



The psychometric performance of the EQ-5D-5L composite and component items in the U.S. General population and by age group

Minh Pham¹ · Benjamin M. Craig¹ · Fanni Rencz^{2,3}

Received: 22 June 2025 / Accepted: 12 August 2025
© The Author(s) 2025

Abstract

Objectives EQ-5D-5L items, self-care (SC), pain/discomfort (PD), and anxiety/depression (AD), are composites of two components. While PD (comprised of pain [PN] and discomfort [DI]) and AD (comprised of anxiety [AN] and depression [DE]) have been extensively studied, SC (comprised of washing self [WA] and dressing self [DR]) remains underexplored. Additionally, to our knowledge, the psychometric performance of composites versus components has never been investigated across age groups. This study examines the three composites and six components overall and by age group.

Methods In 2024, a representative sample of 6,018 US adults participated in an online survey that included the EQ-5D-5L and its six components in fixed order. In the overall sample and separately across age groups (18–29, 30–39, 40–49, 50–59, 60–69, 70+), we evaluated underreporting, ceilings and floors, agreement (weighted Kappa), convergent validity, informativity, explanatory power for EQ VAS (R^2), and structural validity (exploratory factor analysis [EFA]).

Results The ceiling was lower across all three composites compared to their component items. Strong agreement was found between composites and their corresponding component items (SC: 0.78, PD: 0.91, AD: 0.92). Using component items instead of composites increased the informativity and adjusted R^2 values. The first-listed components showed stronger convergent validity with their composites. EFA identified three factors: physical functioning (including MO, UA, SC), PD, and AD. Splitting the AD for younger and PD for older populations showed the greatest added value while there were minimal differences between SC and the module of washing and dressing self.

Conclusions The EQ-5D-5L composites effectively captured the information provided by their individual components, but replacing PD and AD with their components may improve instrument performance and its adaptability across health concerns relevant to different age groups.

Keywords EQ-5D · Composites · Population health · Psychometrics · Health-related quality of life

Background

The EQ-5D-5L is a widely adopted preference-weighted instrument for assessing health-related quality of life (HRQoL) in both clinical research and general population studies [1]. Initially developed to capture and value health

outcomes across diverse conditions, the EQ-5D-5L has become a standard instrument for economic evaluations and is now recommended by health technology assessment (HTA) agencies in multiple countries [1–4]. The instrument includes a descriptive system with five dimensions—mobility (MO), self-care (washing self/dressing self [SC]), usual activities (UA), pain/discomfort (PD), and anxiety/depression (AD)—each rated on a single item across five severity levels, and the EQ visual analogue scale (EQ VAS), which rates health status from 0 (worst health you can imagine) to 100 (best health you can imagine) [2, 3, 5, 6]. Despite its widespread adoption, the instrument has been the subject of ongoing scrutiny, particularly regarding how well its five items capture the complexity of patient experiences [7–10]. Central to this discussion is the assumption that respondents interpret and report on each of its five items in a consistent

✉ Benjamin M. Craig
bcraig@usf.edu

¹ Department of Economics, College of Arts and Sciences, University of South Florida, 4202 E Fowler Ave, Tampa, FL 33620, USA

² Department of Health Policy, Corvinus University of Budapest, Budapest, Hungary

³ EuroQoL Research Foundation, Rotterdam, The Netherlands

and reliable way [11]. However, this assumption becomes particularly challenged when dimension conflate two aspects, as is the case with the three dimensions: SC, PD, and AD. In this study, we referred these three dimensions as “composites” (i.e. broader constructs that encompass multiple underlying subitems) and the subitems that comprised them as “components” (i.e. individual items targeting specific aspects within each composite). It is important to note that while the UA item includes multiple examples in its question (“e.g., work, study, housework, family, or leisure activities”), we did not treat these as distinct components. Rather, they serve as illustrative examples to help guide respondents’ interpretation of the question, not as separate constructs to be measured individually.

By construction, composite labels introduce interpretive uncertainty (e.g., severely anxious or depressed): the adjectival statement does not specify whether respondents should report based on one symptom, both, or the more salient of the two, nor does it clarify how researchers or clinicians should interpret responses described by these terms, which may have vastly different actionable implications [12–14]. This ambiguity complicates both self-report and routine use, highlighting potential considerations about measurement validity and interpretive consistency. As such, there has been increasing interest in examining their underlying components to better understand response patterns and psychometric properties. Notably, the PD and AD have received extensive attention, with researchers using data from various countries and populations exploring their composite structure and potential for greater granularity.

McDonald et al. (UK English, $n=1007$ and $n=1415$) found that ceiling percentages were higher for composites than for their disaggregated components in both current health and worst recalled health assessments [11]. This suggested that respondents did not interpret composites as direct reflections of either individual component. Similarly, Tsuchiya et al. (UK English, $n=2494$) reported that respondents were more likely to select “no problems” on the composites than on both corresponding components when presented separately, with the difference more pronounced for AD than for PD, indicating that PD may be interpreted more consistently with its components [15]. In a Hungarian online survey ($n=1700$), while the differences were smaller between the ceiling percentages of the composites and components compared to other studies, Rencz and Janssen also observed underreporting on composites compared to components and identified systematic ordering effects, where respondents were more likely to respond based on the first-listed component (e.g., pain over discomfort) [16]. In a clinical interview study conducted in Amharic ($n=462$), Belay et al. found that 30.5% of respondents reported no problems on AD, whereas only 13.2% reported no problems

on both anxiety and depression, suggesting that those with moderate symptoms in just one domain may underreport problems on the composite [17]. Overall, despite variation in the number of respondents reporting “no problems” on the composites versus the components, all studies pointed to potential underreporting driven by ambiguity in the composite framing and highlighted the influence of the first-listed components.

Among the four studies, both McDonald et al. and Tsuchiya et al. randomized respondents to different survey versions, including the standard EQ-5D-5L and modified EQ-6D-5 L (i.e. composite was split and replaced with its two components) [11, 15]. Additionally, McDonald et al. also included EQ-4D-5 L+components versions (i.e. composite was split but only one component was retained) [11]. In contrast, Belay et al. and Rencz and Janssen administered both the composites and their corresponding components together in a single survey [16, 17].

Despite these advancements in PD and AD, SC has not been explored with equivalent depth. This composite captures activities such as washing self and dressing self, and consistently shows worse psychometric performance compared to the other EQ-5D-5L items in many populations [18–22]. The evidence on AD and PD leaves open the possibility that separating SC into distinct components, such as “washing self” and “dressing self”, could similarly improve the precision and relevance of the EQ-5D for capturing HRQoL in specific populations. Moreover, despite potential differences in interpretation and reporting, the performance of EQ-5D-5L composites and their components have never been systematically investigated across age groups.

The objective of this study is to describe the psychometric performance of the EQ-5D-5L composites and their components in a U.S. general population sample and by age subgroups: SC, PD, AD, washing self (WA), dressing self (DR), pain (PN), discomfort (DI), anxiety (AN), and depression (DE).

