



Measurement Properties of Commonly Used Generic Preference-Based Measures in East and South-East Asia: A Systematic Review

Xinyu Qian¹ · Rachel Lee-Yin Tan¹ · Ling-Hsiang Chuang² · Nan Luo¹

Published online: 25 November 2019
© The Author(s) 2019

Abstract

Objectives Our aim was to systematically review published evidence on the construct validity, test-retest reliability and responsiveness of generic preference-based measures (PBMs) used in East and South-East Asia.

Methods This systematic review was guided by the COSMIN guideline. A literature search on the MEDLINE, EMBASE, PsycINFO and PubMed databases up to August 2019 was conducted for measurement properties validation papers of the Euro-QoL-5 Dimensions (EQ-5D), Short Form-6 Dimensions (SF-6D), Health Utilities Index (HUI), Quality of Well-Being (QWB), 15-Dimensional (15D) and Assessment of Quality of Life (AQOL) in East and South-East Asian countries. Included papers were disaggregated into individual studies whose results and quality of design were rated separately. The population-specific measurement properties (construct validity, test-retest reliability and responsiveness) of each PBM were assessed separately using relevant studies. The overall methodological quality of the studies used in each of the assessments was also rated.

Results A total of 79 papers containing 1504 studies were included in this systematic review. The methodological quality was ‘very good’ or ‘adequate’ for the majority of the construct validity studies (99%) and responsiveness studies (61%), but for only a small portion of the test-retest reliability studies (23%). EQ-5D was most widely assessed and was found to have ‘sufficient’ construct validity and responsiveness in many populations, while the SF-6D and EuroQoL-Visual Analog Scale (EQ-VAS) exhibited ‘inconsistent’ construct validity in some populations. Scarce evidence was available on HUI and QWB, but current evidence supported the use of HUI.

Conclusions This systematic review provides a summary of the quality of existing generic PBMs in Asian populations. The current evidence supports the use of EQ-5D as the preferred choice when a generic PBM is needed, and continuous testing of all PBMs in the region.

1 Introduction

Preference-based measures (PBMs) provide a convenient approach to deriving health state values for the calculation of quality-adjusted life-years (QALYs) in cost-utility analysis

Xinyu Qian and Rachel Lee-Yin Tan contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40273-019-00854-w>) contains supplementary material, which is available to authorized users.

✉ Nan Luo
ephln@nus.edu.sg

¹ Saw Swee Hock School of Public Health, National University of Singapore, 12 Science Drive 2, Singapore 117549, Singapore

² Pharmerit International, Marten Meesweg 107, 3068 AV Rotterdam, The Netherlands

[1]. The use of a PBM starts with describing health status or health-related quality of life (HRQoL) of individuals using a standardized questionnaire. The HRQoL data can then be converted into health state values using a scoring method (also known as a ‘value set’). The value sets are established using the health preferences of the general public for the health states described by the PBMs. All PBMs use a scale anchored by 0 (corresponding to dead) and 1 (corresponding to full or perfect health), with or without negative values for very poor health states.

PBMs are usually developed for use in one population or culture, and subsequently introduced to other populations after translation or cultural adaptation. Since cultural, environmental, and psychosocial factors may affect the performance of PBMs, the measurement properties of PBMs should be validated in all populations and cultures to which they are introduced. Measurement properties that are relevant to all PBMs include construct validity, test-retest reliability, and responsiveness [2, 3].

Key Points for Decision Makers

Generic preference-based measures (PBMs) play an important role in health technology assessment in Asian countries.

The EuroQol-5 Dimensions (EQ-5D) has shown good construct validity and responsiveness in most countries and most disease groups in East and South-East Asia.

Future research should be expanded to rarely or never tested PBMs, such as the Health Utilities Index, Quality of Well-Being scale, and Assessment of Quality of Life instrument in this region.

In psychometrics, construct validity refers to the extent to which a scale measures what it is supposed to measure, test-retest reliability refers to the ability of a scale to generate reproducible measurement results, and responsiveness or sensitivity to change refers to the ability of a scale to capture the change in the levels of the targeted construct [3]. The testing of all three measurement properties involves collecting individual-level data using the scale, and performing statistical analyses. Construct validity is usually assessed through hypothesis testing because of the absence of a ‘gold standard’ measure [3]. Typically, the hypotheses are that a scale should be correlated with another scale measuring a similar construct (i.e. convergent validity) or that measurement results for groups known to differ in certain characteristics should be different (known-groups validity). The more hypotheses fulfilled, the more likely a scale is valid [3]. Test-retest reliability is assessed by examining the agreement between two different measurements of the same group of individuals whose levels in the targeted construct are the same at the times of the two measurements. Depending on the nature of the scale, statistics such as intraclass correlation coefficient (ICC) can be used as the indicator of test-retest reliability. Responsiveness assessment requires longitudinal data collection of individuals whose levels of the targeted construct change over time. Statistics that can be used to indicate responsiveness include standardized effect size (SES), standardized response mean [3], and receiver operating characteristic analysis [4].

Designed for use in a wide range of therapeutic areas, generic PBMs are particularly useful in economic evaluations informing resource allocations. In the past decades, generic PBMs such as EuroQol-5 Dimensions (EQ-5D) [5] and Short Form-6 Dimensions (SF-6D) [6] have been increasingly used in Asian countries and many validation studies assessing their measurement properties in Asian populations have been published. However, the overall

performance of PBMs in different countries or patient populations in this region is unknown. This is an important knowledge gap since cost-utility analysis is increasingly used to inform reimbursement decision making in Asia [7, 8].

The aim of this systematic review was to review and summarize the current evidence on the measurement properties of generic PBMs in Asian populations.

2 Methods

The COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) guideline for systematic reviews of outcome measurement instruments [4] was used to guide this review. Different from systematic review guidelines that are designed to evaluate interventional studies (e.g. the Cochrane guideline), the COSMIN guideline is specialized for evaluating measurement properties that are usually assessed in observational studies. It provides methods and tools for use in the entire process of systematic reviews, including literature search, selection and evaluation of studies, interpretation of results, and reporting of findings. In this review, two members of the review team worked independently through all phases of the review, and discrepancies were resolved via consensus meetings with the other two members of the review team. The four phases of the review process are described below.

2.1 Identification and Selection of Studies

The search was carried out using online databases, including MEDLINE (OvidSP), EMBASE (OvidSP), PsycINFO (OvidSP), and PubMed, in August 2019. Three groups of search terms were included to describe: (1) country/district, including countries/districts in South-East and East Asia: ‘China’, ‘Korea’, ‘Japan’, ‘Singapore’, ‘Taiwan’, ‘Hong Kong’, ‘Indonesia’, ‘Malaysia’, ‘Philippine’, ‘Thailand’ and ‘Vietnam’; (2) PBMs of interest, including ‘EQ-5D-3L’, ‘EQ-5D-5L’, ‘EQ-VAS’, ‘SF-6D’, ‘HUI2’, ‘HUI3’, ‘QWB’, ‘15D’, and ‘AQOL’; and (3) measurement properties, including ‘construct validity’, ‘test-retest reliability’ and ‘responsiveness’. All spelling variations, acronyms and related terms were included in the search algorithm (Appendix 1 of Supplementary file). The search filter developed by Terwee et al. [9] for the identification of reports on measurement properties of measurement instruments was adapted for use in this review. Although the EuroQol-Visual Analog Scale (EQ-VAS) is not a PBM, it was included as it is a part of EQ-5D.

A set of predefined selection criteria were applied to the hits that were generated by the search terms. Papers that examined the construct validity, test-retest reliability, and/

or responsiveness of any PBMs in any countries/districts of interest were included. Original research using primary data such as interventional and observational studies were included. Secondary research, including reviews, were excluded. Reports on mapping or reports published in a non-English language, as well as commentaries or conference papers (i.e. abstracts) were also excluded.

