



Psychometric properties comparison between EQ-5D-5L and EQ-5D-3L in the general Thai population

Krittaphas Kangwanrattanakul¹ · Pornnip Parmontree¹

Accepted: 25 July 2020
© Springer Nature Switzerland AG 2020

Abstract

Purpose Evidence for the EQ-5D-5L's psychometric properties in the general Thai population is limited. This study aimed to compare ceiling effect, discriminatory power, response redistribution, validity, reliability between the EQ-5D-5L (5L) and the EQ-5D-3L (3L) in the general Thai population.

Methods 1200 participants were randomly selected. The Shannon index (H') and Shannon evenness index (J') determining discriminatory power of both EQ-5D versions in each dimension were compared. Test–retest reliability was evaluated using weighted kappa (k) and intraclass correlation coefficients (ICCs). Validity was evaluated by correlations between similar dimensions of the EQ-5D, WHOQOL-BREF, and SF-12v2 and known-groups validity. The ceiling effects for the 3L and for the 5L were compared.

Results The 5L had lower ceiling effects than the 3L (49.08% vs 57.17%, $p < 0.01$). H' was higher for the 5L than for the 3L, but J' showed otherwise. Moderate correlations were detected between similar dimensions of the EQ-5D and the WHOQOL-BREF and SF-12v2. ICCs and k of the 3L were slightly higher than those of the 5L (ICCs: 0.78 vs 0.71) and (k : 0.42–0.63 vs 0.48–0.61), respectively. Older, female participants and those with comorbidities reported a lower utility index for both versions.

Conclusion Evidence supported use of both EQ-5D versions in the general Thai population. The 5L had better ceiling effects and discriminant activity, while it showed comparable known-groups validity with the 3L. Nevertheless, evidence is limited for the superiority of reliability between these two versions, so more future research is required to investigate it.

Keywords Psychometric properties · EQ-5D-5L · EQ-5D-3L · Comparison · Health-related quality of life

Introduction

The EQ-5D has been used to measure both patients and general populations' health [1–8]. Due to its simplicity, its self-completion with low cost burden, and capacity to generate a preference-weighted index score, known as a utility score, the EQ-5D is commonly used to assess humanistic outcomes for economic appraisal recommended by several HTA guidelines including Thai [9–11].

The first version of the EQ-5D, the EQ-5D-3L (3L), was introduced in 1990 and now has been translated into more

than 170 languages [2]. Its EQ-5D descriptive system has five dimensions, each with three response options including no problem, some/moderate problem, and extreme problem [4]. Nevertheless, previous evidence has revealed some drawbacks of the 3L's use, including high ceiling effect, minor discriminative power, and less sensitivity to clinical changes in both general populations and clinical areas when compared to the SF-6D, SF-12, and SF-36 [12–16]. To solve these problems and still preserve clinical relevance to a wide range of health conditions and populations, a newer version of the EQ-5D, the EQ-5D-5L (5L), was developed and introduced by the EuroQoL group in 2015 [1]. This 5L version includes two additional response options, “slight problem” and “severe problem,” for each of the five dimensions. As a result, the EQ-5D-5L has five response options; no problem, slight problem, moderate problem, severe problem, and extreme/unable to perform [1]. This version is expected to diminish the ceiling effect and improve discriminative power

✉ Krittaphas Kangwanrattanakul
krittaphas@buu.ac.th

¹ Division of Social and Administrative Pharmacy, Faculty of Pharmaceutical Sciences, Burapha University, 169 Long-Hard Bangsaen Rd., Mueang, Chonburi 20131, Thailand

in both general populations and clinical areas. Moreover, this version has now been translated into more than 113 languages including Thai.

To date, several studies examining the 5L's psychometric properties have found it a valid and reliable instrument. Compared to the 3L in both clinical areas and general populations [1, 7, 17–19], it has a lesser ceiling effect but more enhanced discriminative power for clinical changes. Previous evidence has also suggested that the 5L might capture more severe health problems in the patient population, and it might differentiate mild health states, particularly in the pain/discomfort and anxiety/depression dimensions in the general population [20, 21].

In Thailand, evidence is limited for the 5L's psychometric properties. To our knowledge, only two studies have explored the 5L's measurement properties in patients with diabetes [22] and a wide range of chronic diseases [23]. These studies revealed that the 5L was a valid and reliable instrument, with less ceiling effect than the 3L in the patient group. Nevertheless, evidence of the 5L's psychometric properties when administered to the general Thai population has not yet been established. Therefore, this study aimed to assess the 5L's psychometric properties in comparison with the 3L in terms of practicality (administration time and ceiling effect), discriminatory power, response redistribution, test–retest reliability, validity (known-groups and construct validities), and acceptability in the general Thai population.

Methods

Participants and settings

A cross sectional survey study was conducted with study participants ($n = 1200$) randomly selected from five provinces including Nakhon-Srithammarat, Khon-Kaen, Chonburi, Chaing-Mai, and Bangkok (the capital city). Inclusion criteria included (1) age 20–70 years and (2) understanding of the Thai language and the data collection process, as evaluated by the interviewers or the researcher (KK). Exclusion criteria were: (1) being diagnosed with acute or life-threatening illness, (2) having cognitive impairment or (3) disability. Four-stage stratified random sampling was employed to select the provinces, districts, and villages for data collection.

Data collection

Each subject completed the self-administered questionnaire using a paper and pencil as follows: general subject information, EQ-5D-5L, the short form 12 health survey version 2 (SF-12v2), WHOQOL-BREF, EQ-5D-3L, EQ-VAS, and two acceptability questions—(1) ease of

understanding of the EQ-5D and (2) better reflection of health status. Moreover, our interviewers were assigned to be with all subjects to record the time for each part of the questionnaire. Permission to use the Thai version of those instruments was granted by the appropriate officials. All subjects received approximately 3.20 USD (1 USD = 31.19 THB) as compensation for their time. The majority of subjects (95%) completed the questionnaire by themselves; however, for those who had an eyesight problem, our interviewers read all questions and response options without elaborating or interpreting them.