Methods

Data

After independent review board (IRB) at Advarra determined that this research project (Pro00080475; 11 July 2024) was exempt from IRB oversight based on the Department of Health and Human Services regulations found at 45 CFR 46.104(d)(2), we conducted an online cross-sectional study between August 27 and September 15, 2024 through an online marketing panel (Dynata). To align with the 2020 U.S. Census demographics, the recruitment strategy employed 18 quotas stratified by gender (female, male and

other), age (18–34, 35–54, 55+), and race/ethnicity (Hispanic, non-Hispanic Black, non-Hispanic other).

Among the 16,953 unique respondent identifiers representative of U.S. adults, 12,078 completed the screener, 8395 met the inclusion criteria (i.e., residing in one of the 50 U.S. states or Washington, D.C., and meeting the state-specific age of adulthood: ≥ 18 in most states, ≥ 19 in Nebraska and Alabama, and ≥ 20 in Mississippi), and 6467 completed the survey fully. Among those who completed the survey, 449 respondents were excluded because they completed it in less than 13 min to prevent low-effort responses, producing an analytical sample of 6,018 respondents. The survey instrument included the U.S. version of the EQ-5D-5L and the six components in fixed order [6]. Further details are available in the study protocol, including its survey instrument [23].

Statistical analyses

Building on recent studies by McDonald et al., Belay et al., and Rencz and Janssen that have examined the psychometric properties of EQ-5D-5L composites by comparing them to their components, we conducted a series of psychometric, exploratory factor, and regression analyses at the level of component, composite and module formed from the two components [11, 16, 17]. To examine age-related differences in performance, all analyses were repeated within each of six age-stratified subgroups: 18–29, 30–39, 40–49, 50–59, 60–69, and 70 years or older. Our null hypothesis is no difference in psychometric properties by age groups. All statistical analyses were conducted on R version 4.4.2 with packages dplyr version 1.1.4, vegan version 2.6-8, EFA.dimensions version 0.1.8.4, psych version 2.4.6.26 and ggplot2 version 3.3.0 [24–29].

Ceiling and floor

To provide an overview of the dataset and assess the range and potential clustering of responses, we summarized the response frequencies to assess potential underreporting, i.e., respondents reported fewer problems in composites than components. McNemar's test was then used to determine whether the discrepancies between composites and components were statistically significant. We then calculated the percentage of respondents selecting the best (ceiling) and worst (floor) response option for each composite and its components. These analyses were conducted at the individual level for both composites and components, and additionally at the instrument level by comparing the standard EQ-5D-5L profile to module variants: EQ-4D+ components (i.e. replacing each composite by one of its components), EQ-4D+ module (i.e. replacing each composite by both

components). Based on previous studies, we hypothesized that composites would exhibit underreporting and thus resulting in higher ceiling percentages and lower floor percentages compared to their components [16, 17, 30]. By construction, EQ-4D+ module instruments capture more problems than EQ-4D+ components.

Response agreement

To assess the level of agreement between each composite and the worst response of its two associated components, we conducted cross-tabulation between the components and their corresponding composites and calculated percent agreement and Cohen's weighted kappa (i.e. the composite score was compared to the maximum score obtained from their components) [31]. Kappa values were interpreted as follows: ≤ 0 (poor), 0.01–0.20 (slight), 0.21–0.40 (fair), 0.41–0.60 (moderate), 0.61–0.80 (substantial), and 0.81–1.00 (almost perfect) [32]. For the cross-tabulation analysis, we hypothesized that the proportion of respondents reporting some problems on the composite would be lower than on the components, and that response agreement would be moderate across the three dimensions, with higher agreement expected for the first-listed components (i.e., washing self, pain, and anxiety) [11, 17].

Informativity

To assess the degree of uniform distribution of responses across each item and how evenly the responses are distributed, we employed the Shannon Index (H') and the Shannon Evenness Index (J') for the composites, components, and modules. Originally proposed by Claude Shannon in 1948, the Shannon Index has been widely used in the study of the psychometric properties of the EQ-5D-5L [33, 34]. Similar to ceiling and floor, the indices were calculated at the individual level and at the instrument level. We hypothesized that H' and J' of the components would be higher than their composites with first-listed components being the highest [17]. At the instrument level, we hypothesized that EQ-4D-5 L+module instruments would have higher H' and J' compared to the original EQ-5D-5L and EQ-4D-5 L+components. The formulas used for calculating these indices are shown below, where L represents the number of response options for an item, and p_i is the proportion of respondents selecting the i th response option:

$$\text{Shannon Index } (H') : H' = - \sum_{\{i=1\}}^L p_i \log_2 p_i$$

$$\text{Shannon Evenness Index } (J') : J' = \frac{H'}{\log_2 L}$$

Convergent validity

To evaluate the relationship between composites, components, and EQ VAS, we estimated the Spearman's correlation coefficients. Following the guidelines provided by Cohen, correlations ($|r_s|$) between 0.1 and 0.29 were considered weak, those between 0.3 and 0.49 were considered moderate, and correlations of 0.5 or higher were regarded as strong [35]. Based on previous research, we hypothesized strong correlations between composites and their components, with the first-listed components expected to correlate more strongly, as well as strong associations between these items and EQ VAS [16, 30].

Structural validity

To determine if the composites and components effectively represent the same dimension or if the components provide additional, distinct information, we conducted an exploratory factor analysis (EFA). Factor loadings and eigenvalues were examined to identify underlying structures, with decisions on the number of factors guided by a combination of visual inspection of the scree plot and the Kaiser's criterion of retaining factors with eigenvalues greater than 1 [36, 37]. Rotation (e.g., varimax) was also applied to improve interpretability [38]. Factor loadings closer to -1 or $+1$ were interpreted as indicating strong associations between a composite/component and the underlying latent factor, while values near 0 suggested weak or no association. We hypothesized that composites and their components would load onto the same underlying factor.

Explanatory power

To investigate how components and modules contribute to explaining variation in EQ VAS and EQ-5D-5L health profile level-sum scores, we conducted bivariate and multivariate linear regression analyses [39]. These analyses examined the relationships between components (i.e., without interactions between components) and modules (i.e., interactions between components) and EQ VAS without the composites (SC, PD, and AD). Additionally, the effects of splitting the composites into components were also examined at the instrument level (i.e. EQ-4D-5 L+components and EQ-4D-5 L+modules). We hypothesized that including modules in the regression models would increase the explained variance (R^2) of EQ VAS beyond what is explained by the composites alone [40–44].

Results

Characteristics of the study population

Overall, the analytical sample of 6,018 participants was nationally representative and reflected the demographic and regional composition of the U.S. adult population. The analytical sample comprised 866 respondents (14.4%) aged 18–29, 1268 respondents (21%) aged 30–39, 1161 respondents (19.3%) aged 40–49, 997 respondents (16.6%) aged 50–59, 910 respondents (15.1%) aged 60–69, and 802 respondents (13.3%) aged 70 or older (Appendix 1).