2.2 Data Extraction

The COSMIN guideline differentiates papers and studies [4]. Each hypothesis tested, ICC, or SES value reported for assessing construct validity, test-retest reliability, and responsiveness, respectively, is treated as one study. Therefore, a paper can include more than one study.

Information extracted from each study included PBM, sampling country or district, medical condition of study subjects, sample size, sample mean age, sample sex distribution, language of administration, and study design and result (see the following sections for more detail).

2.3 Assessment of Individual Studies

Each study was graded for its result and methodological quality using the methods prescribed in COSMIN [4]. The methods are briefly described below.

The result for *construct validity* was graded based on whether or not it was congruent with a relevant hypothesis formulated by the review team. COSMIN recommends systematic review teams to formulate a set of hypotheses for assessing known groups and convergent validity (including direction and magnitude of correlations) [4]. This is to ensure that results from all studies included in the review are interpreted using the same criteria. In this review, the review team formulated hypotheses based on published papers and on their expert experience. Example hypotheses were ‘patients with worse symptoms would have lower PBM scores’ (for testing known-groups validity) and ‘PBM and Health Assessment Questionnaire (HAQ) scores would be negatively and strongly correlated’ (for testing convergent validity). If the results of a study support the relevant hypothesis, a ‘positive’ rating is given, otherwise, a ‘negative’ rating is given.

Reported results on *test-retest reliability* (i.e. ICC value) were graded using 0.7 as the threshold [4]. A ‘positive’ rating was given if the ICC value was ≥ 0.70 , otherwise a ‘negative’ rating was given. Although area under the curve (AUC) is recommended for assessing responsiveness by COSMIN, the review team used SES because all studies assessing responsiveness included in this review reported either only SES or results that could be used to calculate SES; only one study reported AUC and SES. An SES value below 0.20 has been interpreted as negligible [3, 10]. The

review team assigned studies reporting an SES value < 0.20 a ‘negative’ rating, and those with an SES value ≥ 0.20 were assigned a ‘positive’ rating.

Using the ‘Risk of Bias’ assessment tool, the methodological quality of all studies was rated as ‘very good’, ‘adequate’, ‘doubtful’, or ‘inadequate’ [4]. Different standards were used to assess studies of convergent validity, known-groups validity, test-retest reliability, and responsiveness. These standards targeted various aspects of the design and execution of the studies. For example, measurement properties of the comparator instrument were targeted for assessing studies of convergent validity; characteristics of the comparison groups were targeted for assessing studies of known-groups validity; and stability of patients, time interval between test and retest, and similarity between test conditions were targeted for assessing studies of test-retest reliability. All assessments were made according to COSMIN recommendations, except for one of the standards for assessing convergent validity studies and the standards for assessing responsiveness studies (the modified standards used are shown in Appendices 2 and 3 of Supplementary file).

2.4 Assessment of the Preference-Based Measures (PBMs)

Since measurement properties may vary across populations, the review team assessed the measurement properties of each PBM in different populations separately. In this review, EQ-5D-3L and EQ-5D-5L were treated as one PBM (i.e. EQ-5D), Health Utilities Index (HUI) 2 and HUI3 as HUI, and SF-6Ds derived from SF-12, SF-36, and its descriptive system were not examined separately. For each PBM, different language versions or modes of administration (i.e. self- and interviewer-administered) were not examined separately. The populations were defined first by country/district and then by disease group. The disease groups were defined by the primary medical conditions of study samples included in this review using the International Classification of Diseases, 11th Revision (ICD-11) [11]. Studies on the general population were treated as one group.

For each PBM, separate assessments were performed using relevant studies to evaluate its population-specific measurement properties. Each of the assessments had two components—the measurement property and the quality of the evidence used in the assessment. The measurement property was rated as ‘sufficient’ (if at least 75% of the relevant studies had a ‘positive’ rating), ‘inconsistent’ (if 25–74% of the relevant studies had a ‘positive’ rating), or ‘insufficient’ (if $< 25\%$ of the relevant studies had a ‘positive’ rating) [4]. Using the COSMIN Grading of Recommendation Assessment, Development, and Evaluation (GRADE), the quality of evidence was rated as ‘high’, ‘moderate’, ‘low’, or ‘very low’. To determine the grade for quality of evidence, the

review team first assigned a rating of 'high' and then downgraded the rating based on the methodological quality of included studies (i.e. the 'Risk of Bias' factor) and the sample sizes of the studies (i.e. the 'Imprecision' factor). The review team did not apply the 'Inconsistency' and 'Indirectness' downgrading factors, as recommended by COSMIN [4]. In this review, inconsistency in the characteristics of the study samples was resolved by summarizing the results separately for different populations, and inconsistency in results was used to grade the quality of the PBMs. 'Indirectness' was not used as a downgrading factor because only studies of the populations of interest to the review team (i.e. populations from East and South-East Asia) were included (the modified GRADE criteria can be found in Appendices 4 and 5 of Supplementary file).

3 Results

The search initially identified a total of 1710 papers from four databases, which was reduced to 735 upon removal of duplicates, and further reduced to 114 after assessment of titles and abstracts. After assessment of full-text, 79 papers were retained for this systematic review [12–90]. A Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram for the selection process is shown in Fig. 1.

A total of 1504 individual studies were identified from the 79 retained papers. Table 1 shows the numbers of included papers and studies, organized by measurement property, PBM, and population. EQ-5D was the most studied PBM, construct validity was the most studied measurement property, Singapore and China produced the largest amount of papers, and the general population was the most studied. No relevant studies were found for Assessment of Quality of Life (AQOL), 15-Dimensional (15D) or Phillipines. A more detailed breakdown regarding the distribution of the papers can be found in Appendices 6 and 7 of Supplementary file.

Results were 'positive' in 80% of construct validity studies, 79% of test-retest reliability studies, and 57% of responsiveness studies. While 99% of the construct validity studies and 61% of the responsiveness studies were rated to have 'very good' or 'adequate' methodological quality, only a small portion of test-retest reliability studies (23%) achieved 'very good' or 'adequate' methodological quality.

A total of 729, 38, and 42 studies assessing construct validity, test-retest reliability, and responsiveness of EQ-5D, respectively, were identified. EQ-5D-3L was more commonly studied than EQ-5D-5L. For example, EQ-5D-3L had more than twice the number of studies reported for construct validity than EQ-5D-5L. The results for EQ-5D are summarized in Table 2. 'Sufficient' construct validity exhibits in 6 of 10 countries/districts and 17 of 20 disease groups

assessed; 'sufficient' test-retest reliability exhibits in none of 8 countries/districts and 3 of 10 disease groups assessed; and 'sufficient' responsiveness exhibits in 5 of 6 countries/districts and 8 of 11 disease groups assessed.

A total of 374, 15, and 16 studies assessing construct validity, test-retest reliability, and responsiveness of EQ-VAS, respectively, were identified. The results for EQ-VAS are summarized in Table 3. 'Sufficient' construct validity exhibits in 5 of 10 countries/districts and 8 of 14 disease groups assessed; 'sufficient' test-retest reliability exhibits in 4 of 6 countries/districts and 3 of 5 disease groups assessed; and 'sufficient' responsiveness exhibits in all of 4 countries/districts and 6 of 7 disease groups assessed.

A total of 179, 3, and 15 studies accessing construct validity, test-retest reliability, and responsiveness of SF-6D, respectively, were identified. The results for SF-6D are summarized in Table 4. 'Sufficient' construct validity exhibits in 2 of 5 countries/districts and 6 of 11 different disease groups assessed; 'sufficient' test-retest reliability exhibits in 1 (Hong Kong) of 2 countries/districts and 1 (thyroid) of 2 disease groups assessed; and 'sufficient' responsiveness exhibits in only one (South Korea) of 3 countries/districts and only 2 of 4 disease groups assessed.

