across 31 rare diseases. *Value Health.* 2025;28(5):769–780.
23. Sussex AK, Rencz F, Gaydon M, Lloyd A, Gallop K. Exploring the content validity of the EQ-5D-5L and four bolt-ons (skin irritation, self-confidence, sleep, social relationships) in atopic dermatitis and chronic urticaria. *Qual Life Res.* 2025;34(4):991–1002.
24. Müller EM, Nikl A, Krebs M, et al. Psychometric benefits of adding bolt-ons to the EQ-5D-5L in populations undergoing minimally invasive cosmetic procedures. *Eur J Health Econ.* 2025;26(7):1233–1247.
25. Angyal MM, Janssen MF, Lakatos PL, Brodsky V, Rencz F. The added value of the cognition, dining, gastrointestinal problems, sleep and tiredness bolt-on dimensions to the EQ-5D-5L in patients with coeliac disease. *Eur J Health Econ.* 2025;26(3):473–485.
26. Rencz F, Janssen MF. Testing the psychometric properties of 9 bolt-ons for the EQ-5D-5L in a general population sample. *Value Health.* 2024;27(7):943–954.
27. Bahrapour M, Kochovska S, Currow DC, Viney R, Mulhern B. Testing the psychometric characteristics of EQ-5D-5L and respiratory bolt-ons using a sample of the Australian population. *Qual Life Res.* 2024;34(2):395–403.
28. Dewilde S, Tollenaar NH, Phillips G, Paci S, Janssen MF. Observational cross-sectional study to estimate population norms in eight countries: the POPUP study protocol. *medRxiv.* Published online March 24, 2025, 2025.03.24.25324318.
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. *Arch Public Health.* 2025;83(1):184.
30. Wolfe GI, Herbelin L, Nations SP, Foster B, Bryan WW, Barohn RJ. Myasthenia gravis activities of daily living profile. *Neurology.* 1999;52(7):1487–1489.
31. Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems: health utilities index. *Pharmacoeconomics.* 1995;7(6):490–502.
32. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* 1983;67(6):361–370.
33. 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.
34. Devlin N, Parkin D, Janssen B. Analysis of EQ-5D profiles. In: Devlin N, Parkin D, Janssen B, eds. *Methods for Analysing and Reporting EQ-5D Data.* Cham, Switzerland: Springer International Publishing; 2020:23–49.
35. Feng YS, Jiang R, Pickard AS, Kohlmann T. Combining EQ-5D-5L items into a level summary score: demonstrating feasibility using non-parametric item response theory using an international dataset. *Qual Life Res.* 2022;31(1):11–23.
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.
37. 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 modeling and survey. *Health Technol Assess (Rockv).* 2014;18(9):1–224.
38. 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 Mak.* 2019;39(4):393–404.
39. Yang Y, Brazier J, Tsuchiya A. Effect of adding a sleep dimension to the EQ-5D descriptive system: a “bolt-on” experiment. *Med Decis Mak.* 2014;34(1):42–53.
40. 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.
41. Janssen MF, Pickard AS, Golicki D, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Qual Life Res.* 2013;22(7):1717–1727.
42. Janssen MF, Birnie E, Haagsma JA, Bonsel GJ. Comparing the standard EQ-5D three-level system with a five-level version. *Value Health.* 2008;11(2):275–284.
43. Janssen MF, Birnie E, Bonsel GJ. Evaluating the discriminatory power of EQ-5D, HUI2 and HUI3 in a US general population survey using Shannon’s indices. *Qual Life Res.* 2007;16(5):895–904.
44. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull.* 1959;56(2):81–105.
45. Feng YS, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. *Qual Life Res.* 2021;30(3):647–673.
46. Evans JD. Straightforward statistics for the behavioral sciences. https://books.google.com/books/about/Straightforward_Statistics_for_the_Behav.html?hl=nl&id=8Ca2AAAAIAAJ. Accessed March 13, 2025.
47. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol.* 2010;10(1):22.
48. Lu L, Shara N. Reliability analysis: calculate and compare intraclass correlation coefficients (ICC) in SAS. Department of Biostatistics and Epidemiology, the MedStar Research Institute; Published 2007. Accessed July 24, 2024. <https://lexjansen.com/nesug/nesug07/sa/sa13.pdf>.
49. Cicchetti DV. Guidelines, Criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess.* 1994;6(4):284–290.
50. Makovski TT, Schmitz S, Zeegers MP, Stranges S, van den Akker M. Multimorbidity and quality of life: systematic literature review and meta-analysis. *Ageing Res Rev.* 2019;53:100903.
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):140.
52. Li L, Liu C, Cai X, et al. Validity and reliability of the EQ-5D-5 L in family caregivers of leukemia patients. *BMC Cancer.* 2019;19(1):522.
53. 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.
54. Koszórú K, Hajdu K, Brodsky 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.
55. Golicki D, Niewada M, Buczek J, et al. Validity of EQ-5D-5L in stroke. *Qual Life Res.* 2015;24(4):845–850.
56. 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.
57. Bató A, Brodsky 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.
58. 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.
59. Poór AK, Rencz F, Brodsky 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.
60. 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.
61. Chen G, Olsen JA. Extending the EQ-5D: the case for a complementary set of 4 psycho-social dimensions. *Qual Life Res.* 2023;32(2):495–505.