Distributional characteristics

The response distribution of the overall sample on the individual level is presented in Fig. 1. For SC, 90.4% respondents reported Level 1, while its components, WA and DR, showed slightly higher ceiling percentages at 91.4% and 90.9%, respectively. Additionally, respondents reported extreme problems in SC more than in its components, with a higher percentage at Level 5 (0.3% versus 0.2% for both components). Our hypothesis of underreporting in composites was not supported, as no evidence of underreporting was found in PD and AD composites either. Their four components all showed higher percentages at Level 1 and lower percentages from Level 3 to Level 5. Similar patterns were observed in the analysis by age groups (Appendix 2).

Ceiling percentages were observed to be lower in composites compared to their corresponding components, which further rejected our hypothesis regarding underreporting in composites (Table 1). All increases in ceiling percentages from composites to components were statistically significant ($p < 0.001$), except for DR ($p = 0.08$) and PN ($p = 0.63$).

Figure 1 also illustrates the ceiling and floor percentages of the composites and their corresponding components on individual level by age groups. The ceiling percentages for PN and DI decreased from 53 to 31% and from 54 to 39%, respectively, with increasing age. In contrast, the ceiling percentages for AN and DE increased from 34 to 76% and from 46 to 83%, respectively. The ceiling percentages for WA and DR remained relatively stable.

At the instrument level, there were no statistically significant differences in ceiling percentages between the EQ-5D-5L (25.3%) and either EQ-4D-5 L+WA (25.3%) or EQ-4D-5 L+DR (25.3%) ($p = 1$ for both; Table 1). This remained the case when both WA and DR were used (EQ-4D-5 L+WA+DR: 25.2%). In contrast, ceiling percentages increased significantly with the addition of DI and DE (EQ-4D-5 L+DI: 27.7%, EQ-4D-5 L+DE: 27.9%), while significant decreases were observed for EQ-4D-5 L+PN+DI (24.1%) and EQ-4D-5 L+AN+DE (24.6%). In the age

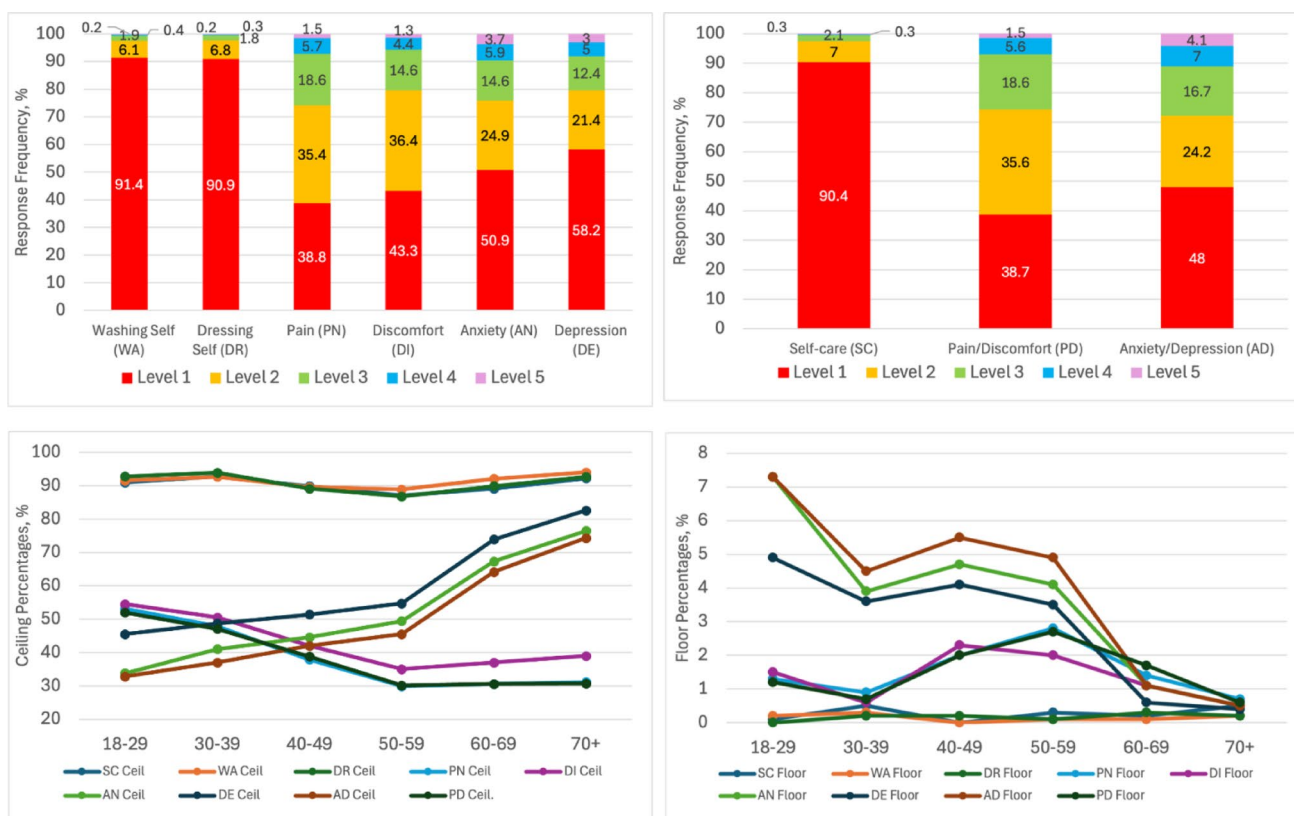


Fig. 1 EQ-5D-5L composites and components response distribution and ceiling and floor percentages by age group, on individual level, %

group analysis, ceiling percentages for EQ-4D-5 L+DI increased steadily with age, while EQ-4D-5 L+DE showed a decline beginning in the 30–39 age group, followed by a modest increase starting from the 50–59 age group (Fig. 2).

Response agreement

Table 2 presents the cross-tabulations, response agreement, and Cohen’s weighted kappa between composites and their components. Moderate to strong concordance was observed, with high diagonal proportions across Levels 1 to 5, except for Level 4 between washing self and self-care (26.9%), likely due to the small sample size ($n=19$). However, the hypothesis that fewer respondents would report problems on composites than on components was not supported. For example, 9.6% reported problems on SC, compared to 8.6% for WA and 9.1% for DR; similarly, PD was 61.3% vs. 61.2% (PN) and 56.8% (DI), and AD was 52.0% vs. 49.1% (AN) and 41.8% (DE). However, although composites had fewer “no problems” responses than their components overall, a small number of respondents who reported “no problems” on the composites subsequently reported problems on the components (WA: 1.6%, DR: 2.1%, PN: 7.2%, DI: 8.0%, AN: 5.3%, DE: 2.4%), suggesting little to no under-reporting. Our hypothesis regarding moderate agreement

with higher concordance for first-listed components was supported. Response agreement ranged from 75.7 to 94.5%, with higher values for WA (94.5%), PN (88.8%), and AN (81.6%) than their counterparts (DR: 94.3%, DI: 77.9%, DE: 75.7%). Weighted kappa values further confirmed this with substantial to almost perfect agreement: 0.78 (SC), 0.91 (PD), and 0.92 (AD).