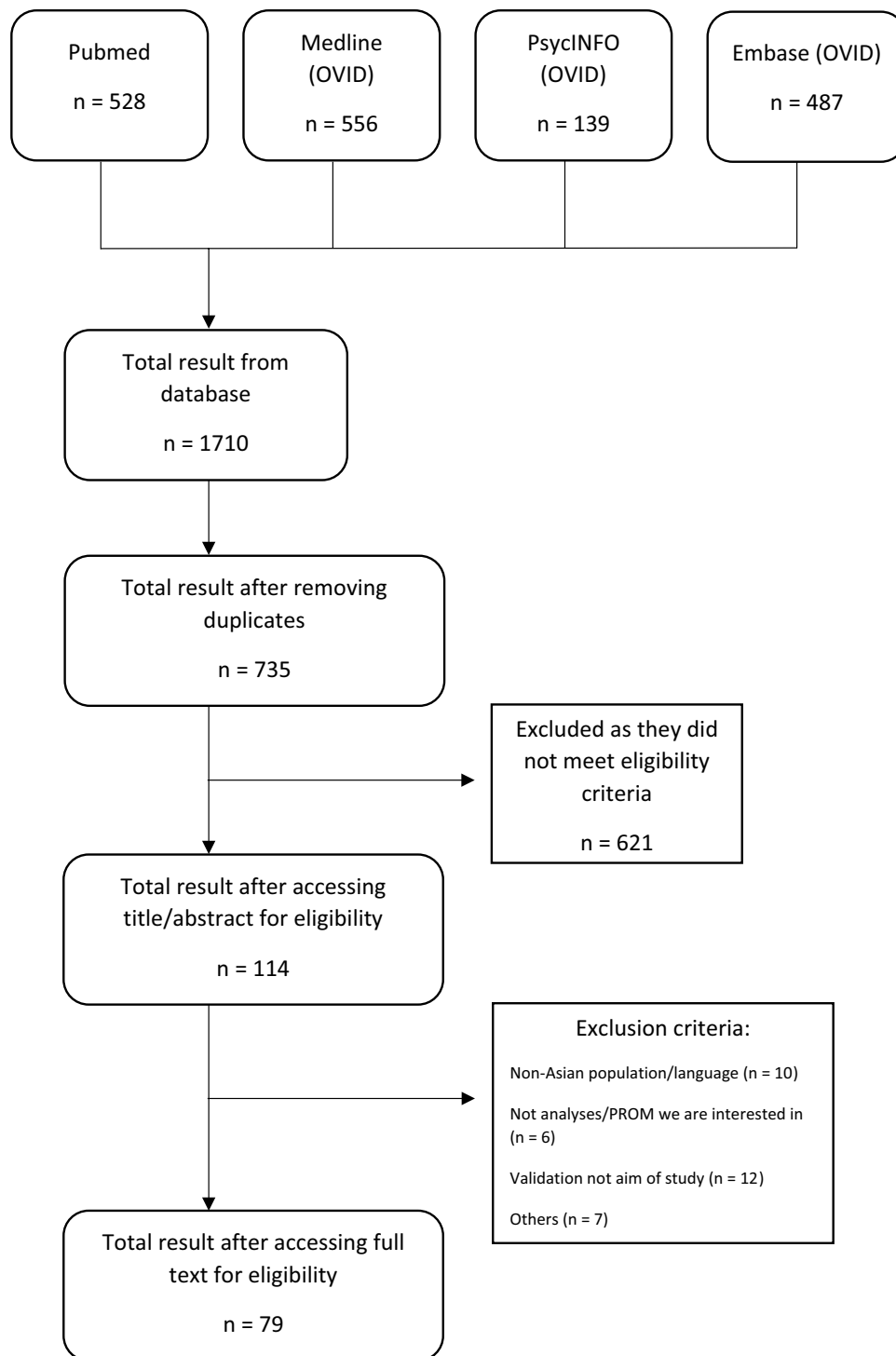
A total of 59, 5, and 7 studies assessing construct validity, test-retest reliability, and responsiveness of HUI, respectively, were identified. The results for HUI are summarized in Table 5. 'Sufficient' construct validity exhibits in all 3 countries/districts and 4 disease groups assessed; 'sufficient' reliability exhibits in 1 (Thailand) of 2 countries/districts and 2 of 3 disease groups assessed; and 'sufficient' responsiveness exhibits in 1 (Thailand) of 2 countries/districts and 2 of 3 disease groups assessed.

A total of 22 studies assessing the construct validity of the Quality of Well-Being (QWB) scale were identified. 'Sufficient' construct validity exhibits in both China and Japan and both neurological and respiratory disease groups.

4 Discussion

This systematic review targets the measurement properties of generic PBMs in East and South-East Asian countries. To the best of the review team's knowledge, this is the first systematic review of its kind. This review found that the generic PBMs that have been tested are EQ-5D, SF-6D, HUI (i.e. HUI2 and HUI3) and QWB, and that EQ-5D (i.e. EQ-5D-3L and EQ-5D-5L) might be the preferred choice when a generic PBM is needed in Asia. First, the evidence for EQ-5D is of the largest amount for all measurement properties and populations assessed. Second, it exhibited 'sufficient' construct validity and responsiveness in the largest number of populations, and 'insufficient' construct validity or responsiveness in none of the populations assessed.

Fig. 1 Chart for search results and selection of papers, *PROMs* patient-reported outcome measures



Satisfactory construct validity and responsiveness were also reported in past systematic reviews of EQ-5D in musculoskeletal [91], schizophrenia [92], skin [93], metabolic [94, 95], and respiratory diseases [96]. However, the current finding that EQ-5D is valid and responsive for patients with eye and heart diseases is at odds with the finding from a systematic review [95] that was mainly based on evidence

from European populations. The contradictory findings from the two systematic reviews suggest that the measurement properties of PBMs might vary from region to region. Therefore, it might be worthwhile to perform similar reviews for other regions to better inform the selection of PBMs for use in different populations.

Table 1 Included papers and studies, by category

| | No. of papers/studies |
|-------------------------|-----------------------|
| Measurement property | |
| Construct validity | 73/1363 |
| Test-retest reliability | 25/61 |
| Responsiveness | 16/80 |
| PBM | |
| EQ-5D-3L | 46/498 |
| EQ-5D-5L | 28/311 |
| EQ-VAS | 37/405 |
| SF-6D | 20/197 |
| HUI2 | 2/16 |
| HUI3 | 6/55 |
| QWB | 2/22 |
| Country/district | |
| China | 19/376 |
| Hong Kong | 10/177 |
| Japan | 5/38 |
| Malaysia | 4/21 |
| Singapore | 19/374 |
| South Korea | 7/159 |
| Taiwan | 6/184 |
| Thailand | 6/146 |
| Vietnam | 1/12 |
| Indonesia | 2/17 |
| Disease groups | |
| Cancer | 10/225 |
| Developmental disease | 1/14 |
| Diabetes | 5/56 |
| Eye disease | 3/32 |
| Gastric disease | 1/6 |
| General population | 17/302 |
| Genitourinary disease | 1/24 |
| Heart disease | 2/47 |
| Hepatitis | 2/31 |
| HIV | 3/39 |
| Injury | 1/60 |
| Kidney disease | 2/15 |
| Mental disorders | 3/65 |
| Multiple conditions | 3/130 |
| Musculoskeletal disease | 6/113 |
| Neurological disease | 3/78 |
| Respiratory disease | 3/32 |
| Rheumatic disease | 9/150 |
| Skin disease | 1/2 |
| Stroke | 3/71 |
| Thyroid disease | 1/12 |