Four hundred subjects were asked to complete both EQ-5D versions at their homes 2–3 weeks after their first interview and to send them back to the researcher (KK) in a prepaid mailing envelope. Subjects were asked to assess whether their individual health status had changed after their first interview, and a five-point Likert scale was used: (1) much better, (2) somewhat better, (3) the same, (4) somewhat worse, and (5) much worse. The researcher (KK) made reminder phone calls 2 weeks after the initial assessments. A questionnaire reaching the researcher (KK) after 21 days was excluded from this analysis.

Instruments

EQ-5D

The EQ-5D is a brief, self-report questionnaire measuring respondents' general health. Respondents were required to rate their health on the day of the questionnaire's administration. The EQ-5D is comprised of two parts; the first part is the EQ-5D descriptive system consisting of five dimensions including mobility (MO), self-care (SC), usual activities (UA), pain/discomfort (PD), and anxiety/depression (AD). This part is generally used to calculate the EQ-5D index using a country-specific value set for economic analyses. At present, both 3L and 5L Thai versions have individual value sets for calculating the EQ-5D index [24, 25], which generally ranges from 0 to 1, where 1.00 represents perfect health and 0 represents death. The lowest of the Thai index scores are -0.454 and -0.283 for the health state “33333” and “55555” of the 3L and 5L, respectively, while the maximum Thai index score for both versions is 1.00. Moreover, a negative value represents the health state as worse than death. The second part is the EQ-VAS, that is, the respondent's self-rated health on a 20-cm vertical line (visual analog scale) measuring the current respondent's self-rated health, where endpoints are labeled “worst imaginable health state” at 0 and “best imaginable health state” at 100 [2]. The scores of EQ-VAS range from 0 to 1 and are obtained from dividing the number marked on the scale by 100.

WHOQOL-BREF

The WHOQOL-BREF is a shorter version of the WHO-QoL-100, developed by the World Health Organization (WHO) collecting data from 15 countries including Thailand [26]. This instrument requires respondents to rate their HRQoL levels during the past 2 weeks. The WHOQOL-BREF contains 24 items grouped into four dimensions as follow: physical (7 items), psychological (6 items), social (3 items), and environmental (8 items). Two other items are one for general health and another for overall quality of life. Response options are on a 5-point Likert scale; 1 = not at all, 2 = not much, 3 = moderately, 4 = a great deal, and 5 = completely. WHOQOL results are reported as raw scores for each dimension, calculated by multiplying the mean score of all items by four, so the score can range from 4 to 20 for the four domains. It can also be converted to a transformed score ranging from 0 (the worst possible health status) to 100 (the best possible health status). The official Thai version of WHOQOL-BREF is available [27].

SF-12 version 2

The generic health profile “12-item Short Form Survey version 2” is a short version of the 36-item Short Form Survey (SF-36) for measuring health status in large surveys [28]. It has been proven to be valid and reliable in Thai patients with chronic diseases [23, 29, 30]. It consists of 12 items further grouped into eight dimensions including Physical Functioning (PF: 2 items), Role limitations due to physical problems (RP: 2 items), Bodily Pain (BP: 1 item), General Health (GH: 1 item), Vitality (VT: 1 item), Social Functioning (SF: 1 item), Role limitations due to emotional problems (RE: 2 items), and Mental Health (MH: 2 items). SF-12 scores can be transformed from 0 (the worst possible health status) to 100 (the best possible health status) for each health dimension, and they can be converted to norm-based scoring, which is referred to 50 ± 10 (mean \pm SD). Moreover, scores of those eight dimensions can be summarized into two major scales, the Physical Component Summary (PCS) and the Mental Component Summary (MCS) [31]. In this study, we used the 4-week standard recall period of the SF-12v2.

Data analyses

Practicality

The ceiling effect was computed as the proportion of subjects reporting “no problem” (level 1) for both versions in each dimension and across all five dimensions divided by the total number of subjects. An acceptable percentage of ceiling effect was set as less than 15% [32]. We hypothesized that by adding two more levels of impairment to the 3L,

the ceiling effect of the 5L would be diminished. Absolute and relative reductions of this effect from 3L to 5L were computed and reported. The average time of both EQ-5D versions’ completion was also reported and compared.

Discriminatory power

Two indices, Shannon entropy (Shannon index (H')) and information efficiency (Shannon evenness index (J')), were employed to determine each dimension’s discriminatory power. The Shannon index is defined as follows:

$$H' = - \sum_{i=1}^C P_i \log_2 P_i,$$

where H' is the absolute amount of informativity captured, C is the number of levels in this study, and $P_i = n_i/N$ is the proportion of observations at the i th level ($i = 1, \dots, C$) among our study samples, where n_i is the number of responses at the i th level and N is the total sample size. A higher Shannon index (H') indicates more information captured by the instrument and better discriminant activity.

Shannon evenness index (J') means evenness of information distribution regardless of the number of response options [33]. J' was calculated as H'/H'_{\max} , and its value ranges from 0 to 1, where 1 means all response options selected with the same frequency. We hypothesized that the 5L would have higher H' , and J' of 5L would remain equal to or decrease slightly from the 3L.

Response redistribution

Response redistribution was used to determine both versions’ response consistency. To quantify consistency, the 3L response level was recoded to the 5L (the 3L_{5L}) response level as follows: 1 = 1, 2 = 3, and 3 = 5 [7, 22, 34]. Inconsistency size was calculated as $|3L_{5L} - 5L| - 1$, which means zero or less indicated consistency. The mean of EQ-VAS from the 5L for each pair was also quantified to ensure response redistribution’s validity. The mean of individuals’ VAS scores remaining at the same level was hypothesized to be higher than those selecting more severe problems on the 5L or to be lower than those selecting milder problems on the 5L.

Validity

Construct validity was evaluated in terms of convergent and discriminant validity via correlations between the five dimensions of 3L and 5L and other well-established HRQoL instruments, WHOQOL-BREF and SF-12v2, using Spearman’s rank rho correlations. Colton’s rule was used to determine the strength of correlation as follows: weak or no ($r < 0.25$), moderate ($0.25 \leq r < 0.50$), moderate to strong

($0.50 \leq r < 0.75$), and strong ($r \geq 0.75$) [35]. Convergent validity represents a high correlation between the dimensions of these instruments measuring similar constructs, whereas discriminant validity represents otherwise [32]. Hypothesized strong correlations were expected among these pairs, including MO/PF/Physical dimension, PD/BP/Physical dimension, and AD/MH/Psychological dimension. The correlation level between EQ-VAS scores was also determined and reported using Pearson's correlation.