Informativity

Absolute informativity (Shannon H') and relative informativity (Shannon J') are presented in Table 3. At the individual level, our hypothesis (i.e. components have higher H' and J' than their composites) was only supported for PN ($H' = 1.8383$, $J' = 0.7917$); all other components showed lower informativity than their corresponding composites. While PN and AN ($H' = 1.8182$, $J' = 0.7830$) demonstrated greater informativity than DI ($H' = 1.7403$, $J' = 0.7495$) and DE ($H' = 1.6711$, $J' = 0.7197$), this pattern did not hold for the SC module. At the instrument level, replacing composites with their corresponding modules resulted in higher absolute informativity compared to the original EQ-5D-5L and versions that included only one component. However, this improvement was not consistent in terms of relative informativity; only EQ-4D-5 L+AN+DE demonstrated

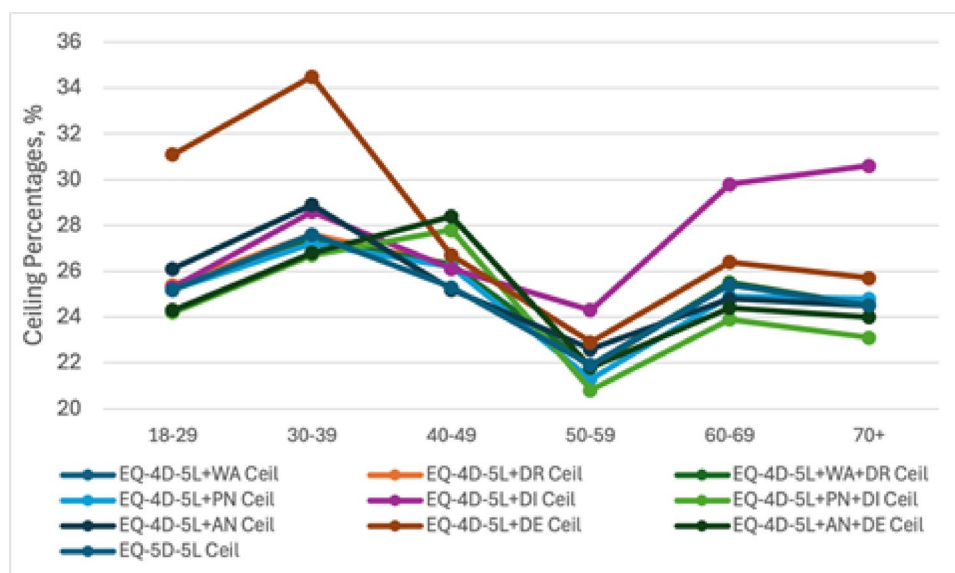
Table 1 Results of EQ-5D-5L ceiling and floor percentages at the individual level and instrument level in overall sample, % (n)

	EQ-5D-5L	Ceiling	Floor	Components	Ceiling	Floor	<i>p</i> -value [†]	
Individual level	MO	74.4 (4475)	0.5 (33)					
	SC	90.4 (5439)	0.3 (16)	WA	91.4 (5499)	0.2 (10)	<0.001	
				DR	90.9 (5468)	0.2 (11)	0.08	
	UA	74.2 (4466)	0.5 (29)					
	PD	38.7 (2327)	1.5 (89)	PN	38.8 (2337)	1.5 (92)	0.63	
				DI	43.3 (2603)	1.3 (81)	<0.001	
	AD	48 (2887)	4.1 (247)	AN	50.9 (3061)	3.7 (222)	<0.001	
				DE	58.2 (3501)	3.0 (179)	<0.001	
		Original instrument	Ceiling	Floor		Ceiling	Floor	<i>p</i> -value [‡]
	Instrument level	EQ-5D-5L	25.3 (1523)	0.0 (2)	EQ-4D-5L+ components			
				EQ-4D-5L+WA	25.3 (1522)	0 (1)	1	
				EQ-4D-5L+DR	25.3 (1524)	0 (1)	1	
				EQ-4D-5L+PN	25.0 (1506)	0 (2)	0.17	
				EQ-4D-5L+DI	27.7 (1667)	0 (2)	<0.001	
				EQ-4D-5L+AN	25.8 (1553)	0 (2)	0.03	
				EQ-4D-5L+DE	27.9 (1676)	0 (1)	<0.001	
				EQ-4D-5L+Modules				
				EQ-4D-5L+WA+DR	25.2 (1519)	0 (1)	0.39	
				EQ-4D-5L+PN+DI	24.1 (1452)	0 (1)	<0.001	
				EQ-4D-5L+AN+DE	24.6 (1478)	0 (1)	<0.001	

[†]: *p* values from McNemar's test for differences between components and composites

[‡]: *p* values from McNemar's test for differences between original instrument and EQ-4D-5L + components, EQ-4D-5L + modules

MO, mobility; SC, self-care; UA, usual activity; PD, pain/discomfort; AD, anxiety/depression; WA, washing self; DR, dressing self; PN, pain; DI, discomfort; AN, anxiety; DE, depression

Fig. 2 EQ-5D-5L ceiling percentages by age group, on instrument level, %

comparable relative informativity to EQ-5D-5L, while the other two modules showed lower values. Among the versions with only one component, only EQ-4D-5 L+PN yielded higher absolute informativity than EQ-5D-5L.

Age-specific analysis revealed that the first-listed components exhibited higher informativity than the second for all age groups and across all three modules, except for WA and DR in respondents aged 40 and above. Among the

Table 2 Results of distribution of responses of components by EQ-5D-5L composites, % (n)