PBM preference-based measures, *EQ-5D-3L* EuroQol-5 Dimensions, 3-Level Version, *EQ-5D-5L* EuroQol-5 Dimensions, 5-Level Version, *EQ-VAS* EuroQol-Visual Analog Scale, *SF-6D* Short Form-6 Dimensions, *HUI* Health Utilities Index, *QWB* Quality of Well-Being

The test-retest reliability of EQ-5D was found to be either ‘inconsistent’ or ‘insufficient’ for almost all populations, which is largely inconsistent with past systematic reviews [91, 94, 96]. The inferior test-retest reliability of EQ-5D revealed in this review could be related to suboptimal quality of evidence, which was attributable to the imperfect study design. In many studies included in this review, the ‘test’ was conducted when subjects visited a health institution, in the mode of face-to-face interview or self-completion, while the ‘retest’ was conducted over the telephone or via post when subjects were rested in their homes. The change in the data collection mode and setting from test to retest could have negatively affected the assessment result. Moreover, the test-retest reliability of EQ-5D could be underestimated due to the long duration used in those studies. Most studies included in this systematic review conducted the retest 1–2 weeks after the first test, as recommended [97]. While an interval of 1–2 weeks is appropriate for testing scales using a recall period of 1–4 weeks, it may be too long for EQ-5D because its recall period is only one day (‘today’). It is very possible that the health status of patients experiencing episodic symptoms in a particular day would change after 1 or 2 weeks, thus violating the assumption of unchanged health status needed for test-retest reliability testing, and leading to a worse test result.

The results for EQ-VAS are not entirely surprising because a visual analogue scale is not as easy to understand or use as verbal or categorical rating scales, where each response option is attached to an explanatory label [98]. It is possible that Asians, on average, have more difficulty with the EQ-VAS than Westerners because of their relatively lower education levels [99]. The suboptimal construct validity could also be caused by the vagueness of the labels used by EQ-VAS. In a qualitative study of Asians from Singapore [100], great variations in the interpretation of ‘best imaginable health’ were observed, which casts doubt on the comparability of EQ-VAS scores across individuals. However, a ‘sufficient’ result on responsiveness suggests that the EQ-VAS can be useful in evaluating individual-level change in HRQoL.

The suboptimal construct validity results for SF-6D are somewhat surprising. The descriptive system of SF-6D is more comprehensive than EQ-5D, and worldwide studies comparing SF-6D and EQ-5D found the two PBMs to have comparable measurement properties. One possible explanation can be due to elderly patients in Asia having a relatively lower literacy rate. According to UNESCO data [101], the elderly in European countries, such as Italy and Romania, have a literacy rate of > 85%. On the other hand, the literacy rate for the elderly in Asian countries, such as Thailand and Malaysia, is below 40%. The data collection for SF-6D is usually through SF-36, which contains 36 questions using relatively long sentence structures, which in turn might be

Table 2 Grading results for EQ-5D in different countries/districts and different disease groups

| | Quality of PBM, quality of evidence, and references | | |
|-------------------------|---|--|--------------------------------------|
| | Construct validity | Test-retest reliability | Responsiveness |
| China | + H [21, 27, 28, 41–43, 45, 57, 72, 75, 76, 82, 83, 85, 87, 89, 90] | ± H [27, 41, 75] | |
| Hong Kong | + H [17, 18, 20, 70, 79] | ± <i>L</i> ^b [20, 70, 79] | + <i>V</i> ^{b,c} [19] |
| Japan | ± <i>M</i> ^a [56, 66, 67] | | + <i>H</i> [56] |
| Malaysia | + H [53, 65, 71] | – <i>V</i> ^{b,c} [53] | |
| Singapore | + H [12–14, 38, 40, 47, 50, 68, 74, 77, 84, 86] | ± <i>L</i> ^b [12, 37, 38, 47] | ± <i>H</i> [12–14, 38, 51, 68] |
| South Korea | + H [29–32, 34, 39] | ± <i>L</i> ^b [29–32, 34] | + <i>H</i> [29, 33] |
| Taiwan | ± H [15, 16, 26, 36, 44, 88] | ± <i>M</i> ^a [15, 36] | + <i>M</i> ^a [16, 26, 44] |
| Thailand | + H [35, 58, 60–63] | ± <i>L</i> ^b [58, 63] | + <i>M</i> ^a [60, 63] |
| Vietnam | ± H [69] | | |
| Indonesia | ± <i>H</i> [59] | – <i>L</i> ^b [59] | |
| Cancer | + H [30, 32, 34, 36, 38, 42, 67] | ± <i>L</i> ^b [30, 32, 34, 36–38] | ± <i>M</i> ^c [38] |
| Diabetes | + H [39, 57, 58, 74] | ± <i>L</i> ^b [58] | |
| Eye disease | + H [13, 14] | | ± <i>H</i> [13, 24] |
| Gastric disease | + H [53] | – <i>V</i> ^{b,c} [53] | |
| General population | + H [15, 28, 31, 35, 41, 45, 59, 65, 66, 71, 75, 83, 85, 88, 89] | ± <i>M</i> ^a [15, 31, 41, 59, 75] | |
| Genitourinary disease | ± H [90] | | |
| Heart disease | + H [60, 82] | | + <i>M</i> ^c [60] |
| Hepatitis | + <i>L</i> ^b [27] | + <i>H</i> [27] | |
| HIV | + H [62, 69, 72] | | |
| Injury | + H [26] | | + <i>L</i> ^b [26] |
| Kidney disease | + H [86] | | |
| Mental disorders | ± H [12, 68] | ± <i>L</i> ^b [12] | ± <i>M</i> ^a [12, 68] |
| Multiple conditions | + H [61, 63, 77] | + <i>L</i> ^b [63] | + <i>L</i> ^b [63] |
| Musculoskeletal disease | + H [18, 20, 21, 87] | + <i>L</i> ^b [20] | + <i>V</i> ^{b,c} [19] |
| Neurological disease | + H [50] | | + <i>L</i> ^d [51] |
| Respiratory disease | + <i>M</i> ^a [17, 56] | | + <i>H</i> [56] |
| Rheumatic disease | + H [29, 40, 47, 70, 76, 84] | ± <i>L</i> ^b [29, 47, 70] | + <i>M</i> ^c [29] |
| Skin disease | + <i>H</i> [43] | | |
| Stroke | ± H [16, 44] | | + <i>H</i> [16, 33, 44] |
| Thyroid disease | + H [79] | – <i>L</i> ^b [79] | |

Quality of PBM: + indicates sufficient results; ± indicates inconsistent results; – indicates insufficient results

Italicised font indicates that grading is based on no more than three studies

Quality of evidence: H indicates high; M indicates moderate; L indicates low; V indicates very low

EQ-5D EuroQol-5 Dimensions, *PBM* preference-based measure, *ROB* risk of bias

^aQuality downgraded by 1 level due to ROB

^bQuality downgraded by 2 levels due to ROB

^cQuality downgraded by 1 level due to imprecision

^dQuality downgraded by 2 levels due to imprecision

difficult for some respondents with a lower literacy level [99].