Known-group validity was performed to investigate utility index changes against participant sub-groups defined by demographic characteristics. We hypothesized that low utility scores would be observed among women, smokers/ex-smokers, drinkers/ex-drinkers, older samples (≥ 54 years old), and those with lower education levels (no schooling or primary school), lower incomes ($< 30,000$ THB or 990 USD), and higher numbers of comorbidities. Multivariable analyses were used to investigate the associations between the demographic characteristics and 3L and 5L index scores.

Reliability

Test–retest reliability was assessed among subjects with stable health between initial and second assessments. The reliability of the EQ-5D index and the EQ-VAS scores were examined using intraclass correlation coefficients (ICCs), while the reliability of each dimension in the EQ-5D descriptive system of both versions was assessed and compared using a weighted kappa coefficient. Rosner's guideline was used to determine the agreement level for both ICCs and weighted kappa coefficients as follows: poor reproducibility (< 0.4), good reproducibility ($0.4–0.75$) and excellent reproducibility (≥ 0.75) [7, 19, 36, 37].

Acceptability

Responses to the two acceptability questions, including ease of understanding and better reflection of health status, were summarized and reported in terms of percentages.

All statistical analyses were performed using IBM SPSS version 23, with the p value < 0.05 generally considered statistically significant.

Results

Characteristics of study subjects

Table 1 displays basic characteristics of all subjects ($n = 1200$). Most were female ($n = 640$, 53.3%), married ($n = 765$, 63.7%), and educated at secondary school ($n = 514$, 42.9%). Subjects' mean age and household income were 42.7 years ($SD = 13.7$) and 12,631.50 ($SD = 10,276.5$) THB/

Table 1 Characteristics of study subjects ($n = 1200$)

Characteristics	n (%)
Gender, n (%)	
Male	560 (46.7)
Female	640 (53.3)
Age (years)	
Mean \pm SD	42.7 \pm 13.7
Median (interquartile range)	43 (31–54)
Marital status, n (%)	
Single	347 (28.9)
Married	765 (63.7)
Widowed	41 (3.4)
Divorced/separated	47 (3.9)
Education, n (%)	
Primary school or lower	414 (34.5)
Secondary school	514 (42.9)
College degree	102 (8.5)
Bachelor's degree	150 (12.5)
Postgraduate degree(s)	10 (1.7)
Health insurance, n (%)	
Civil servant medical benefit scheme	59 (4.9)
Universal coverage	737 (61.4)
Social security	364 (30.3)
Others	40 (3.3)
Average household income (baht/month)	
Mean \pm SD	12,631.5 \pm 10,276.5
Median (interquartile range)	10,000 (7000–15,000)

month, respectively. Moreover, most of them reported as healthy ($n = 844$, 70.33%). There were no missing values from either EQ-5D version.

Practicality

As shown in Table 2, both 5L and 3L showed that the highest and lowest proportions of subjects rating “no problem (level 1)” were SC (97.3% vs 97.5%, $p > 0.05$) and PD (64.8% vs 57.8%, $p < 0.01$), respectively, and PD showed the highest relative reduction of 10.68%. Moreover, overall ceiling effects reduced from 57.17% for the 3L to 49.08% for the 5L, with a relative reduction of 14.14%. The average times for subjects to complete the 3L and the 5L were 2.08 ± 1.03 and 2.20 ± 1.04 min, respectively.

Discriminatory power

Table 3 presents the Shannon index and the Shannon evenness index of the two EQ-5D versions. Our results revealed that the Shannon index (H') increased when samples rated two more severity levels of the 5L for all five dimensions (range 0.19–1.37). As expected, the Shannon evenness index

Table 2 The absolute and relative reductions of ceiling from the 3L to the 5L and descriptive statistics of both EQ-5D versions

Dimensions	No problems; <i>n</i> (%)		<i>p</i> value ^a	Absolute (%)	Relative (%)
	3L	5L			
Mobility (MO)	977 (81.4%)	925 (77.1%)	<0.001	4.30	5.32
Self-care (SC)	1168 (97.3%)	1170 (97.5%)	0.774	-0.20	-0.17
Usual activities (UA)	1073 (89.4%)	1065 (88.8%)	0.280	0.90	0.75
Pain/discomfort (PD)	777 (64.8%)	694 (57.8%)	<0.001	7.00	10.68
Anxiety/depression (AD)	996 (83.0%)	962 (80.2%)	<0.001	2.80	0.40
Full health	686 (57.17%)	589 (49.08%)	<0.001	8.09	14.14

^aMcNemar test**Table 3** Discriminant power measured by Shannon index (*H'*) and Shannon evenness index (*J'*) for the 5L compared to the 3L (*n* = 1200)

Domains	<i>H'</i>		<i>J'</i>	
	3L	5L	3L	5L
MO	0.69	1.07	0.44	0.46
SC	0.18	0.19	0.11	0.08
UA	0.49	0.60	0.31	0.26
PD	0.96	1.37	0.61	0.59
AD	0.68	0.88	0.43	0.38

MO mobility, SC self-care, UA usual activities, PD pain/discomfort, AD anxiety/depression

(*J'*) was lower in the 5L than in the 3L for all dimensions except MO. The percentage of relative Shannon evenness index reduction showed that its maximum and minimum were found in PD (3.28%) and SC (27.27%), respectively.

Response redistribution

As shown in Table 4, most samples reporting level 1-3L remained at level 1-5L for all five dimensions (87.3–99.6%). Of the study samples answering level 2-3L, from 57.8% for MO to 71.2% for PD shifted their answers to level 2-5L, whereas approximately 14.5% for AD—27.4% for MO upgraded their answers to level 3-5L. The proportion of samples marked for AD (25.0%) and PD (100%) reporting level 3-3L redistributed their answers to level 4-5L. Moreover, only two samples (50.0%) shifted their answers from level 3-3L to level 5-5L for AD. Of 6,000 redistribution pairs, small proportions of inconsistent pairs were observed in the AD (*n* = 30, 0.5%) and the SC (*n* = 7, 0.12%) dimensions.