Components		Composites					Problem reporting rate, %	Response agreement, %	Cohen's Kappa
		SC					SC=9.60		
		Level 1	Level 2	Level 3	Level 4	Level 5			
WA	Level 1	97.3 (5351)	2.3 (125)	0.3 (19)	0	0.1 (4)	8.62	94.5	0.78
	Level 2	18.6 (69)	70.5 (261)	9.5 (35)	0.8 (3)	0.5 (2)			
	Level 3	15 (17)	25.7 (29)	54.9 (62)	4.4 (5)	0			
	Level 4	7.7 (2)	19.2 (5)	30.8 (8)	26.9 (7)	15.4 (4)			
	Level 5	0	0	0	40 (4)	60 (6)			
DR	Level 1	97.4 (5326)	2.2 (120)	0.3 (18)	0 (1)	0.1 (3)	9.13	94.3	
	Level 2	23.1 (95)	65.5 (270)	10.4 (43)	1 (4)	0			
	Level 3	14.2 (15)	24.5 (26)	56.5 (60)	2.8 (3)	1.9 (2)			
	Level 4	4.8 (1)	14.3 (3)	14.3 (3)	47.6 (10)	19 (4)			
	Level 5	18.2 (2)	9.1 (1)	0	9.1 (1)	63.3 (7)			
		PD					PD=61.33		
		Level 1	Level 2	Level 3	Level 4	Level 5			
PN	Level 1	92.4 (2160)	6.9 (161)	0.6 (14)	0.1 (2)	0	61.17	88.8	0.91
	Level 2	6.5 (138)	88.5 (1884)	5 (106)	0.1 (2)	0			
	Level 3	2.1 (23)	8.1 (91)	85.3 (954)	4.4 (49)	0.1 (1)			
	Level 4	1.5 (5)	2.1 (7)	12.9 (44)	79.5 (271)	4.1 (14)			
	Level 5	1.1 (1)	0	3.3 (3)	15.2 (14)	80.4 (74)			
DI	Level 1	82.2 (2140)	16.2 (422)	1.5 (38)	0.1 (3)	0	56.75	77.9	
	Level 2	7.6 (167)	75 (1644)	16.9 (371)	0.4 (9)	0			
	Level 3	1.5 (13)	8 (70)	75.8 (667)	13.8 (121)	1 (9)			
	Level 4	2.3 (6)	1.9 (5)	15.6 (41)	69.6 (183)	10.6 (28)			
	Level 5	1.2 (1)	2.5 (2)	4.9 (4)	27.2 (22)	64.2 (52)			
		AD					AD=52.03		
		Level 1	Level 2	Level 3	Level 4	Level 5			
AN	Level 1	89.3 (2734)	8.2 (252)	2 (61)	0.3 (9)	0.2 (5)	49.14	81.6	0.92
	Level 2	8.8 (132)	74.9 (1108)	15 (225)	1.9 (29)	0.4 (6)			
	Level 3	2.1 (18)	10.1 (89)	73.5 (645)	11.5 (101)	2.8 (25)			
	Level 4	0.3 (1)	2.2 (8)	17.6 (63)	69.7 (249)	10.1 (36)			
	Level 5	0.9 (2)	0.9 (2)	4.5 (10)	14.9 (33)	78.8 (175)			
DE	Level 1	80.5 (2817)	15.2 (532)	3.4 (119)	0.8 (27)	0.2 (6)	41.82	75.7	
	Level 2	4.4 (57)	66.4 (857)	24.7 (319)	3.6 (46)	0.9 (11)			
	Level 3	1.1 (8)	8.1 (61)	70.1 (525)	17.1 (128)	3.6 (27)			
	Level 4	1.7 (5)	2.7 (8)	13 (39)	66.6 (199)	16.1 (48)			
	Level 5	0	0.6 (1)	1.1 (2)	11.7 (21)	86.6 (155)			

Level 1: No, Level 2: Slight, Level 3: Moderate, Level 4: Severe, Level 5: Extreme/Unable to.

MO, mobility; SC, self-care; UA, usual activity; PD, pain/discomfort; AD, anxiety/depression; WA, washing self; DR, dressing self; PN, pain; DI, discomfort; AN, anxiety; DE, depression.

EQ-4D-5 L+modules versions, only EQ-4D-5 L+WA+DR exhibited lower relative informativity than EQ-5D-5L across age groups (Appendix 3).

Convergent validity

Spearman's correlation coefficients indicated strong correlations between composites and their components, as well as between components within the same modules (Table 4). In all modules, the first-listed components showed stronger correlations with their respective composites than the

second-listed components. Among the components, WA and DR demonstrated weak to moderate correlations with the EQ VAS ($|r_s| = 0.29$ and $|r_s| = 0.31$, respectively), whereas the remaining components showed moderate to strong correlations. When analyzed by age groups, the correlation between SC and WA/DR showed a marked increase with age ($|r_s| = 0.65$ – 0.82 and $|r_s| = 0.64$ – 0.81 , respectively). In contrast, the correlations between AD and AN/DE, as well as between PD and PN/DI, remained relatively stable across age groups (Appendix 4).

Table 5 Results of EQ-5D-5L Factor loadings with and without components overall and by age groups

Factors	Whole sample		18–29		30–39		40–49	
	Without	With	Without	With	Without	With	Without	With
<i>Physical functioning</i>								
MO	0.73	0.67	0.61	0.67	0.72	0.70	0.72	0.58
SC	0.79	0.90	0.68	0.91	0.86	0.92	0.71	0.85
UA	0.72	0.70	0.72	0.68	0.70	0.68	0.68	0.57
WA	–	0.88	–	0.89	–	0.90	–	0.83
DR	–	0.87	–	0.88	–	0.89	–	0.80
<i>Pain and discomfort</i>								
PD	0.62	0.88	0.49	0.86	0.48	0.87	0.22	0.89
PN	–	0.89	–	0.87	–	0.88	–	0.89
DI	–	0.80	–	0.78	–	0.78	–	0.82
<i>Anxiety and depression</i>								
AD	0.61	0.94	0.63	0.94	0.66	0.93	0.58	0.92
AN	–	0.91	–	0.86	–	0.84	–	0.88
DE	–	0.88	–	0.84	–	0.84	–	0.85
<hr/>								
		50–59		60–69		70+		
	Without	With	Without	With	Without	With	Without	With
<i>Physical functioning</i>								
MO	0.81	0.64	0.74	0.66	0.71	0.71	0.71	0.71
SC	0.74	0.87	0.79	0.93	0.80	0.91	0.80	0.91
UA	0.73	0.67	0.87	0.76	0.64	0.70	0.64	0.70
WA	–	0.88	–	0.91	–	0.87	–	0.87
DR	–	0.92	–	0.88	–	0.86	–	0.86
<i>Pain and discomfort</i>								
PD	0.48	0.86	0.41	0.88	0.69	0.91	0.69	0.91
PN	–	0.88	–	0.89	–	0.91	–	0.91
DI	–	0.80	–	0.78	–	0.82	–	0.82
<i>Anxiety and depression</i>								
AD	0.64	0.91	0.63	0.95	0.47	0.95	0.47	0.95
AN	–	0.89	–	0.91	–	0.90	–	0.90
DE	–	0.88	–	0.90	–	0.88	–	0.88

MO, mobility; SC, self-care; UA, usual activity; PD, pain/discomfort; AD, anxiety/depression; WA, washing self; DR, dressing self; PN, pain; DI, discomfort; AN, anxiety; DE, depression

EQ-4D-5 L+DE showed slightly higher adjusted R^2 values than the original EQ-5D-5L (0.3956). Contrary to part of our hypothesis, adjusted R^2 values only modestly improved for the PD and AD modules when both components were included additively (i.e., components+components). However, including interaction terms between components (i.e., components \times components) resulted in higher adjusted R^2 values, with EQ-4D-5 L+WA \times DR outperforming the standard EQ-5D-5L at both the individual and instrument levels. Compared to the EQ-5D-5L, EQ-4D-5 L+AN \times DE exhibited the largest increase in adjusted R^2 value in the 18–29 age group, while EQ-4D-5 L+PN \times DI showed the largest increase in the 60–69 age group.

