Comparing the psychometric properties of the EQ-5D-3L and EQ-5D-5L in cancer patients in Korea

Seon Ha Kim · Hwa Jung Kim · Sang-il Lee ·
Min-Woo Jo

Abstract

Purpose This study was conducted to assess the redistribution properties of the EQ-5D-3L when using the EQ-5D-5L and to compare the validity, informativity, and reliability of both EQ-5D-3L and EQ-5D-5L in Korean cancer patients.
Methods Patients visiting one ambulatory cancer center self-administered the two versions of the EQ-5D and the EORTC QLQ-C30 questionnaire. Redistribution properties in each dimension of EQ-5D were analyzed between EQ-5D-3L and EQ-5D-5L. Informativity was evaluated using the Shannon entropy and ceiling effect. Convergent validity was evaluated by comparing the EQ-VAS, ECOG performance status, and EORTC QLQ-C30 subscales. Reliability was also evaluated in terms of test–retest reliability.
Results All levels of the EQ-5D-3L substantially partitioned into associated levels of the EQ-5D-5L. The average inconsistency rate of the two versions was 3.5%. Absolute informativity was higher for the EQ-5D-5L than for the EQ-5D-3L, but their informative efficiency tended to be similar. The proportion of 'perfect health' (11111) decreased from 16.8% in the EQ-5D-3L to 9.7% in the EQ-5D-5L. EQ-5D-5L demonstrated similar or higher correlations with the EQ-VAS, ECOG performance status, and EORTC QLQ-C30, than the EQ-5D-3L. The intraclass correlation coefficient of the EQ-5D-5L index was 0.77.
Conclusions The EQ-5D-5L had greater informativity and lower rate in the ceiling effect than those values of the EQ-5D-3L. The EQ-5D-5L showed good construct validity and reasonable reliability. Therefore, considering these findings, the EQ-5D-5L may be preferable to the EQ-5D-3L.

Keywords EQ-5D · Health-related quality of life · Cancer · Psychometrics

Abbreviations

ECOG	Eastern Cancer Oncology Group
EQ-5D-3L	3-level version of EQ-5D
EQ-5D-5L	5-level version of EQ-5D
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30
EU ISPOR	European Union International Society for Pharmacoeconomics and Outcomes Research
HRQoL	Health-related quality of life
HUI2	Health utility index mark 2
HUI3	Health utility index mark 3
ICC	Intraclass correlation coefficient
SF-6D	Short form 6 dimension
VAS	Visual analogue scale

S. H. Kim · S. Lee · M.-W. Jo (✉)
Department of Preventive Medicine, University of Ulsan College
of Medicine, 86, Asanbyeongwon-gil, Songpa-gu,
Seoul 138-736, Korea
e-mail: jominwoo@amc.seoul.kr

H. J. Kim
Department of Clinical Epidemiology and Biostatistics,
Asan Medical Center, 86, Asanbyeongwon-gil, Songpa-gu,
Seoul 138-736, Korea

Introduction

The EQ-5D-3L, the original version of the EQ-5D, is an instrument widely used to measure and evaluate general health status [1–4]. The EQ-5D-3L descriptive system

describes general health in terms of five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has three levels, indicating no problems, some or moderate problems, and extreme problems, resulting in a total of 243 (i.e., 3^5) unique health states. EQ-5D-3L provides a simple descriptive profile and a single index for health status that can be used in the clinical and economic evaluation of health care, as well as population health surveys [5].

The EQ-5D-3L has good psychometric properties and is able to detect small changes in chronic diseases [6, 7]. However, there has been some evidence that the SF-6D derived from SF-36 was more discriminative than the EQ-5D-3L index [8, 9], although Cunillera et al. [10] showed that the SF-6D derived from SF-12 was less discriminative than the EQ-5D-3L index. In addition, it lacks descriptive richness, when compared with other generic preference-based instruments, including the Health Utilities Index Mark 2 and 3 (HUI2 and HUI3) and the Short Form 6D (SF-6D), which define 24,000, 972,000, and 18,000 unique health states, respectively [11]. Moreover, EQ-5D-3L suffers from ceiling effects [8, 9, 12].

An expanded descriptive system, with more response categories (i.e., levels) per dimension, may improve the ability of the EQ-5D to reliably discriminate among different levels of health and to detect changes in health [13]. The EuroQol Group designed a new questionnaire, the EQ-5D-5L version (i.e., the level 5 of EQ-5D), to improve the sensitivity and reduce the ceiling effects of the EQ-5D-3L version. At present, versions of the EQ-5D-5L in 57 languages are available at the EuroQol website (http://www.euroqol.org), and studies have compared the psychometric properties of the EQ-5D-3L and EQ-5D-5L [11, 13]. However, these studies used a prototype version of the EQ-5D-5L, not the official version.

We therefore aimed to assess the response redistribution of EQ-5D-3L when using the EQ-5D-5L. In addition, we compared the psychometric properties of the EQ-5D-3L and EQ-5D-5L from the perspective of informativity, validity, and reliability.