Validity

Table 5 shows convergent and discriminant validity. The physical dimension of the WHOQOL-BREF had moderate correlations with the MO ($r = -0.32$ for the 3L, $r = -0.33$ for the 5L, all $p < 0.01$), PD ($r = -0.36$ for the 3L, $r = -0.35$

for the 5L all $p < 0.01$), while MO and PD had moderate correlations with PF ($r = -0.42$ for the 3L, $r = -0.47$ for the 5L, all $p < 0.01$) and BP ($r = -0.33$ for the 3L, $r = -0.35$ for the 5L, all $p < 0.01$) of the SF-12v2, respectively. Nevertheless, AD had moderate correlations with the psychological dimension of the WHOQOL-BREF ($r = -0.34$ for the 3L, $r = -0.33$ for the 5L, $p < 0.01$), and it had weak and moderate correlations with MH of the SF-12v2 for the 3L and the 5L ($r = -0.24$ for the 3L, $r = -0.30$ for the 5L, $p < 0.01$), respectively. The EQ-VAS produced the highest correlation with the physical dimension of the WHOQOL-BREF ($r = 0.40$, $p < 0.01$), while it yielded the strongest correlation with GH of the SF-12v2 ($r = 0.35$, $p < 0.01$).

As displayed in Table 6, we found that both EQ-5D versions could discriminate utility scores well in regard to gender, age, education level, household income, smoking, alcohol, and number of comorbidities. As expected, these following hypotheses were confirmed for both EQ-5D versions since we found that female, elderly, and those with one or more comorbidities tended to have a lower mean of utility index, with all $p < 0.05$.

Reliability

All 400 subjects completed the questionnaire at 2–3 weeks after the initial assessment, and all retest questionnaires were returned to researchers within 21 days. Of 400 subjects, 239 (59.75%) reported themselves with no health status change from the first measurement (Table 7). The 5L's weighted kappa coefficients ranged from 0.48 to 0.61, while the 3L's ranged from 0.42 to 0.63. Moreover, the MO from both versions had the highest reproducibility with weighted kappa coefficients of 0.63 (95% CI 0.51–0.76) for the 3L and 0.61 (95% CI 0.49–0.72) for the 5L, while the lowest reproducibility was observed in AD of the 3L and UA of the 5L, with weighted kappa coefficients of 0.42 (95% CI 0.29–0.55) and 0.48 (95% CI 0.35–0.60), respectively. Percentage agreements across five dimensions ranged from 0.81 to 0.97 for the 3L and from 0.75 to 0.97 for the 5L. The ICCs of 3L and 5L indexes and EQ-VAS

Table 4 Response redistribution from 3L to 5L and mean of EQ-VAS

Dimensions	3L	5L			
		Level 1	Level 2	Level 3	Level 4
Mobility (MO) <i>n</i> (%) ^a [mean EQ-VAS ^b , size of inconsistency ^c]					
Level 1	920 (94.2%) [84.53, -1]	47 (4.8%) [77.02, 0]	8 (0.8%) [76.25, 1]	2 (0.2%) [55.00, 2]	0 (0.0%) [0.00, 3]
Level 2	5 (2.1%) [72.00, 1]	129 (57.8%) [78.25, 0]	61 (27.4%) [71.23, -1]	27 (12.1%) [63.89, 0]	1 (0.4%) [70.00, 1]
Level 3	0 (0.0%) [0.00, 3]	0 (0.0%) [0.00, 2]	0 (0.0%) [0.00, 1]	0 (0.0%) [0.00, 0]	0 (0.0%) [0.00, -1]
Self-care (SC)					
Level 1	1163 (99.6%) [82.67, -1]	5 (0.4%) [76.00, 0]	0 (0.0%) [0.00, 1]	0 (0.0%) [0.00, 2]	0 (0.0%) [0.00, 3]
Level 2	7 (21.9%) [79.29, 1]	19 (59.4%) [67.11, 0]	5 (15.6%) [61.00, -1]	1 (3.1%) [40.00, 0]	0 (0.0%) [0.00, 1]
Level 3	0 (0.0%) [0.00, 3]	0 (0.0%) [0.00, 2]	0 (0.0%) [0.00, 1]	0 (0.0%) [0.00, 0]	0 (0.0%) [0.00, -1]
Usual activities (UA)					
Level 1	1048 (97.7%) [83.63, -1]	23 (2.1%) [75.00, 0]	2 (0.2%) [77.50, 1]	0 (0.0%) [0.00, 2]	0 (0.0%) [0.00, 3]
Level 2	17 (13.4%) [79.41, 1]	82 (64.6%) [73.29, 0]	27 (21.3%) [65.74, -1]	1 (0.8%) [40.00, 0]	0 (0.0%) [0.00, 1]
Level 3	0 (0.0%) [0.00, 3]	0 (0.0%) [0.00, 2]	0 (0.0%) [0.00, 1]	0 (0.0%) [0.00, 0]	0 (0.0%) [0.00, -1]
Pain/discomfort (PD)					
Level 1	678 (87.3%) [85.87, -1]	87 (11.2%) [81.44, 0]	10 (1.3%) [78.00, 1]	2 (0.3%) [100.00, 2]	0 (0.0%) [0.00, 3]
Level 2	16 (3.8%) [83.93, 1]	299 (71.2%) [79.25, 0]	93 (22.1%) [71.24, -1]	12 (2.9%) [52.92, 0]	0 (0.0%) [0.00, 1]
Level 3	0 (0.0%) [0.00, 3]	0 (0.0%) [0.00, 2]	0 (0.0%) [0.00, 1]	3 (100.0%) [0.00, 0]	0 (0.0%) [0.00, -1]
Anxiety/depression (AD)					
Level 1	935 (93.9%) [84.02, -1]	59 (5.9%) [80.08, 0]	1 (0.1%) [80.00, 1]	1 (0.1%) [90.00, 2]	0 (0.0%) [0.00, 3]
Level 2	27 (13.5%) [81.22, 1]	140 (70.0%) [75.50, 0]	29 (14.5%) [67.41, -1]	4 (2.0%) [72.50, 0]	0 (0.0%) [0.00, 1]
Level 3	0 (0.0%) [0.00, 3]	0 (0.0%) [0.00, 2]	1 (25.0%) [80.00, 1]	1 (25.0%) [50.00, 0]	2 (50.0%) [55.00, -1]

Inconsistencies are presented in bold

^aPercentage in each level of the 3L

^bMean of EQ-VAS from the 5L

^cSize of inconsistent response for each pair

were 0.78 (95% CI 0.71–0.83), 0.71 (95% CI 0.63–0.78), and 0.82 (95% CI 0.77–0.86), respectively.