Discussion

This psychometric analysis evaluated the performance of EQ-5D-5L composites and their components in a general population sample of U.S. adults and by age group. We used a range of analytical methods to examine how well the composites captured the information provided by their respective components in terms of descriptive patterns and dimensional structure.

The first finding of our study concerned how composites performed relative to their components. At the individual level, response distributions and cross-tabulations showed that fewer respondents reported “no problems” and more reported some problems on the composites (SC, PD, and AD) than on their corresponding components. Ceiling and floor percentages confirmed that composites reflected a wide range of health problems than their individual components, which supports the interpretation of little to no underreporting. The patterns suggest that respondents

Table 6 Results of bivariate and multivariate regression analysis between EQ-5D-5L, components, modules and EQ VAS overall and by age groups

Selection of items	EQ VAS (Adjusted R ²)						
	Whole sample	18–29	30–39	40–49	50–59	60–69	70+
<i>EQ-5D-5L</i>							
MO	0.2063	0.1185	0.1368	0.1991	0.3007	0.3071	0.2750
SC	0.1391	0.0723	0.1124	0.1497	0.2160	0.1729	0.1303
UA	0.2465	0.1547	0.2066	0.2391	0.3252	0.2971	0.2846
PD	0.2723	0.2316	0.2092	0.2930	0.3182	0.3115	0.2933
AD	0.2052	0.2021	0.2235	0.2377	0.2995	0.1587	0.1491
<i>Components</i>							
WA	0.1109	0.0745	0.0979	0.1134	0.1642	0.1279	0.0891
DR	0.1228	0.5336	0.7515	0.1305	0.1885	0.1753	0.1216
PN	0.2776	0.2418	0.2043	0.2986	0.3218	0.3222	0.2948
DI	0.2921	0.2445	0.2299	0.3105	0.3322	0.3450	0.2894
AN	0.1856	0.1954	0.1989	0.2162	0.2733	0.1530	0.1318
DE	0.2017	0.2018	0.2105	0.2302	0.2812	0.1717	0.1445
<i>Modules</i>							
WA+DR	0.1344	0.0814	0.1064	0.1418	0.1939	0.1825	0.1253
PN+DI	0.3047	0.2717	0.2370	0.3235	0.3481	0.3626	0.3194
AN+DE	0.2169	0.2270	0.2405	0.2459	0.3074	0.1852	0.1656
WA×DR	0.1380	0.0896	0.1106	0.1480	0.1951	0.1896	0.1316
PN×DI	0.3060	0.2873	0.2419	0.3375	0.3578	0.3737	0.3313
AN×DE	0.2218	0.2499	0.2549	0.2699	0.3261	0.2021	0.1806
EQ-5D-5L	0.3956	0.3251	0.3474	0.4064	0.4909	0.4253	0.4161
<i>EQ-4D-5L + components</i>							
EQ-4D-5L+WA	0.3934	0.3246	0.3534	0.3977	0.4860	0.4317	0.4158
EQ-4D-5L+DR	0.3932	0.3260	0.3430	0.3966	0.4881	0.4318	0.4173
EQ-4D-5L+PN	0.3996	0.3280	0.3501	0.4068	0.4927	0.4361	0.4183
EQ-4D-5L+DI	0.3998	0.3291	0.3500	0.4060	0.4918	0.4429	0.4166
EQ-4D-5L+AN	0.3880	0.3215	0.3369	0.3967	0.4820	0.4311	0.4127
EQ-4D-5L+DE	0.3977	0.3265	0.3487	0.4057	0.4962	0.4378	0.4152
<i>EQ-4D-5L + modules</i>							
EQ-4D-5L+(WA+DR)	0.3944	0.3290	0.3552	0.3987	0.4888	0.4355	0.4202
EQ-4D-5L+(PN+DI)	0.4042	0.3413	0.3529	0.4110	0.4963	0.4505	0.4267
EQ-4D-5L+(AN+DE)	0.4024	0.3364	0.3606	0.4100	0.5007	0.4417	0.4207
EQ-4D-5L+(WA×DR)	0.4044	0.3758	0.3940	0.4378	0.5215	0.4666	0.4446
EQ-4D-5L+(PN×DI)	0.4142	0.3832	0.3926	0.4540	0.5351	0.4937	0.4497
EQ-4D-5L+(AN×DE)	0.4174	0.4084	0.4060	0.4361	0.5269	0.4838	0.4447

All reported regression models are statistically significant ($p < 0.001$, ANOVA F-test). The Cartesian product $A \times B$ implies $(a,b) | a \in A, b \in B$ (i.e. all combinations of levels from two component items, not the simple multiplication of item levels.)

MO, mobility; SC, self-care; UA, usual activity; PD, pain/discomfort; AD, anxiety/depression; WA, washing self; DR, dressing self; PN, pain; DI, discomfort; AN, anxiety; DE, depression

interpreted composites as a logical combination of their corresponding components (e.g., pain OR discomfort) rather than focusing solely on a single component. These findings contradict four previous studies that reported fewer problems on composites than on their components [11, 15–17]. This discrepancy may be due to methodological differences, such as the separate presentation of composites and components in McDonald et al. and Tsuchiya et al. [11, 15], the use of disease-specific populations in Belay et al. or the use of Hungarian language in Rencz and Janssen [16, 17]. At the instrument level, the standard EQ-5D-5L generally

outperformed the versions with only one component added (e.g., EQ-4D-5 L+WA, EQ-4D-5 L+DR), which further supports the idea that the composites might function better than individual components on their own. However, significant improvements were observed when both components of a composite were included (e.g., EQ-4D-5 L+PN+DI and EQ-4D-5 L+AN+DE). It is important to note that while composites outperformed individual components as stand-alone items, the combined inclusion of both components in a module enhanced the instrument's overall informativity and explanatory power. This suggests that composites

are meaningful and efficiently capture health problems, but disaggregated components may offer additional granularity when both are included. This interpretation is also supported by the lower correlations between the components in the same module compared to the correlations between each component and its composites, except for the PD module. These findings align with Belay et al., who reported increased informativity when replacing composites with their corresponding components [17].

On the other hand, our study supports the systematic order effect mentioned by Rencz and Janssen (i.e. respondents' answers to the composites appeared to be primarily influenced by the first-listed components) [16]. Cross-tabulations showed that PN and AN shared a higher proportion of responses with their respective composites, PD and AD, than DI and DE did. Convergent validity results also indicated stronger correlations between PN and PD, and between AN and AD, compared to DI and DE. A similar, though less consistent, pattern was observed for WA and DR. Taken together with our earlier finding that composites captured more reported problems than their corresponding components at the individual level, these results may reflect an ordering bias in the survey structure. Respondents completed the EQ-5D-5L composites before their split components, and the components were presented in a fixed order. Future research should examine whether this pattern holds when the order of components is randomized and when the components are presented between the core EQ-5D-5L dimensions.

