

## **P105 A systematic review and head-to-head comparison of EQ-5D and PROPr: Psychometric performance and distributional patterns**

Xin Zhang<sup>1</sup>, Annushiah Vasan Thakumar<sup>2</sup>, Nan Luo<sup>1</sup>, Ling Jie Cheng<sup>3,1,4</sup>

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

**Objective:** This systematic review compared the distributional characteristics and measurement properties of the EQ-5D-5L/3L and Patient-Reported Outcomes Measurement Information System® (PROMIS®), 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 15, 2025. We extracted and evaluated 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.

**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 patient samples. Most studies (8/11) applied the US EQ-5D-5L value set, and all used the PROPr US value set. All studies (14 estimates) reported ceiling effects for index scores. Of these, 12 estimates showed lower ceiling for PROPr (0.0–12.7%) compared to EQ-5D-5L/3L (0.0–86.1%). Floor effects were reported in seven studies (nine estimates), 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 a negative skew (-2.7 to -0.55) for EQ-5D-5L/3L while PROPr showed near-normal distributions (-0.1–1.33). Six studies involving 5,847 self-reported cases assessed convergent validity with 14 estimates, demonstrating high-certainty evidence (93%) for both instruments. Ten studies involving 124,572 self-reported cases evaluated known-group validity across sociodemographic characteristics, self-perceived health status, and health-condition subgroups. Both instruments showed high-certainty evidence, with slightly higher ratings for PROPr (≥80%) compared to EQ-5D (≥76%).

**Conclusions:** 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 is limited: no studies assessed test-retest reliability or responsiveness, and only one examined informativity. Further empirical studies are needed to evaluate these properties and generalizability beyond the US value set.