Acceptability

Most subjects ($n = 589$, 49.1%) thought that the 5L was easier to understand than the 3L, while 34.8% reported no difference. Conversely, 29.8% of subjects thought that the 5L could better reflect their health than the 3L; however, 31.5% indicated that the two versions were similar.

Discussion

Ours is the first study investigating psychometric analyses including practicality, discriminatory power, response redistribution, validity, reliability, and acceptability of the 5L compared to the 3L in the general Thai population.

Like previous studies [6, 7, 17–19, 22, 38–42], adding two more levels of severity to the 3L could reduce the overall ceiling effect by 8.09%, with the relative reduction of 14.14%. Our percentage of ceiling effect reduction

Table 5 Comparison of convergent and discriminant validity of 5L and 3L with WHOQOL-BREF and SF-12v2

Dimensions	Correlation coefficients (3L/5L)					EQ-VAS ^b
	MO ^a	SC ^a	UA ^a	PD ^a	AD ^a	
WHOQOL-BREF						
Physical	-0.32/-0.33	-0.22/-0.22	-0.29/-0.30	-0.36/-0.35	-0.29/-0.27	0.40
Psychological	-0.23/-0.23	-0.20/-0.19	-0.24/-0.24	-0.28/-0.25	-0.34/-0.33	0.34
Environmental	-0.19/-0.19	-0.17/-0.18	-0.26/-0.25	-0.25/-0.20	-0.28/-0.27	0.23
Social	-0.22/-0.21	-0.11/-0.12	-0.24/-0.25	-0.25/-0.23	-0.23/-0.20	0.25
Overall QoL	-0.28/-0.29	-0.22/-0.22	-0.23/-0.23	-0.29/-0.28	-0.25/-0.23	0.38
SF-12v2						
PF	-0.42/-0.47	-0.16/-0.18	-0.31/-0.31	-0.37/-0.37	-0.28/-0.27	0.26
RP	-0.29/-0.31	-0.12/-0.13	-0.28/-0.30	-0.33/-0.33	-0.26/-0.28	0.22
BP	-0.22/-0.26	-0.10/-0.13	-0.21/-0.22	-0.33/-0.35	-0.24/-0.27	0.25
GH	-0.24/-0.28	-0.09/-0.08	-0.20/-0.19	-0.29/-0.30	-0.25/-0.24	0.35
VT	-0.14/-0.16	-0.01/-0.01	-0.07/-0.07	-0.19/-0.13	-0.14/-0.16	0.13
SF	-0.17/-0.15	-0.09/-0.10	-0.15/-0.15	-0.20/-0.17	0.26/-0.25	0.15
RE	-0.20/-0.22	-0.09/-0.09	-0.25/-0.24	-0.25/-0.24	-0.32/-0.34	0.18
MH	-0.16/-0.18	-0.02/-0.01	-0.03/-0.04	-0.17/-0.14	-0.24/-0.30	0.10
MCS	-0.12/-0.13	-0.03/-0.02	-0.08/-0.08	-0.16/-0.12	-0.27/-0.32	0.10
PCS	-0.32/-0.36	-0.13/-0.15	-0.29/-0.29	-0.38/-0.40	-0.24/-0.25	0.34

All values were significant at $p < 0.01$

PF physical functioning, RP role limitations due to physical problems, BP bodily pain, GH general health perceptions, SF social functioning, VT vitality, RE role limitations due to emotional problems, MH mental health, MCS Mental Component Score, PCS Physical Component Score, MO mobility, SC self-care, UA usual activities, PD pain/discomfort, AD anxiety/depression

^aSpearman's rank correlation

^bPearson's correlation

(3L–5L) was lower than those in previous studies that ranged from 9.7 to 20% [38, 43–45]. However, we confirmed that our results are valid since our percentage of ceiling reductions was similar to those in a previous study conducted with the general Korean population [7].

Not surprisingly, most samples retained their answers at level 1 for both EQ-5D versions in all five dimensions, consistent with various previous studies conducted in both general populations and patient groups [7, 18, 19, 22, 46]. This might be due to most recruited samples being relatively healthy, so they rated themselves with “no problems” for both EQ-5D versions. We also found inconsistent responses to these two versions in our samples, at an average proportion of 1.7%, highest in AD (0.5%) and lowest in SC (0.12%). This was similar to that reported in the previous studies [41, 47], thus indicating that our samples answered the two EQ-5D versions consistently.

As expected, adding two more levels of severity to the EQ-5D's descriptive system increased discriminative activity (H') in all dimensions from the 3L, with incremental values ranging from 0.01 to 0.41. Conversely, the Shannon evenness index (J') was lower in the 5L than in the 3L for all dimensions, except MO. Notably, our H' and J' values were slightly lower than the findings from previous studies [6,

17, 19, 39, 42, 46]. Because those studies were conducted in clinical areas and in general populations with a large sample size ($n = 7554$), samples with moderate/extreme conditions were more likely to be recruited. However, our values were similar to those reported in Pattanaphesaj et al. (0.21–1.40 for H' , 0.09–0.60 for J') [22]. This ascertains that our results are valid and that the 5L showed improvement in discriminant activity across a wide range of population in Thailand.

As for construct validity, hypothesized correlations between both EQ-5D versions and WHOQOL-BREF and SF-12v2 were confirmed because two similar dimensions from those instruments yielded a higher correlation coefficient than two dissimilar dimensions. However, the strength of correlation was not as strong as anticipated. We reasoned that both EQ-5D versions asked respondents to rate their current health status, whereas the WHOQOL-BREF and SF-12v2 asked respondents to rate their health with a 2-week and a 4-week recall period, respectively. Nevertheless, the correlation pattern was like those reported in previous studies [7, 19, 22, 23, 42], implying that our results are valid.