Our findings offer practical implications for tailoring EQ-5D-5L applications to specific age groups. Ceiling percentages for PD, PN, and DI decreased with age, suggesting that older adults may benefit from expanded assessment of pain-related issues. In contrast, younger adults, particularly those aged 18–29, reported more mental health concerns, indicating greater relevance of the AD composite and its components in this population. Supporting these patterns, the inclusion of interaction terms further highlighted age-specific differences in explanatory power: the 60–69 age group experienced the largest increase in adjusted R^2 when $PN \times DI$ interactions were included, while the 18–29 age group showed the greatest increase in explanatory power when $AN \times DE$ interactions were added. Additionally, EFA showed that the underlying structure of the EQ-5D-5L remained stable with or without the component items, indicating that splitting the composites can enhance measurement precision without altering the instrument's dimensional integrity. However, as our study was based on a general population sample using an online survey, it is unclear whether these findings generalize to clinical populations. Future research should examine the performance and factor structure of the

composite and component items in disease-specific populations to assess the broader applicability.

Lastly, given the performance of WA and DR across our analyses, we did not observe substantial improvements in psychometric performance when analyzing them separately from SC. However, as this study was conducted in a general population sample, these findings may not generalize to clinical populations with higher prevalence of self-care problems (e.g., individuals with stroke, Parkinson's disease, or multiple sclerosis). Moreover, the correlation between WA and the SC composite was the lowest among the three modules, which suggests that respondents may distinguish between these aspects of self-care. Additionally, variations in how SC is interpreted across different language versions of the EQ-5D-5L may affect the functioning of its components. For example, the U.S. version uses the phrase "washing or dressing myself," while other translations differ in interpretation. In the Chinese version, the term "洗澡" is used, which means "bathing," whereas in the Vietnamese version, the term "giặt giũ" is used, which refers to "doing laundry." [45, 46] Such inconsistencies may limit the reliability and comparability of SC component items across settings.

There are limitations to our study that warrant consideration for future research. First, the use of an online survey may have introduced sampling bias, potentially underrepresenting individuals with limited internet access or lower digital literacy, thereby affecting the generalizability of the findings [47]. Future research is encouraged to combine online surveys with face-to-face interviews to improve the validity of the data. Second, during the data collection process, we employed quota sampling based on broader age bands (18–34, 35–54, and 55+) but later analyzed and reported results using narrower bands. This could limit the generalizability of age-specific findings and introduce sampling bias, especially if certain age subgroups are over- or underrepresented due to differential response rates. Third, we did not include utility analyses in this study, as our primary focus was on measurement properties and psychometric performance. However, given the widespread use of EQ-5D-5L in health technology assessment, future research should explore how the component- and composite-level data can be leveraged for greater precision when summarizing effectiveness.

Conclusions

This study found that the EQ-5D-5L composites generally captured the information provided by their components, with comparable or even broader coverage in many aspects. It was also noted that there were little to no statistically

significant differences observed between SC and its components (WA and DR). The overall construct of the EQ-5D-5L remained stable with or without components, suggesting robust dimensionality. Age group analyses highlighted the instrument's potential adaptability, with younger adults reporting more mental health issues and older adults more pain-related problems. Future research should assess the performance and structure of composite and component items in disease-specific populations and explore the effects of randomizing the order and placement of component items within the instrument.

Acknowledgements This work was presented at the 9th EuroQol Academy Meeting held March 11–13, 2025, in Barcelona, Catalonia, Spain. We thank Jan Abel Olsen for leading the discussion and the attendees for their insightful comments, particularly those on the measurement of activities for daily living (ADLs) in older adults.

Author contributions BMC was responsible for the original grant proposal, study design, instrument construction and online survey. MP, BMC, and FR developed the grant proposal for this secondary data analysis, developed the data analysis strategy, conducted the analysis, and drafted the manuscript.

Funding This study is funded by the EuroQol Research Foundation (464-RA[original study]; 1970-RA[secondary analysis]). We are also grateful to the Foundation for the travel grant for Minh Pham to attend the 5th EuroQol Early Career Researcher (ECR) Meeting 2025 and present this research at the 9th EuroQol Academy 2025.

Data availability Data will be made available upon reasonable request.

Declarations

Conflict of interest Benjamin M. Craig and Fanni Rencz are members of the EuroQol Group and its Descriptive Systems Working Group (DSWG). Fanni Rencz is employed by the EuroQol Research Foundation. Views expressed by the authors in the publication do not necessarily reflect the views of the EuroQol Research Foundation or the DSWG.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Kennedy-Martin, M., Slaap, B., Herdman, M., van Reenen, M., Kennedy-Martin, T., Greiner, W., et al. (2020). Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *European Journal of Health Economics*, *21*, 1245–1257.
- EuroQol—a new facility for the measurement of health-related quality of life. (1990). *Health Policy*, *16*, 199–208.
- Devlin, N. J., & Brooks, R. (2017). EQ-5D and the EuroQol group: Past, present and future. *Applied Health Economics and Health Policy*, *15*, 127–137.
- Rencz, F., Gulácsi, L., Drummond, M., Golicki, D., Prevolnik Rupel, V., Simon, J., et al. (2016). EQ-5D in central and eastern Europe: 2000–2015. *Quality of Life Research*, *25*, 2693–2710.
- Brooks, R. (1996). EuroQol: The current state of play. *Health Policy*, *37*, 53–72.
- Herdman, M., Gudex, C., Lloyd, A., Janssen, M. F., Kind, P., Parkin, D., et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, *20*, 1727–1736.
- Whalley, D., Globe, G., Crawford, R., Doward, L., Tafesse, E., Brazier, J., et al. (2018). Is the EQ-5D fit for purpose in asthma? Acceptability and content validity from the patient perspective. *Health and Quality of Life Outcomes*, *16*, 160.
- Lin, F.-J., Pickard, A. S., Krishnan, J. A., Joo, M. J., Au, D. H., Carson, S. S., et al. (2014). Measuring health-related quality of life in chronic obstructive pulmonary disease: Properties of the EQ-5D-5L and PROMIS-43 short form. *BMC Medical Research Methodology*, *14*, 78.
- Efthymiadou, O., Mossman, J., & Kanavos, P. (2019). Health related quality of life aspects not captured by EQ-5D-5L: Results from an international survey of patients. *Health Policy (Amsterdam)*, *123*, 159–165.
- Matza, L. S., Boye, K. S., Stewart, K. D., Curtis, B. H., Reaney, M., & Landrian, A. S. (2015). A qualitative examination of the content validity of the EQ-5D-5L in patients with type 2 diabetes. *Health and Quality of Life Outcomes*, *13*, 192.
- McDonald, R., Mullett, T. L., & Tsuchiya, A. (2020). Understanding the composite dimensions of the EQ-5D: An experimental approach. *Social Science & Medicine*, *265*, Article 113323.
- Peasgood, T., Mukuria, C., Carlton, J., Connell, J., & Brazier, J. (2021). Criteria for item selection for a preference-based measure for use in economic evaluation. *Quality of Life Research*, *30*, 1425–1432.
- Grant Levy, S. (2019). Deconstructing a double-barreled alternative: Evolution and creationism. *Psychological Reports*, *122*, 1995–2004.
- Menold, N. (2020). Double barreled questions: An analysis of the similarity of elements and effects on measurement quality. *Journal of Official Statistics*, *36*, 855–886.
- Tsuchiya, A., Bansback, N., Hole, A. R., & Mulhern, B. (2019). Manipulating the 5 dimensions of the EuroQol instrument: The effects on self-reporting actual health and valuing hypothetical health states. *Medical Decision Making*, *39*, 380–392.
- Rencz, F., & Janssen, M. F. (2022). Analyzing the pain/discomfort and anxiety/depression composite domains and the meaning of discomfort in the EQ-5D: A mixed-methods study. *Value in Health*, *25*, 2003–2016.
- Belay, Y. B., Mihalopoulos, C., Lee, Y. Y., Mulhern, B., & Engel, L. (2023). Examining the psychometric properties of a split version of the EQ-5D-5L anxiety/depression dimension in patients with anxiety and/or depression. *Quality of Life Research*, *32*, 2025–2036.