This study provides some directions for future research on generic PBMs in Asia. First, future research should be expanded to rarely or never tested PBMs such as HUI, QWB, and AQOL. HUI (i.e. HUI2 and HUI3) is especially worth more research since ‘sufficient’ support has been shown for

most measurement properties in all populations assessed. Second, researchers are strongly recommended to use a better design in future studies of test-retest reliability and responsiveness, such as using the same data collection mode in all time points. Last, studies should be conducted to ascertain the reasons for the suboptimal construct validity of the

Table 3 Measurement properties of EQ-VAS in different countries/districts and disease groups

| | Quality of PBM and evidence | | | | | | | | |
|-------------------------|-----------------------------|----------------|--|-------------------------|----------------|------------------|----------------|----------------|--------------|
| | Construct validity | | | Test-retest reliability | | | Responsiveness | | |
| China | + | H | [21, 41, 72, 73, 75, 85] | + | M ^a | [21, 41, 73, 75] | | | |
| Hong Kong | + | H | [17, 18, 70] | | | | | | |
| Japan | ± | M ^c | [25] | | | | | | |
| Malaysia | ± | H | [22, 53, 65] | | | | | | |
| Singapore | ± | H | [13, 38, 46, 48–50, 77] | + | L ^b | [38] | + | H | [13, 38, 51] |
| South Korea | ± | H | [29–32] | – | L ^b | [29, 30, 32] | + | M ^c | [29] |
| Taiwan | ± | H | [15, 16, 26, 36, 88] | + | H | [15] | + | L ^b | [16, 26] |
| Thailand | + | H | [35, 61–63] | + | L ^b | [63] | + | L ^b | [63] |
| Vietnam | + | H | [69] | | | | | | |
| Indonesia | + | H | [59, 64] | ± | M ^a | [59, 64] | | | |
| Cancer | ± | H | [30, 32, 36, 38, 64] | ± | L ^b | [30, 32, 38, 64] | + | M ^c | [38] |
| Diabetes | ± | H | [46] | | | | | | |
| Eye disease | ± | H | [13] | | | | – | H | [13] |
| Gastric disease | + | H | [53] | | | | | | |
| General population | + | H | [15, 25, 31, 35, 41, 59, 65, 75, 85, 88] | + | M ^a | [15, 41, 59, 75] | | | |
| HIV | + | H | [62, 69, 72] | | | | | | |
| Injury | ± | H | [26] | | | | + | L ^b | [26] |
| Kidney disease | + | H | [22] | | | | | | |
| Multiple conditions | + | H | [61, 63, 77] | + | L ^b | [63] | + | L ^b | [63] |
| Musculoskeletal disease | + | H | [18, 21] | – | M ^c | [21] | | | |
| Neurological disease | ± | H | [50] | | | | + | L ^d | [51] |
| Respiratory disease | + | H | [17] | | | | | | |
| Rheumatic disease | + | H | [29, 48, 49, 70, 73] | + | L ^b | [29, 73] | + | M ^c | [29] |
| Stroke | – | H | [16] | | | | + | M ^c | [16] |

Quality of PBM: + indicates sufficient results; ± indicates inconsistent results; – indicates insufficient results

Quality of evidence: H indicates high; M indicates moderate; L indicates low

Italicised font indicates that grading is based on no more than three studies

EQ-VAS EuroQol-Visual Analog Scale, PBM preference-based measure, ROB risk of bias

^aQuality downgraded by 1 level due to ROB

^bQuality downgraded by 2 levels due to ROB

^cQuality downgraded by 1 level due to imprecision

^dQuality downgraded by 2 levels due to imprecision

SF-6D and EQ-VAS, and to explore ways to improve their performance in Asian populations.

This study has three limitations. First, since some of the COSMIN methods and tools do not apply to a systematic review of multiple measures in multiple populations, it was necessary for the review team to modify the original methods. Due to these modifications, it may not be meaningful to compare the results from this review with those from other reviews that applied the original COSMIN methods. These modifications, however, are unlikely to favour any of the PBMs included in this study. The second limitation is the exclusion of papers published in non-English journals due to limited manpower and resources. There are databases

in the Chinese, Japanese, and Korean languages that could include validation studies of PBMs. Therefore, the results of this review might not truly reflect the performance of the generic PBMs in China, Japan, and South Korea. Third, since different language versions were not differentiated, results from this review for Singapore and Malaysia might not be accurate for all language versions of the studied instruments. Despite the effort that has been put into translation, psychometric equivalence between source and target languages might not necessarily occur [102]. Nevertheless, studies have shown measurement equivalence between different language versions of EQ-5D and SF-6D in Singapore [103–106].

Table 4 Measurement properties of SF-6D in different countries/districts and different disease groups

| | Quality of PBM and evidence | | | | | | | | |
|-------------------------|-----------------------------|----------|--------------------------|-------------------------|-------------------------|------|----------------|-----------------------|----------|
| | Construct validity | | | Test-retest reliability | | | Responsiveness | | |
| China | ± | H | [28, 42, 82, 87, 89, 90] | | | | | | |
| Hong Kong | + | H | [17, 18, 78, 80] | + | <i>L</i> ^a | [79] | ± | <i>L</i> ^a | [81] |
| Japan | – | <i>H</i> | [66] | | | | | | |
| Singapore | + | H | [12, 40, 84, 86] | ± | <i>V</i> ^{a,b} | [12] | ± | <i>L</i> ^a | [12, 24] |
| South Korea | | | | | | | + | H | [33] |
| Thailand | ± | H | [61] | | | | | | |
| Cancer | + | H | [73] | | | | ± | <i>L</i> ^a | [81] |
| Eye disease | | | | | | | + | <i>L</i> ^a | [24] |
| General population | ± | H | [28, 66, 80, 89] | | | | | | |
| Genitourinary disease | ± | H | [90] | | | | | | |
| Heart disease | ± | H | [82] | | | | | | |
| Hepatitis | ± | H | [78] | | | | | | |
| Kidney disease | + | H | [86] | | | | | | |
| Mental disorders | + | H | [12] | ± | <i>V</i> ^{a,b} | [12] | ± | <i>L</i> ^a | [12] |
| Multiple conditions | ± | H | [61] | | | | | | |
| Musculoskeletal disease | + | H | [18, 87] | | | | | | |
| Respiratory disease | + | H | [17] | | | | | | |
| Rheumatic disease | + | H | [40, 84] | | | | | | |
| Stroke | | | | | | | + | H | [33] |
| Thyroid disease | | | | + | <i>L</i> ^a | [79] | | | |

Quality of PBM: + indicates sufficient results; ± indicates inconsistent results; – indicates insufficient results

Italicised font indicates that grading is based on no more than three studies

SF-6D Short Form-6 Dimensions, PBM preference-based measure, ROB risk of bias

Quality of evidence: H indicates high; L indicates low; V indicates very low

^aQuality downgraded by 2 levels due to ROB

^bQuality downgraded by 1 level due to imprecision

Table 5 Measurement properties of HUI in different countries/districts and different disease groups

| | Quality of PBM and evidence | | | | | | | | |
|-----------------------|-----------------------------|---|--------------|-------------------------|-------------------------|----------|----------------|-----------------------|----------|
| | Construct validity | | | Test-retest reliability | | | Responsiveness | | |
| Hong Kong | + | H | [54] | | | | | | |
| Singapore | + | H | [12, 47, 52] | ± | <i>L</i> ^b | [12, 47] | ± | <i>M</i> ^a | [12, 24] |
| Thailand | + | H | [60] | + | <i>H</i> | [60] | + | H | [60] |
| Developmental disease | + | H | [54] | | | | | | |
| Eye disease | | | | | | | + | <i>H</i> | [24] |
| Heart disease | + | H | [60] | + | <i>H</i> | [60] | + | H | [60] |
| Mental disorders | + | H | [12, 52] | ± | <i>V</i> ^{b,c} | [12] | ± | <i>L</i> ^b | [12] |
| Rheumatic disease | + | H | [47] | + | <i>V</i> ^{b,c} | [47] | | | |

Quality of PBM: + indicates sufficient results; ± indicates inconsistent results; – indicates insufficient results

Quality of evidence: H indicates high; M indicates moderate; L indicates low; V indicates very low

Italicised font indicates that grading is based on no more than three studies

HUI Health Utilities Index, PBM preference-based measure, ROB risk of bias

^aQuality downgraded by 1 level due to ROB

^bQuality downgraded by 2 levels due to ROB

^cQuality downgraded by 1 level due to imprecision

5 Conclusions

This systematic review provides a summary of the quality of existing generic PBMs in Asian populations from different countries and different disease groups. The current evidence supports the use of EQ-5D as the preferred choice, when a generic PBM is needed, and the continuous testing of all PBMs in the region.