For known-group validity, both EQ-5D versions showed that decreases in utility index were observed among female, elderly, and those with higher number of comorbidities. These findings were consistent with previous studies [39,

Table 6 Known-group validity of 5L and 3L index scores using real Thai value sets using multivariable analyses

Sample characteristics	Sample size (<i>n</i>)	Standardized coefficients (β)		
		3L	5L	EQ-VAS
Gender (Ref: female)				
Male	560	0.114***	0.089*	0.014
Age (years) grouped by quartile (Ref: < 31 years group)				
≥ 31 and < 43	307	- 0.108***	- 0.053*	- 0.093**
≥ 43 and < 54	302	- 0.161***	- 0.097**	- 0.112**
≥ 54	303	- 0.243***	- 0.167***	- 0.129**
Education level (Ref: none or primary school level)				
Secondary school	512	0.032	0.061	0.048
College degree	102	0.039	0.068*	0.037
≥ Bachelor's degree	172	0.059	0.076*	- 0.021
Household income (baht/month) (Ref: low: 0–30,000 baht/month)				
High: ≥ 30,000 baht/month	89	0.030	0.010	0.049
Smoking (Ref: non-smokers)				
Smokers	335	- 0.022	0.018	0.036
Alcohol (Ref: non-drinkers)				
Drinkers	452	0.010	- 0.030	0.005
Reported comorbidities (Ref: none)				
1–2 comorbidities	312	- 0.185***	- 0.274***	- 0.195***
3–4 comorbidities	41	- 0.176***	- 0.240***	- 0.176***
5 or more comorbidities	3	- 0.065*	- 0.066*	- 0.007

Ref reference value

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 7 Comparison of test–retest reliability for the EQ-5D descriptive system and the utility index between 3L and 5L

Dimensions	Weighted kappa coefficient (95% CI)		Percentage agreement	
	3L	5L	3L	5L
Mobility (MO)	0.63 (0.51–0.76)	0.61 (0.49–0.72)	0.90	0.85
Self-care (SC)	n/a	n/a	0.97	0.97
Usual activities (UA)	0.60 (0.48–0.72)	0.48 (0.35–0.60)	0.95	0.93
Pain/discomfort (PD)	0.58 (0.45–0.70)	0.51 (0.38–0.63)	0.81	0.75
Anxiety/depression (AD)	0.42 (0.29–0.55)	0.54 (0.41–0.67)	0.90	0.86
Intraclass correlation coefficient (ICC)				
EQ-5D index	0.78 (0.71–0.83)	0.71 (0.63–0.78)		
EQ-VAS	0.82 (0.77–0.86)			

n/a non-assessment

42, 48]. Moreover, known-groups revealed that smokers and drinkers had higher utility scores than their counterparts for 5L and 3L, respectively. Similar to a previous study, they revealed that smokers and drinkers reported more health problems than non-smokers and non-drinkers on the two bolt-on dimensions, interpersonal relationships, and activities related to bending knees on the EQ-5D-5L among Thai diabetic patients [49], while another previous study showed that smokers and drinkers reported higher scores on the SF-36v2 in the general Thai population [50]. This might

be due to the Thai population's specific characteristics, so these associations should be reinvestigated through further research.

Regarding reliability, 3L and 5L indexes and EQ-VAS showed good to excellent reproducibility, and all five dimensions produced good reproducibility for both EQ-5D versions. Compared with studies by Pattanaphesaj et al. [22] and Sakthong et al. [23], our weighted kappa values were similar or slightly higher, while our ICCs of the EQ-5D index and EQ-VAS were slightly higher than Pattanaphesaj

et al., but were less than those reported in Sakthong et al. Moreover, Pattanaphesaj et al. and our study similarly reported that SC was not computed due to the high ceiling effect for both versions, resulting in lack of variance in our dataset. A possible explanation is that our study was conducted in the general Thai samples with limited range of health states, while Pattanaphesaj et al. conducted their study in diabetic patients without complications. This contrasted with the sample reported in Sakthong et al., as they reported the weighted kappa coefficient of SC with the value of 0.57 (95% CI 0.44–0.70) since they conducted the study in Thai patients with many chronic diseases and different levels of impairments.

Moreover, our study showed that the values of weighted kappa and ICCs for 5L were lower than those for the 3L, indicating that the 5L seemed less reliable than the 3L. These resembled that reported in Pattanaphesaj et al. [22]. However, these findings contrasted with those reported by Kim et al. [19], Corner-Spady et al. [47], and Jia et al. [38]. We explained that the long time (14–21 days) between the two assessments might contribute to recall bias for the respondents to judge whether their health status had changed after the first assessments. Furthermore, some respondents (40%) assigned to complete the second set of questionnaire were unhealthy, so their health status might have changed during this long time interval. Previous evidence has also suggested that approximately 2 weeks were considered the appropriate time interval for the retest reliability [51, 52]. Therefore, further studies investigating samples' reliability with a shorter time interval are warranted.

One limitation that should be addressed is our time interval was 2–3 weeks for evaluating test–retest reliability, and results were inconsistent with findings reported in other studies. Therefore, future research investigating the effect of various time intervals on test–retest reliability is greatly encouraged.

Conclusion

Evidence supported that the 5L had an acceptable level of validity and reliability in the general Thai population. In addition, we found that the 5L was slightly better than the 3L in ceiling effect, discriminatory power and in convergent validity, while it showed comparable known-groups validity with the 3L. However, evidence to distinguish the superiority of the 5L over the 3L for test–retest reliability was limited. To confirm our results, therefore, it should be reinvestigated with a larger number of subjects having various levels of health impairment.

Acknowledgements This work was financially supported by the Research Grant of Burapha University through National Research

Council of Thailand (Grant No. Rx2/2562). However, the results and opinions in this report have not been endorsed by the above funding agency or elsewhere. We would like to express our gratitude to the EuroQoL group, OPTUM and Suan Prung Psychiatric Hospital's director, Thailand for allowing us to use both EQ-5D versions, Thai SF-12v2 and WHOQOL-BREF-THAI, respectively. We would also like to thank all local village leaders and participants from the provinces of Nakhon-Srithammarat, Khon-Kaen, Chonburi, Chaing-Mai and Bangkok in Thailand who facilitated or participated in the data collection process. Special thanks to all trained interviewers for assistance with the interview process.

Author contributions Krittaphas Kangwanrattanakul (KK) was only involved in conception, study design, data collection, data analyses, interpretation and drafting of this manuscript, while Porntip Parmon-tree (PP) rechecked the data analyses and interpretation. However, all authors were involved in the final approval of this manuscript.

