A Systematic Review and Head-to-Head Comparison of EQ-5D and PROPr: Psychometric Performance and Distributional Patterns

Xin ZHANG,[1] Annushiah VASAN THAKUMAR,[2] Nan LUO,[1] Ling Jie CHENG[3,4]

[1] Saw Swee Hock School of Public Health, National University of Singapore, Singapore [2] School of Pharmacy, Faculty of Health & Medical Sciences, Taylor's University, Malaysia
[3] National Perinatal Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, United Kingdom
[4] Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

Objectives

This systematic review compared the distributional patterns and measurement properties of EQ-5D-5L/3L and Patient-Reported Outcomes Measurement Information System, preference-weighted scoring system (PROPr) using published head-to-head comparison studies.

Methods

Following PRISMA-COSMIN guideline, we searched eight databases for English-language studies comparing EQ-5D-5L/3L and PROPr from 2018 (the first PROPr value set published) to March 2025. We extracted data on ceiling/floor effects, skewness, informativity, construct validity, reliability and responsiveness. Measurement quality was rated based on the proportion of estimates meeting acceptable thresholds for each measurement property. An estimate was defined as an independent assessment of a specific psychometric property.

Table 1 Ceiling and floor effect compared between EQ-5D and PROPr

	Full HS	Top 20 HS on the scale
Ceiling (%)		
EQ-5D range (%)	0.0 to 40.3	30.5 to 86.1
PROPr Range (%)	0.0 to 2.3	2.1 to 12.7
Relative range (EQ-5D minus PROPr, %)	0.0 to 38.0	28.4 to 73.4
S (E) reported ceiling effects (N)	10 (10)	2 (4)
S (E) reported lower in PROPr (N)	8 (8)	2 (4)
S (E) reported lower in EQ-5D, (N)	0 (0)	0 (0)
S (E) reported equal ceiling, (N)	2 (2)	0 (0)
Floor (%)		
EQ-5D range (%)	0.0 to 2.1	
PROPr Range (%)	0.0 to 41.8	
Relative range (EQ-5D minus PROPr, %)	-39.7 to 0.0	
S (E) reported ceiling effects (N)	5 (5)	2 (4)
S (E) reported lower in PROPr (N)	0 (0)	0 (0)
S (E) reported lower in EQ-5D, (N)	0 (0)	2 (4)
S (E) reported equal [unknow*] ceiling, (N)	4 (4) [1 (1)]	0 (0)

HS: health state, S: studies, E: estimates, * unknow to missing report on PROPr

Conclusion

- Both EQ-5D-5L/3L and PROPr demonstrated sufficient construct validity, with PROPr showing marginally better performance. The more normally distributed PROPr scores and lower ceiling effects are likely attributable to its broader domain coverage, though may increase respondent burden.
- Evidence on other properties such as test-retest reliability and responsiveness is limited. Further empirical studies are needed to evaluate these properties and generalizability beyond the US value set.

Results

- Twelve studies were included, from the following regions: North America (n=4), Europe (n=5), Oceania (n=1), and East Asia (n=1). Seven studies sourced general-population samples, five involved patients (cancer, rheumatological, psychosomatic conditions, low back pain, chronic skin conditions, spinal muscular atrophy, haemodialysis or kidney transplant).
- All studies used PROPr US value set (-0.022 to 1); Most studies (8) applied the US EQ-5D-5L value set which ranged from -0.573 to 1, other EQ-5D value sets included EQ-5D-5L China's, Canada's, 3L crosswalk to 5L and EQ-5D-3L US.

Distributional patterns (Table 1)

- Distributional patterns were contrasted between PROPr and EQ-5D. All studies (14 estimates) reported ceiling effects on index scores, 12 estimates reported lower for PROPr (0.0–12.7%) than EQ-5D-5L/3L (0.0–86.1%). Seven studies (nine estimates) reported floor effects, consistently lower or absent in EQ-5D-5L/3L (0.0–2.1%) compared to PROPr (0.0–41.8%).
- Four studies reported Pearson's coefficients of skewness for index scores, indicating near-normal distribution for PROPr (-0.1–1.33), while negative skew (-2.7 to -0.55) for EQ-5D-5L/3L.

Construct validity (Table 2)

- Six studies (14 estimates) including 5,847 self-reported cases assessed convergent validity, high certainty of evidence (93%) supporting sufficient convergent validity for both instruments.
- Ten studies (169 estimates) included 12,572 self-reported cases evaluated known-group validity across sociodemographic characteristics, self-perceived health status, and health-condition subgroups. High certainty of evidence supports sufficient known-group validity for both instruments, with PROPr (80%) slightly higher than EQ-5D (76%).

Other psychometric properties

- No studies assessed test–retest reliability or responsiveness, only one examined informativity

Table 2 Construct validity of EQ-5D and PROPr

EQ-5D (5L/3L)					PROPr				
N (S/E)	'+' rating (%)	N (cases)	Range of Statistics	Quality (COE)	N (S/E)	'+' rating (%)	N (cases)	Range of Statistics	Quality (COE)
Convergent									
6 (14)	13 (93)	5,847	0.18-0.79[A]	+ (H)	6 (14)	13 (93)	5,847	0.17-0.86[A]	+ (H)
Known-Group									
10 (169)	129 (76)	12,572	0.0 to 0.23[B]	+ (H)	10 (169)	135 (80)	14,643	0.001 to 0.28[B]	+ (H)
			0.001 to 1.2[C]					0.02 to 1.27[C]	
			0.55 to 0.68[D]					.0.52 to 0.66[D]	
			0.004 to 0.44[E]					0.001 to 0.43[E]	

S: study, E: estimate, A: Correlation coefficients, B: Regression Coefficients, C: Cohen's d, D: AUC, E: Mean difference