Construct Validity and Test-Retest Reliability of the Three-level and Five-level Versions of the EQ-5D-Y: A Systematic Review and Meta-regression of Head-to-head comparison studies

SA65

Ling Jie CHENG[1], Le Ann CHEN[1], Jing Ying CHENG[2], Michael HERDMAN[1], Nan LUO[1]

[1] Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore; [2] Khoo Teck Puat Hospital, Yishun Health, National Healthcare Group, Singapore

INTRODUCTION & AIM

- The three-level EQ-5D-Y (Y-3L) is a health-related quality of life instrument designed for children and adolescents.
- With the introduction of its five-level version (Y-5L), it is important to compare the performance of the two versions.
- An ongoing systematic review showed moderate construct validity and test-retest reliability for both the Y-3L and Y-5L.[1]
- Therefore, this review aims to explore the factors influencing the construct validity and test-retest reliability of Y-3L and Y-5L using results from published head-to-head studies.

METHOD

- Eight databases were searched for validation papers comparing the Y-3L and Y-5L, published in English up to February 14, 2024.
- The screening process involved identifying articles, screening titles and abstracts, and assessing full-text articles for eligibility, with two reviewers conducting this independently.
- The quality of measurement properties was rated as **"sufficient" (good)**, **"inconsistent" (moderate)**, or **"insufficient" (poor)** based on the proportion of studies meeting acceptable levels as per the COSMIN guideline.
- We chose to focus on seven factors that were commonly associated with quality ratings: (1) nature of population (general/patient), (2) geographical region (South Africa/ Southeast or East Asia), (3) age of the children (8-11/12-18), (4) survey language (English/Chinese/Bahasa Indonesian), (5) respondent (child/proxy), (6) interval period [*only for reliability*] (<2 weeks/≥2 weeks).
- Additionally, we examined the impact of these factors on construct validity using meta-regression.

RESULTS

- The review included 17 studies from Southeast/East Asia (N=9) and Africa (N=5), mostly cross-sectional (N=12) with consecutive sampling (N=16).
- High-certainty evidence consistently supports moderate construct validity for both versions (N=14).
- Despite few differences between the factors for both versions, the meta-regression, using 788 tests (Y-3L) and 866 tests (Y-5L), indicated that general populations (Y-3L: OR 1.83, p<0.001), adolescents aged 12-18 (Y-3L: OR 1.83; Y-5L: OR 3.07, p<0.05), and the English (Y-3L: OR 1.91; Y-5L: OR 1.72, p<0.05) and Chinese language versions (Y-3L: OR 3.29; Y-5L: OR 5.52, p<0.001) were associated with positive construct validity test results after adjusting for study clustering **(Table 1).**
- Low-certainty evidence supports moderate test-retest reliability for both Y-3L and Y-5L (N=11, 59 tests). Specifically, Y-3L showed better reliability than Y-5L in studies conducted in Africa and Oceania, as well as among child respondents **(Table 2).**

Table 1. Factors associated with construct validity of EQ-5D-Y

EQ-5D-Y	Y-3L				Y-5L			
	N (Articles)	'+' rating (%)	Adjusted OR ¥ (95% CI)	Quality (COE)	N (Articles)	'+' rating (%)	Adjusted OR ¥ (95% CI)	Quality (COE)
Overall	788 (14)	418 (53.0)		Moderate (H)	866 (14)	420 (48.5)		Moderate (H)
Nature of population								
General populations	319 (6)	207 (64.9)	1.83 (1.35, 2.49) ***	Moderate (H)	415 (6)	206 (49.6)	1.13 (0.84, 1.50)	Moderate (H)
Patient populations	469 (8)	211 (45.0)	Ref	Moderate (H)	451 (8)	214 (47.5)	Ref	Moderate (H)
Geographical regions								
Africa	494 (5)	236 (47.8)	Ref	Moderate (H)	496 (5)	248 (50.0)	Ref	Moderate (H)
South-east/ East Asia	222 (6)	110 (49.5)	1.05 (0.76, 1.44)	Moderate (H)	205 (6)	109 (53.2)	1.07 (0.77, 1.49)	Moderate (H)
Western Pacific	72 (3)	72 (100.0)	-		165 (3)	63 (38.2)	0.33 (0.19, 0.57) ***	Moderate (H)
Age of the Children								
<12 years	694 (11)	353 (50.9)	Ref	Moderate (H)	770 (11)	351 (45.6)	Ref	Moderate (H)
12-18 years	94 (3)	65 (69.1)	1.83 (1.13, 2.95) *	Moderate (H)	96 (3)	69 (71.9)	3.07 (1.92, 4.90) ***	Moderate (H)
Survey language								
English	566 (8)	308 (54.4)	1.91 (1.26, 2.91) **	Moderate (H)	661 (8)	311 (47.0)	1.72 (1.12, 2.65) *	Moderate (H)
Bahasa Indonesian	127 (2)	44 (34.6)	Ref	Moderate (H)	109 (2)	38 (34.9)	Ref	Moderate (H)
Chinese	95 (4)	66 (69.5)	3.29 (1.82, 5.95) ***	Moderate (H)	96 (4)	71 (74.0)	5.52 (3.00, 10.15) ***	Moderate (H)
Respondent								
Child	686 (11)	353 (51.5)	Ref	Moderate (H)	766 (11)	356 (46.5)	Ref	Moderate (H)
Proxy	102 (5)	65 (63.7)	1.03 (0.64, 1.65)	Moderate (H)	100 (5)	64 (64.0)	2.14 (1.38, 3.33) ***	Moderate (H)