Funding This work was financially supported by the Research Grant of Burapha University through National Research Council of Thailand under Grant No. Rx2/2562.

Data availability The data analyzed and reported in this manuscript is not available for public sharing.

Compliance with ethical standards

Conflict of interest All authors declare that they have no competing interests in this study.

Ethical approval This work was approved by the Burapha University Institutional Review Board (BUU-IRB): 108/2562 before the study commenced.

Informed consent The written consent form was obtained from each participant before the study commenced; however, they were informed to be able to withdraw from this study at any time if they felt uncomfortable.

References

1. 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(10), 1727–1736.
2. Oemar, M., & Janssen, B. (2013). *EQ-5D-5L user guide-basic information on how to use the EQ-5D-5L instrument* (p. 28). Rotterdam: EuroQol Group.
3. Brooks, R. (1996). EuroQol: The current state of play. *Health Policy*, 37(1), 53–72.
4. EuroQoL Group. (1990). EuroQol—A new facility for the measurement of health-related quality of life. *Health Policy*, 16(3), 199–208.
5. Rabin, R., & De Charro, F. (2001). EQ-5D: A measure of health status from the EuroQol Group. *Annals of Medicine*, 33(5), 337–343.
6. Janssen, M. F., Pickard, A. S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., et al. (2013). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Quality of Life Research*, 22(7), 1717–1727.
7. Kim, T. H., Jo, M. W., Lee, S. I., Kim, S. H., & Chung, S. M. (2013). Psychometric properties of the EQ-5D-5L in the general

- population of South Korea. *Quality of Life Research*, 22(8), 2245–2253.
8. König, H. H., Bernert, S., Angermeyer, M. C., Matschinger, H., Martinez, M., Vilagut, G., et al. (2009). Comparison of population health status in six European countries: results of a representative survey using the EQ-5D questionnaire. *Medical Care*, 47(2), 255–261.
 9. Rawlins, M. D., & Culyer, A. J. (2004). National Institute for Clinical Excellence and its value judgments. *British Medical Journal*, 329(7459), 224–227.
 10. Weinstein, M. C., Siegel, J. E., Gold, M. R., Kamlet, M. S., & Russell, L. B. (1996). Recommendations of the panel on cost-effectiveness in health and medicine. *JAMA*, 276(15), 1253–1258.
 11. Sakthong, P. (2008). Measurement of clinical-effect: Utility. *Journal of the Medical Association of Thailand*, 91(Suppl 2), S43–52.
 12. Johnson, J. A., & Coons, S. J. (1998). Comparison of the EQ-5D and SF-12 in an adult US sample. *Quality of Life Research*, 7(2), 155–166.
 13. Petrou, S., & Hockley, C. (2005). An investigation into the empirical validity of the EQ-5D and SF-6D based on hypothetical preferences in a general population. *Health Economics*, 14(11), 1169–1189.
 14. McCrone, P., Patel, A., Knapp, M., Schene, A., Koeter, M., Amadeo, F., et al. (2009). A comparison of SF-6D and EQ-5D utility scores in a study of patients with schizophrenia. *The Journal of Mental Health Policy and Economics*, 12(1), 27–31.
 15. Petrou, S., Morrell, J., & Spiby, H. (2009). Assessing the empirical validity of alternative multi-attribute utility measures in the maternity context. *Health Qual Life Outcomes*, 7, 40.
 16. Kularatna, S., Senanayake, S., Gunawardena, N., & Graves, N. (2019). Comparison of the EQ-5D-3L and the SF-6D (SF-36) contemporaneous utility scores in patients with chronic kidney disease in Sri Lanka: a cross-sectional survey. *British Medical Journal Open*, 9(2), e024854.
 17. Pickard, A. S., De Leon, M. C., Kohlmann, T., Cella, D., & Rosenbloom, S. (2007). Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. *Medical Care*, 45(3), 259–263.
 18. Scalone, L., Ciampichini, R., Fagioli, S., Gardini, I., Fusco, F., Gaeta, L., et al. (2013). Comparing the performance of the standard EQ-5D-3L with the new version EQ-5D-5L in patients with chronic hepatic diseases. *Quality of Life Research*, 22(7), 1707–1716.
 19. Kim, S. H., Kim, H. J., Lee, S. I., & Jo, M. W. (2012). Comparing the psychometric properties of the EQ-5D-3L and EQ-5D-5L in cancer patients in Korea. *Quality of Life Research*, 21(6), 1065–1073.
 20. Craig, B. M., Pickard, A. S., & Lubetkin, E. I. (2014). Health problems are more common, but less severe when measured using newer EQ-5D versions. *Journal of Clinical Epidemiology*, 67(1), 93–99.
 21. Feng, Y., Devlin, N., & Herdman, M. (2015). Assessing the health of the general population in England: how do the three- and five-level versions of EQ-5D compare? *Health Qual Life Outcomes*, 13, 171.
 22. Pattanaphesaj, J., & Thavorncharoensap, M. (2015). Measurement properties of the EQ-5D-5L compared to EQ-5D-3L in the Thai diabetes patients. *Health and Quality of Life Outcomes*, 13, 14.
 23. Sakthong, P., Sonsa-Ardjit, N., Sukarnjanaset, P., & Munpan, W. (2015). Psychometric properties of the EQ-5D-5L in Thai patients with chronic diseases. *Quality of Life Research*, 24(12), 3015–3022.
 24. Pattanaphesaj, J., Thavorncharoensap, M., Ramos-Goni, J. M., Tongsiri, S., Ingsrisawang, L., & Teerawattananon, Y. (2018). The EQ-5D-5L valuation study in Thailand. *Expert Review of Pharmacoeconomics & Outcomes Research*, 18(5), 551–558.
 25. Tongsiri, S., & Cairns, J. (2011). Estimating population-based values for EQ-5D health states in Thailand. *Value Health*, 14(8), 1142–1145.
 26. The World Health Organization Quality of Life Group. (1998). The World Health Organization quality of life assessment (WHOQOL): development and general psychometric properties. *Social Science & Medicine*, 46(12), 1569–1585.
 27. Mahatnirundkul, S. (1998). Comparison of the WHOQOL-100 and the WHOQOL-BREF (26 items). *Journal of Mental Health of Thailand*, 5, 4–15.
 28. Ware, J. E., Jr., Kosinski, M., & Keller, S. D. (2002). *SF-12: how to score the SF-12 physical and mental health summary scales*. Boston: Health Assessment Lab, QualityMetric Inc.
 29. Chariyalertsak, S., Wansom, T., Kawichai, S., Ruangyuttikarna, C., Kemerer, V., & Wu, A. (2011). Reliability and validity of Thai versions of the MOS-HIV and SF12 quality of life questionnaires in people living with HIV/AIDS. *Health and Quality of Life Outcomes*, 9, 1–9.
 30. Phantipa, S., Vijj, K., & Win, W.-W. (2017). Assessment of health-related quality of life in Thai patients after heart surgery. *Asian Biomedicine*, 9(2), 203–210.
 31. Ware, J., Kosinski, M., Turner-Bowker, D., & Gandek, B. (2002). How to score SF-12 items *SF-12 v2 How to Score Version 2 of the SF-12 Health Survey*, 29–38.
 32. Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60(1), 34–42.
 33. Kang, E. J., & Ko, S. K. (2009). A catalogue of EQ-5D utility weights for chronic diseases among noninstitutionalized community residents in Korea. *Value Health*, 12(Suppl 3), S114–117.
 34. Janssen, M. F., Birnie, E., Haagsma, J. A., & Bonsel, G. J. (2008). Comparing the standard EQ-5D three-level system with a five-level version. *Value in Health*, 11(2), 275–284.
 35. Cohen, P. (1974). *Regression and correlation*. In *Statistic in medicine*. Boston: Little Brown and Company.
 36. Fleiss, J. L., Levin, B., & Paik, M. C. (1981). The measurement of interrater agreement. *Statistical Methods for Rates and Proportions*, 2(212–236), 22–23.
 37. Golicki, D., Niewada, M., Karlinska, A., Buczek, J., Kobayashi, A., Janssen, M. F., et al. (2015). Comparing responsiveness of the EQ-5D-5L, EQ-5D-3L and EQ-VAS in stroke patients. *Quality of Life Research*, 24(6), 1555–1563.
 38. Jia, Y. X., Cui, F. Q., Li, L., Zhang, D. L., Zhang, G. M., Wang, F. Z., et al. (2014). Comparison between the EQ-5D-5L and the EQ-5D-3L in patients with hepatitis B. *Quality of Life Research*, 23(8), 2355–2363.
 39. Poor, A. K., Rencz, F., Brodzky, V., Gulacsi, L., Beretzky, Z., Hidvegi, B., et al. (2017). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L in psoriasis patients. *Quality of Life Research*, 26(12), 3409–3419.
 40. Ferreira, L. N., Ferreira, P. L., Ribeiro, F. P., & Pereira, L. N. (2016). Comparing the performance of the EQ-5D-3L and the EQ-5D-5L in young Portuguese adults. *Health and Quality of Life Outcomes*, 14, 89.
 41. Buchholz, I., Janssen, M. F., Kohlmann, T., & Feng, Y. S. (2018). A systematic review of studies comparing the measurement properties of the three-level and five-level versions of the EQ-5D. *Pharmacoeconomics*, 36(6), 645–661.
 42. Yfantopoulos, J., Chantzaras, A., & Kontodimas, S. (2017). Assessment of the psychometric properties of the EQ-5D-3L and EQ-5D-5L instruments in psoriasis. *Archives of Dermatological Research*, 309(5), 357–370.
 43. Pan, C. W., Sun, H. P., Wang, X., Ma, Q., Xu, Y., Luo, N., et al. (2015). The EQ-5D-5L index score is more discriminative than