18. Feng, Y.-S., Kohlmann, T., Janssen, M. F., & Buchholz, I. (2021). Psychometric properties of the EQ-5D-5L: A systematic review of the literature. *Quality of Life Research*, *30*, 647–673.
19. Hinz, A., Kohlmann, T., Stöbel-Richter, Y., Zenger, M., & Brähler, E. (2014). The quality of life questionnaire EQ-5D-5L: Psychometric properties and normative values for the general German population. *Quality of Life Research*, *23*, 443–447.
20. Liu, X., Chan, W. W., Tang, E. H., Suen, A. H., Fung, M. M., Woo, Y. C., et al. (2023). Psychometric properties of EQ-5D-5L for use in patients with Graves' disease. *Health and Quality of Life Outcomes*, *21*, 90.
21. Rencz, F., Brodzky, V., & Janssen, M. F. (2023). A direct comparison of the measurement properties of EQ-5D-5L, PROMIS-29+2 and PROMIS Global Health Instruments and EQ-5D-5L and PROPr utilities in a general population sample. *Value in Health*, *26*, 1045–1056.
22. Konnopka, A., & Koenig, H.-H. (2017). The, “no problems”-problem: An empirical analysis of ceiling effects on the EQ-5D 5L. *Quality of Life Research*, *26*, 2079–2084.
23. Craig, B. M. (2025). Health valuation protocol for dual discrete choice experiment (dual-DCE) surveys to estimate the effects of different scenarios and attributes on main effects. 2025 [cited 2025 Mar 30]; Available from: <https://bmjopen.bmj.com/content/15/2/e091097>
24. R: The R Project for Statistical Computing [Internet]. [cited 2025 Jun 5]. Available from: <https://www.r-project.org/>
25. Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., Software, P., et al. (2023) dplyr: A Grammar of Data Manipulation [Internet]. [cited 2025 Jun 5]. Available from: <https://cran.r-project.org/web/packages/dplyr/index.html>
26. Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., et al. (2025) vegan: Community Ecology Package [Internet]. [cited 2025 Jun 5]. Available from: https://cran.r-project.org/web/packages/vegan/index.html?utm_source=chatgpt.com
27. O'Connor, B. P. (2024) EFA.dimensions: Exploratory Factor Analysis Functions for Assessing Dimensionality [Internet]. [cited 2025 Jun 5]. Available from: <https://cran.r-project.org/web/packages/EFA.dimensions/index.html>
28. Revelle, W. (2025) psych: Procedures for Psychological, Psychometric, and Personality Research [Internet]. [cited 2025 Jun 5]. Available from: <https://cran.r-project.org/web/packages/psych/index.html>
29. Create Elegant Data Visualisations Using the Grammar of Graphics [Internet]. [cited 2025 Jun 5]. Available from: <https://ggplot2.tidyverse.org/>
30. Engel, L., Whitehurst, D. G. T., Haagsma, J., Janssen, M. F., & Mulhern, B. (2023). What is measured by the composite, single-item pain/discomfort dimension of the EQ-5D-5L? An exploratory analysis. *Quality of Life Research*, *32*, 1175–1186.
31. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46.
32. McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*, 276–282.
33. Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423.
34. BasJanssen, M. F., Birnie, E., & Bonsel, G. J. (2007). Evaluating the discriminatory power of EQ-5D, HUI2 and HUI3 in a US general population survey using Shannon's indices. *Quality of Life Research*, *16*, 895–904.
35. Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.
36. Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Publications.
37. Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141–151.
38. Costello, A. B., Osborne, J. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. [cited 2025 May 29]; Available from: <https://openpublishing.library.umass.edu/pare/article/id/1650/>
39. Rencz, F., & Janssen, M. F. (2024). Testing the psychometric properties of 9 bolt-ons for the EQ-5D-5L in a general population sample. *Value in Health*, *27*, 943–954.
40. Balicki, P., Sołtysik, B. K., Borowiak, E., Kostka, T., & Kostka, J. (2025). Activities of daily living limitations in relation to the presence of pain in community-dwelling older adults. *Scientific Reports*, *15*, 15027.
41. Lay, K., Crocker, M., Engel, L., Ratcliffe, J., Charlton, S., & Hutchinson, C. (2023). How do older adults receiving aged care services understand and respond to the EQ-5D-5L? A think-aloud study in residential care. *Quality of Life Research*, *32*, 3161–3170.
42. Lopez, J., Perez-Rojo, G., Noriega, C., Sánchez-Cabaco, A., Sitges, E., & Bonete, B. (2024). Quality-of-life in older adults: Its association with emotional distress and psychological wellbeing. *BMC Geriatrics*, *24*, 815.
43. van Oppen, J. D., Conroy, S. P., Coats, T. J., Mackintosh, N. J., & Valderas, J. M. (2023). Measuring health-related quality of life of older people with frailty receiving acute care: Feasibility and psychometric performance of the EuroQol EQ-5D. *BMC Emergency Medicine*, *23*, 137.
44. Krawczyk-Suszek, M., & Kleinrok, A. (2022). Health-related quality of life (HRQoL) of people over 65 years of age. *International Journal of Environmental Research and Public Health*, *19*, 625.
45. Yang, F., Jiang, S., He, X. N., Li, H. C., Wu, H. Y., Zhang, T. T., et al. (2020). Do rural residents in China understand EQ-5D-5L as intended? Evidence from a qualitative study. *PharmacoEconomics*. <https://doi.org/10.1007/s41669-020-00212-z>
46. Mai, V. Q., Sun, S., Minh, H. V., Luo, N., Giang, K. B., Lindholm, L., et al. (2020). An EQ-5D-5L value set for Vietnam. *Quality of Life Research*, *29*, 1923–1933.
47. Bethlehem, J. (2010). Selection bias in web surveys. *International Statistical Review*, *78*, 161–188.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.