Table 2. Factors associated with test-retest reliability of EQ-5D-Y

EQ-5D-Y	Y-3L			Y-5L		
	N (Articles)	'+' rating (%)	Quality (COE)	N (Articles)	'+' rating (%)	Quality (COE)
Overall	59 (11)	42 (71.2)	Moderate (L [b])	59 (11)	37 (62.7)	Moderate (L [b])
Nature of population						
General populations	20 (3)	13 (65.0)	Moderate (L [b])	20 (3)	9 (45.0)	Moderate (L [b])
Patient populations	39 (8)	29 (74.4)	Moderate (M [a])	39 (8)	28 (71.8)	Moderate (M [a])
Geographical region						
Africa	12 (1)	10 (83.3)	Good (L [b])	12 (1)	8 (66.7)	Moderate (L [b])
South-east/ East Asia	37 (7)	29 (78.4)	Good (M [a])	37 (7)	28 (75.7)	Good (M [a])
Western Pacific	10 (3)	3 (30.0)	Moderate (L [b])	10 (3)	1 (10.0)	Poor (L [b])
Age of the children						
<12 years	44 (8)	27 (61.4)	Moderate (L [b])	44 (8)	22 (50.0)	Moderate (L [b])
12-18 years	15 (3)	15 (100.0)	Good (M [a])	15 (3)	15 (100.0)	Good (M [a])
Survey language						
English	22 (4)	13 (59.1)	Moderate (L [b])	22 (4)	9 (40.9)	Moderate (L [b])
Chinese	31 (5)	23 (74.2)	Moderate (L [b])	31 (5)	22 (71.0)	Moderate (L [b])
Bahasa Indonesian	6 (2)	6 (100.0)	Good (H)	6 (2)	6 (100.0)	Good (H)
Interval period						
< 2 weeks	36 (4)	26 (72.2)	Moderate (L [b])	36 (4)	22 (61.1)	Moderate (L [b])
>= 2 weeks	23 (7)	16 (69.6)	Moderate (L [b])	23 (7)	15 (65.2)	Moderate (L [b])
Respondent						
Child	36 (8)	28 (77.8)	Good (L [b])	36 (8)	25 (69.4)	Moderate (L [b])
Proxy	23 (6)	14 (60.9)	Moderate (M [a])	23 (6)	12 (52.2)	Moderate (M [a])

COE: Certainty of evidence; CI: Confidence Interval; Quality of EQ-5D: + indicates 'sufficient' results; ± indicates 'inconsistent' results; – indicates 'insufficient' results; Certainty of evidence: H indicates 'high'; M indicates 'moderate'; L indicates 'low'; VL indicates 'very low'; [a] Quality downgraded by 1 level due to ROB; [b] Quality downgraded by 2 level due to ROB
[c] Quality downgraded by 1 level due to imprecision; [d] Quality downgraded by 2 levels due to imprecision

CONCLUSIONS

- Both versions of EQ-5D-Y showed similar results for construct validity, while Y-3L outperformed Y-5L in terms of test-retest reliability.
- This review identified high heterogeneity in the validity and reliability test results across populations, and highlighted the need for more high-quality, head-to-head validation studies of the two instruments.

REFERENCES

[1] Cheng LJ, Kreimeier S, Chen LA, Herdman M, Luo N. Head-to-head comparisons of the measurement properties of the three-level and five-level versions of the EQ-5D-Y: a systematic review. Oral presentation at the ISOQOL Annual Conference, 2023, Calgary, Canada & 41st EuroQol Plenary Meeting 2024, Rotterdam, the Netherlands.

CONTACT INFORMATION

Twitter: @JeremyChengLJ
Email Address: Sphclj@nus.edu.sg