- the EQ-5D-3L index score in diabetes patients. *Quality of Life Research*, 24(7), 1767–1774.
44. Greene, M. E., Rader, K. A., Garellick, G., Malchau, H., Freiberg, A. A., & Rolfson, O. (2015). The EQ-5D-5L improves on the EQ-5D-3L for health-related quality-of-life assessment in patients undergoing total hip arthroplasty. *Clinical Orthopaedics and Related Research*, 473(11), 3383–3390.
 45. Buchholz, I., Thielker, K., Feng, Y. S., Kupatz, P., & Kohlmann, T. (2015). Measuring changes in health over time using the EQ-5D-3L and 5L: A head-to-head comparison of measurement properties and sensitivity to change in a German inpatient rehabilitation sample. *Quality of Life Research*, 24(4), 829–835.
 46. Martí-Pastor, M., Pont, A., Ávila, M., Garin, O., Vilagut, G., Forero, C. G., et al. (2018). Head-to-head comparison between the EQ-5D-5L and the EQ-5D-3L in general population health surveys. *Population Health Metrics*, 16(1), 14.
 47. Conner-Spady, B. L., Marshall, D. A., Bohm, E., Dunbar, M. J., Loucks, L., Al Khudairy, A., et al. (2015). Reliability and validity of the EQ-5D-5L compared to the EQ-5D-3L in patients with osteoarthritis referred for hip and knee replacement. *Quality of Life Research*, 24(7), 1775–1784.
 48. Yfantopoulos, J. N., & Chantzaras, A. E. (2017). Validation and comparison of the psychometric properties of the EQ-5D-3L and EQ-5D-5L instruments in Greece. *The European Journal of Health Economics*, 18(4), 519–531.
 49. Kangwanrattanakul, K., Gross, C. R., Sunantiwat, M., & Thavorncharoensap, M. (2019). Adding two culture-specific 'bolt-on' dimensions on the Thai version of EQ-5D-5L: an exploratory study in patients with diabetes. *Expert Rev Pharmacoecon Outcomes Res*, 19(3), 321–329.
 50. Kangwanrattanakul, K., & Auamnoy, T. (2019). Psychometric testing of the health-related quality of life measurement, SF-36v2, in the general population of Thailand. *Expert Review of Pharmacoeconomics & Outcomes Research*, 19(3), 313–320.
 51. Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: a practical guide to their development and use*. New York: Oxford University Press.
 52. Marx, R. G., Menezes, A., Horovitz, L., Jones, E. C., & Warren, R. F. (2003). A comparison of two time intervals for test-retest reliability of health status instruments. *Journal of Clinical Epidemiology*, 56(8), 730–735.

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