Author Contributions As co-first authors, XQ and RL-YT were involved in data analysis, reporting of results, and drafting of the manuscript. L-HC was involved in critically reviewing the manuscript for important intellectual content, while NL, as the corresponding author, was responsible for conceptualizing and overseeing the project to ensure its quality and integrity.

Funding This study was funded by the EuroQol Research Foundation (Grant no. EQ Project 2016230).

Data Availability The datasets generated during the current study are available from the corresponding author on reasonable request.

Compliance with Ethical Standards

Conflicts of interest Xinyu Qian and Rachel Lee-Yin Tan have no conflicts of interest to declare. Nan Luo and Ling-Hsiang Chuang are members of the EuroQol Group.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. *Annu Rev Public Health*. 2000;21:587–611.
2. 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: a systematic review, statistical modelling and survey. Southampton: NIHR Journals Library. <https://www.journalslibrary.nihr.ac.uk/hta/hta18090/#/abstract>. Accessed Feb 2014
3. Fayers PM, Machin D. Quality of life: the assessment, analysis and interpretation of patient-reported outcomes. 2nd ed. Chichester: Wiley; 2007.
4. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs). https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf. Accessed 1 Jul 2018.
5. The EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16:199–208.
6. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ*. 2002;21:271–92.
7. Thorat T, Lin P, Neumann PJ. The state of cost-utility analyses in Asia: a systematic review. *Value Health Reg Issues*. 2015;6:7–13.
8. Butt T, Liu GG, Kim DD, Neumann PJ. Taking stock of cost-effectiveness analysis of healthcare in China. *BMJ Glob Health*. 2019;4:e001418.
9. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measure instruments. *Qual Life Res*. 2009;18:1115–23.
10. Cohen J. Statistical power analysis for the behavioural sciences. 2nd ed. Hillsdale: Lawrence Earlbaum Associates; 1988.
11. International Classification of Diseases 11th Revision (ICD 11). <https://icd.who.int/en/>. Accessed 16 Aug 2019.
12. Abidin E, Chong SA, Seow E, et al. A comparison of the reliability and validity of SF-6D, EQ-5D and HUI3 utility measures in patients with schizophrenia and patients with depression in Singapore. *Psychiatry Res*. 2019;274:400–8.
13. Ang M, Fenwick E, Wong TY, Lamoureux E, Luo N. Utility of EQ-5D to assess patients undergoing cataract surgery. *Optomet Vis Sci*. 2013;90(8):861–6.
14. Au EKG, Chan EW, Luo N, et al. Validity of EuroQOL-5D, time trade-off, and standard gamble for age-related macular degeneration in the Singapore population. *Eye (London)*. 2012;26(3):379–88.
15. Chang TJ, Tarn YH, Hsieh CL, et al. Taiwanese version of the EQ-5D: validation in a representative sample of the Taiwanese population. *J Formosan Med Assoc*. 2007;106(12):1023–31.
16. Chen P, Lin KC, Liing RJ, et al. Validity, responsiveness, and minimal clinically important difference of EQ-5D-5L in stroke patients undergoing rehabilitation. *Qual Life Res*. 2016;25(6):1585–96.
17. Chen J, Wong CK, McGhee SM, Pang PK, Yu WC. A comparison between the EQ-5D and the SF-6D in patients with chronic obstructive pulmonary disease (COPD). *PLoS One*. 2014;9(11):1–9.
18. Cheung PWH, Wong CKH, Cheung JPY. Differential psychometric properties of EuroQol 5-dimension 5-level and short-form 6-dimension utility measures in low back pain. *Spine*. 2019;44(11):E679–86.
19. Cheung PWH, Wong CKH, Lau ST, Cheung JPY. Responsiveness of the EuroQol 5-dimension (EQ-5D) in adolescent idiopathic scoliosis. *Eur Spine J*. 2018;27(2):278–85.
20. Cheung PWH, Wong CKH, Samartzis D, et al. Psychometric validation of the EuroQol 5-Dimension 5-Level (EQ-5D-5L) in Chinese patients with adolescent idiopathic scoliosis. *Scoliosis Spinal Disord*. 2016;11(19):1–12.
21. Fang H, Farooq U, Wang D, et al. Reliability and validity of the EQ-5D-3L for Kashin-Beck disease in China. *SpringerPlus*. 2016;5(1):1924.
22. Faridah A, Jamaayah H, Goh A, Soraya A. The validation of the EQ-5D in Malaysian dialysis patients. *Med J Malay*. 2010;65:114–9.
23. Gao K, Xia L, Pan SQ, Xiong T, Li SC. Validation of a Chinese Version of the Quality of Well-Being Scale-Self-Administered (QWB-SA) in patients with epilepsy. *Epilepsia*. 2013;54(9):1647–57.
24. Gandhi M, Ang M, Teo K, et al. EQ-5D-5L is more responsive than EQ-5D-3L to treatment benefit of cataract surgery. *Patient*. 2019;12(4):383–92.
25. Hisashige A, Mikasa H, Katayama T. Description and valuation of health-related quality of life among the general public in Japan by the EuroQol. *J Med Investig*. 1998;45(11–4):123–9.
26. Hung MC, Lu WS, Chen SS, et al. Validation of the EQ-5D in patients with traumatic limb injury. *J Occup Rehabil*. 2015;25(2):387–93.

27. Jia YX, Cui FQ, Zhang DL, et al. Comparison between the EQ-5D-5L and the EQ-5D-3L in patients with hepatitis B. *Qual Life Res.* 2014;23(8):2355–63.
28. Jin H, Wang B, Gao Q, et al. Comparison between EQ-5D and SF-6D in rural residents of Jiangsu Province, China. *PLoS One.* 2012;7(7):e41550.
29. Kim MH, Cho YS, Uhm WS, Kim S, Bae SC. Cross-cultural adaptation and validation of the Korean version of the EQ-5D in patients with rheumatic diseases. *Qual Life Res.* 2003;14(5):1401–6.
30. Kim SH, Hwang JS, Kim TW, Hong YS, Jo MW. Validity and reliability of the EQ-5D for cancer patients in Korea. *Support Care Cancer.* 2012;20(12):3155–60.
31. Kim TH, Jo MW, Lee SI, Kim SH, Chung SM. Psychometric properties of the EQ-5D-5L in the general population of South Korea. *Qual Life Res.* 2013;22(8):2245–53.
32. Kim SH, Jo MW, Lee JW, Lee HJ, Kim JK. Validity and reliability of EQ-5D-3L for breast cancer patients in Korea. *Health Qual Life Outcomes.* 2015;13(203):1–7.
33. Kim SK, Kim SH, Jo MW, Lee SI. Estimation of minimally important differences in the EQ-5D and SF-6D indices and their utility in stroke. *Health Qual Life Outcomes.* 2015;13(32):1–6.
34. Kim SH, Kim HJ, Lee SI, Jo MW. Comparing the psychometric properties of the EQ-5D-3L and EQ-5D-5L in cancer patients in Korea. *Qual Life Res.* 2012;21(6):1065–73.
35. Kimman M, Vathesatogkit P, Woodward M, et al. Validity of the Thai EQ-5D in an occupational population in Thailand. *Qual Life Res.* 2013;22(6):1499–506.
36. Lang HC, Chuang L, Shun SC, Hsieh CL, Lan CF. Validation of EQ-5D in patients with cervical cancer in Taiwan. *Support Care Cancer.* 2010;18(10):1279–86.
37. Lee CF, Luo N, Ng R, et al. Comparison of the measurement properties between a short and generic instrument, the 5-level EuroQoL Group's 5-dimension (EQ-5D-5L) questionnaire, and a longer and disease-specific instrument, the Functional Assessment of Cancer Therapy-Breast (FACT-B), in Asian breast cancer patients. *Qual Life Res.* 2013;22(7):1745–51.
38. Lee CF, Ng R, Luo N, et al. The English and Chinese versions of the five-level EuroQoL Group's five-dimension questionnaire (EQ-5D) were valid and reliable and provided comparable scores in Asian breast cancer patients. *Support Care Cancer.* 2013;21(1):201–9.
39. Lee WJ, Song KH, Noh JH, Choi YJ, Jo MW. Health-related quality of life using the EuroQol 5D questionnaire in Korean patients with type 2 diabetes. *J Korean Med Sci.* 2012;27(3):255–60.
40. Leung YY, Png ME, Wee HL, Thumboo J. Comparison of EuroQol-5D and short form-6D utility scores in multi-ethnic Asian patients with psoriatic arthritis: a cross-sectional study. *J Rheumatol.* 2013;40(6):859–65.
41. Li L, Liu C, Cai X, et al. Validity and reliability of the EQ-5D-5L in family caregivers in leukemia patients. *BMC Cancer.* 2019;19:522.
42. Li S, Wang M, Liu L, Chen G. Which approach is better in eliciting health state utilities from breast cancer patients? Evidence from mainland China. *Eur J Cancer Care.* 2019;28(2):e12965.
43. Liu L, Li S, Zhao Y, Zhang J, Chen G. Health state utilities and subjective well-being among psoriasis vulgaris patients in mainland China. *Qual Life Res.* 2018;27(5):1323–33.
44. Lu WS, Huang SL, Yang JF, et al. Convergent validity and responsiveness of the EQ-5D utility weights for stroke survivors. *J Rehabil Med.* 2016;48(4):346–51.
45. Lu P, Liang Y. Health-related quality of life of young Chinese civil servants working in local government: comparison of SF-12 and EQ5D. *Appl Res Qual Life.* 2016;11(4):1445–64.
46. Luo N, Cang SQ, Quah HMJ, How CH, Tay EG. The discriminative power of the EuroQol visual analog scale is sensitive to survey language in Singapore. *Health Qual Life Outcomes.* 2012;10(32):1–6.
47. Luo N, Chew LH, Fong KY, et al. A comparison of the EuroQol-5D and the Health Utilities Index mark 3 in patients with rheumatic disease. *J Rheumatol.* 2003;30(10):2268–74.
48. Luo N, Chew LH, Fong KY, et al. Validity and reliability of the EQ-5D self-report questionnaire in Chinese-speaking patients with rheumatic diseases in Singapore. *Ann Acad Med Singap.* 2003;32(5):685–90.
49. Luo N, Chew LH, Fong KY, et al. Validity and reliability of the EQ-5D self-report questionnaire in English-speaking patients with rheumatic diseases in Singapore. *Qual Life Res.* 2003;12(1):87–92.
50. Luo N, Low S, Lau PN, Au WL, Tan LC. Is EQ-5D a valid quality of life instrument in patients with Parkinson's disease? A study in Singapore. *Ann Acad Med Singap.* 2009;38(6):521–8.
51. Luo N, Ng WY, Lau PN, Au WL, Tan LC. Responsiveness of the EQ-5D and 8-item Parkinson's Disease Questionnaire (PDQ-8) in a 4-year follow-up study. *Qual Life Res.* 2010;19(4):565–9.
52. Luo N, Seng BK, Thumboo J, Feeny D, Li SC. A study of the construct validity of the Health Utilities Index Mark 3 (HUI3) in patients with schizophrenia. *Qual Life Res.* 2006;15(5):889–98.
53. Mahadeva S, Wee HL, Goh KL, Thumboo J. The EQ-5D (Euroqol) is a valid generic instrument for measuring quality of life in patients with dyspepsia. *BMC Gastroenterol.* 2009;9(1):20.
54. Mok WK, Wong WH, Mok GT, et al. Validation and application of health utilities index in Chinese subjects with down syndrome. *Health Qual Life Outcomes.* 2014;12(144):1–8.
55. Nishimura K, Oga T, Ikeda A, et al. Comparison of health-related quality of life measurements using a single value in patients with asthma and chronic obstructive pulmonary disease. *J Asthma.* 2008;45(7):615–20.
56. Oga T, Nishimura K, Tsukino M, et al. A comparison of the responsiveness of different generic health status measures in patients with asthma. *Qual Life Res.* 2003;12(5):555–63.
57. Pan CW, Sun HP, Wang X, et al. The EQ-5D-5L index score is more discriminative than the EQ-5D-3L index score in diabetes patients. *Qual Life Res.* 2015;24(7):1767–74.
58. Pattanaphesaj J, Thavorncharoensap M. Measurement properties of the EQ-5D-5L compared to EQ-5D-3L in the Thai diabetes patients. *Health Qual Life Outcomes.* 2015;13(14):1–8.
59. Purba FD, Hunfeld JAM, Iskandarsyah A, et al. Quality of life of the Indonesian general population: test-retest reliability and population norms of the EQ-5D-5L and WHOQOL-BREF. *PLoS One.* 2018;13(5):e0197098.
60. Saiguay W, Sakhong P. The psychometric testing of the Thai version of the Health Utilities Index in patients with ischemic heart disease. *Qual Life Res.* 2013;22(7):1753–9.
61. Sakhong P, Munpan W. A head-to-head comparison of UK SF-6D and Thai and UK EQ-5D-5L value sets in Thai patients with chronic diseases. *Appl Health Econom Health Pol.* 2017;15(5):669–79.
62. Sakhong P, Schommer JC, Gross CR, Prasithsirikul W, Sakulbumrungsil R. Health utilities in patients with HIV/AIDS in Thailand. *Value Health.* 2009;12(2):377–84.
63. Sakhong P, Sonsa-Ardjit N, Sukarnjanaset P, Munpan W. Psychometric properties of the EQ-5D-5L in Thai patients with chronic diseases. *Qual Life Res.* 2015;24(12):3015–22.
64. Setiawan D, Dusafitri A, Galistiani G, et al. Health-related quality of life of patients with HPV-related cancers in Indonesia. *Value Health Reg Issues.* 2018;15:63–9.
65. Shafie AA, Hassali MA, Liaw SY. A cross-sectional validation study of EQ-5D among the Malaysian adult population. *Qual Life Res.* 2011;20(4):593–600.
66. Shirowa T, Fukuda T, Ikeda S, et al. Japanese population norms for preference-based measures: EQ-5D-3L, EQ-5D-5L, and SF-6D. *Qual Life Res.* 2017;25(3):707–19.

67. Shiroiwa T, Fukuda T, Shimozuma K, et al. Comparison of EQ-5D scores among anthracycline-containing regimens followed by taxane and taxane-only regimens for node-positive breast cancer patients after surgery: the N-SAS BC 02 trial. *Value Health*. 2011;14(5):746–51.
68. Subramaniam M, Abdin E, Poon LY, et al. EQ-5D as a measure of programme outcomes: results from the Singapore early psychosis intervention programme. *Psychiatry Res*. 2014;215(1):46–51.
69. Tran BX, Ohinmaa A, Nguyen LT. Quality of life profile and psychometric properties of the EQ-5D-5L in HIV/AIDS patients. *Health Qual Life Outcomes*. 2012;10(132):1–8.
70. Tsang HHL, Cheung JPY, Wong CKH, et al. Psychometric validation of the EuroQol 5-dimension (EQ-5D) questionnaire in patients with spondyloarthritis. *Arthritis Res Ther*. 2019;21(1):41.
71. Varatharajan S, Chen WS. Reliability and validity of EQ-5D in Malaysian population. *Appl Res Qual Life*. 2011;7:209–21.
72. Wang X, Guo G, Zhou L, et al. Health-related quality of life in pregnant women living with HIV: a comparison of EQ-5D and SF-12. *Health Qual Life Outcomes*. 2017;15(1):1–10.
73. Wang SL, Hsieh E, Zhu LA, Wu B, Lu LJ. Comparative assessment of different health utility measures in systematic lupus erythematosus. *Sci Rep*. 2015;5(13297):1–6.
74. Wang P, Luo N, Tai ES, Thumboo J. The EQ-5D-5L is more discriminative than the EQ-5D-3L in patients with diabetes in Singapore. *Value Health Res Issues*. 2016;9:57–62.
75. Wang HM, Patrick DL, Edwards TC, et al. Validation of the EQ-5D in a general population sample in urban China. *Qual Life Res*. 2012;21(1):155–60.
76. Wang SL, Wu B, Zhu LA, et al. Construct and criterion validity of the Euro Qol-5D in patients with systematic lupus erythematosus. *PLoS One*. 2014;9(6):e98883.
77. Wee HL, Loke WC, Li SC, et al. Cross-cultural adaptation and validation of Singapore Malay and Tamil versions of the EQ-5D. *Ann Acad Med Singap*. 2007;36(6):403–8.
78. Wong CK, Lam ET, Lam CL. Comparison of direct-measured and derived short form six dimensions (SF-6D) health preference values among chronic hepatitis B patients. *Qual Life Res*. 2013;22(10):2973–81.
79. Wong CKH, Lang BHH, Yu HMS, Lam CLK. EQ-5D-5L and SF-6D utility measures in symptomatic benign thyroid nodules: acceptability and psychometric evaluation. *Patient*. 2017;10(4):447–54.
80. Wong CKH, Mulhern B, Cheng GHL, Lam CLK. SF-6D population norms for the Hong Kong Chinese general population. *Qual Life Res*. 2018;27(9):2349–59.
81. Wong CK, Mullhern B, Wan YF, Lam CL. Responsiveness was similar between direct and mapped SF-6D in colorectal cancer patients who declined. *J Clin Epidemiol*. 2014;67(2):219–27.
82. Wu J, Han Y, Zhao FL, et al. Validation and comparison of EuroQoL-5 dimension (EQ-5D) and Short Form-6 dimension (SF-6D) among stable angina patients. *Health Qual Life Outcomes*. 2014;12:156.
83. Wu C, Gong Y, Wu J, et al. Chinese version of the EQ-5D preference weights: applicability in a Chinese general population. *PLoS One*. 2016;11(10):e0164334.
84. Xie F, Li SC, Luo N, et al. Comparison of the EuroQol and short form 6D in Singapore multi-ethnic Asian knee osteoarthritis patients schedules for total knee replacement. *Arthritis Rheumatol*. 2007;57(6):1043–9.
85. Yang Z, Busschbach J, Liu G, Luo N. EQ-5D-5L norms for the urban Chinese population in China. *Health Qual Life Outcomes*. 2018;16:210.
86. Yang F, Lau T, Lee E, et al. Comparison of the preference-based EQ-5D-5L and SF-6D in patients with end-stage renal disease (ESRD). *Eur J Health Econ*. 2015;16(9):1019–26.
87. Ye Z, Sun L, Wang Q. A head-to-head comparison of EQ-5D-5L and SF-6D in Chinese patients with low back pain. *Health Qual Life Outcomes*. 2019;17(1):57.
88. Yu ST, Chang HY, Yao KP, Lin YH, Hurng BS. Validity of EQ-5D in general population of Taiwan: results of the 2009 national health interview and drug abuse survey of Taiwan. *Qual Life Res*. 2015;24(10):2541–8.
89. Zhao L, Liu X, Liu D, et al. Comparison of the psychometric properties of the EQ-5D-3L and SF-6D in the general population of Chengdu city in China. *Medicine*. 2019;98(11):e14719.
90. Zhao FL, Yue M, Yang H, et al. Validation and comparison of EuroQol and short form 6D in chronic prostatitis patients. *Value Health*. 2010;13(5):649–56.
91. Grobet C, Marks M, Tecklenburg L, Audige L. Application and measurement properties of EQ-5D to measure quality of life in patients with upper extremity orthopaedic disorders: a systematic literature review. *Achieves Orthopaedic Trauma Surg*. 2019;138:953–61.
92. Papaioannou D, Brazier J, Parry G. How valid and responsive are generic health status measures, such as EQ-5D and SF-36, in Schizophrenia? A systematic review. *Value Health*. 2011;14:907–20.
93. Yang Y, Brazier J, Longworth L. EQ-5D in skin conditions: an assessment of validity and responsiveness. *Eur J Health Econ*. 2015;16:927–39.
94. Janssen MF, Lubetkin EI, Sekhobo JP, Pickard AS. The use of the EQ-5D preference-based health status measure in adults with type 2 diabetes mellitus. *Diabetic Med*. 2010;28:395–413.
95. Finch AP, Brazier JE, Mukuria C. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. *Eur J Health Econ*. 2017;19:557–70.
96. Pickard S, Wilke C, Jung E, et al. Use of a preference-based measure of health (EQ-5D) in COPD and asthma. *Respir Med*. 2008;102:519–36.
97. Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use*. Oxford: Oxford University Press; 2015.
98. Kind P, Brooks R, Rabin R. *EQ-5D concepts and methods: a developmental history*. Amsterdam: Springer; 2005.
99. Cheung YB, Thumboo J. Developing health-related quality-of-life instruments for use in Asia: the issues. *PharmacoEconomics*. 2006;24(7):643–50.
100. Tan RLY, Lee GL, Herdman M, Luo N. Assessing the appropriateness of the EQ-5D-5L questionnaire in Asia: a qualitative study of Singaporeans. Presented at the 1st EuroQol Academy Meeting 2019; Guangzhou, China.
101. United Nation Educational Scientific and Cultural Organisation (UNESCO) Institute for Statistics (UIS). http://uis.unesco.org/en/home#tabs-0-uis_home_top_menus-3. Assessed 9 Sep 2019.
102. Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Qual Life Res*. 1998;7(4):323–35.
103. Luo N, Chew L-H, Fong K-Y, Koh D-R, Ng S-C, Yoon K-H, et al. Do English and Chinese EQ-5D versions demonstrate measurement equivalence? An exploratory study. *Health Qual Life Outcomes*. 2003;1(1):7.
104. Wang Y, Tan N-C, Tay E-G, Thumboo J, Luo N. Cross-cultural measurement equivalence of the 5-level EQ-5D (EQ-5D-5L) in patients with type 2 diabetes mellitus in Singapore. *Health Qual Life Outcomes*. 2015;13(1):103.
105. Wee H-L, Cheung Y-B, Fong K-Y, Luo N, Machin D, Thumboo J. Are English- and Chinese-language versions of the SF-6D equivalent? A comparison from a population-based study. *Clin Ther*. 2004;26(7):1137–48.
106. Tan ML, Wee H-L, Lee J, Ma S, Heng D, Tai E-S, et al. The Short Form 36 English and Chinese versions were equivalent in a multiethnic Asian population. *J Clin Epidemiol*. 2013;66(7):759–67.