Methods

Subjects and settings

This study was approved by the Institutional Review Board of Asan Medical Center (approval number: 2010-0546), and all participants provided written informed consent.

We consecutively recruited 901 cancer patients aged over 18 years who were receiving chemotherapy at an ambulatory cancer center in Korea, over 1-month period.

Patients were excluded if they had a performance status of 4 on the ECOG Scale, or if there were missing or duplicate responses on the EQ-5D-3L or EQ-5D-5L questionnaire. Participants filled in the questionnaire just before or during chemotherapy.

To assess reliability, 250 patients conveniently selected from the first survey subjects were asked to retest brief questionnaires, which included EQ-5D-3L and EQ-5D-5L, but not EORTC QLQ-C30, at 1–4 weeks interval, and return them by mail.

Information

ECOG performance scale was evaluated by one research nurse. General characteristics, including gender, age, and clinical information, were obtained from the cancer registry in the center.

Every participant completed the Korean versions of the EORTC QLQ-C30, the EQ-5D-5L, and the EQ-5D-3L, consecutively. The Korean version of EQ-5D-3L was previously validated [14]. The EQ-5D-5L used in this study was the official version provided by EuroQol. Its dimensions are the same as those of the EQ-3D-3L, but include five response levels. The descriptors for levels 1, 3, and 5 on the EQ-5D-5L were similar to the wording for levels 1, 2, and 3, respectively, on the EQ-5D-3L, but not identical. For instance, 'some' of level 2 in EQ-5D-3L changed to 'moderate' or 'moderately' of level 3 in EQ-5D-5L. Level 3 in the mobility domain of EQ-5D-3L was described as 'I am confined to bed', whereas level 5 in the mobility domain of EQ-5D-5L was described as 'I am unable to walk about'. Level 2 on the EQ-5D-5L was labeled 'slightly' for anxiety/depression and 'slight' for the remaining four dimensions. Level 4 on the EQ-5D-5L was labeled 'severely' for anxiety/depression and 'severe' for the other 4 dimensions. Further detailed comparisons between the two EQ-5D instruments are available at http://www.euroqol.org.

The EQ-5D-3L index was calculated using the valuation set from Lee et al. [15], whereas the EQ-5D-5L index was calculated by applying the indirect interim mapping method presented by the EuroQoL group at the 13th EU ISPOR meeting. According to the files provided by the EuroQoL group, EQ-5D-5L Crosswalk Project was conducted in 3,691 patients across six countries. Using these data, they obtained the transition probability matrix between the EQ-5D-3L and EQ-5D-5L states. Then, the EQ-5D-5L index was calculated by summing each of moving probability from EQ-5D-3L states to specific EQ-5D-5L states multiplied by each utility weight in EQ-5D-3L states expressed by the following equation:

EQ - 5D - 5L$_{21243}$

$$= \sum_{a=1}^{3} \sum_{b=1}^{3} \sum_{c=1}^{3} \sum_{d=1}^{3} \sum_{e=1}^{3} (P_{3L\_abcde \rightarrow 5L\_21243} \times U_{3L\_abcde})$$

$P_{3L\_abcde \rightarrow 5L\_21243}$: transition probability from the 'abcde' state in EQ-5D-3L to '21243' state in EQ-5D-5L

U_{3L_abcde}: utility weight in the 'abcde' state in EQ-5D-3L

The EORTC QLQ-C30 is an integrated system for assessing the HRQoL of cancer patients. It includes five functional scales (i.e., physical, role, emotional, cognitive, and social), three symptom scales, a global health status, and a number of single items [16].

All HRQoL instruments were self-administered; if necessary, the research nurse assisted participants in completing the questionnaires. Respondents could change their answer before submitting it to the research nurse.

Analyses

Response redistribution

On any item in the EQ-5D-5L or EQ-5D-3L, missing or multiple answers were excluded from analyses. Redistribution properties were described as proportions of the 3L-5L response pairs within each 3L response level, and the corresponding mean and median VAS scores were calculated for each subgroup of paired responses, except for inconsistent pairs. Inconsistency and its size were defined as in Janssen et al. [11]. Briefly, after projecting the 3L response scale on a 5L response scale (i.e., producing $3L_{5L}$ by recoding $1 = 1$, $2 = 3$, $3 = 5$), the size of inconsistency was calculated as $|3L_{5L} - 5L| - 1$. If inconsistency size was 0 or less than zero, it was considered as consistency. For example, when level 1 in EQ-5D-3L was redistributed as level 1 or 2 in EQ-5D-5L, it was considered as consistent response. However, if the response was redistributed as level 3, 4, or 5 in EQ-5D-5L, it was considered as inconsistent response and the size of inconsistency was 1, 2, and 3, respectively.

Informativity

The ceiling effect was calculated as the proportion of 'no problem' responses on each dimension and the proportion of 'no problem' in all dimensions. Reduction in the ceiling effect suggested enhancement of discriminant ability. Informativity was assessed using the Shannon entropy (Shannon index) and the information efficiency (Shannon evenness index) [17, 18].

The Shannon entropy is calculated as

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

where H' is the absolute amount of informativity captured, C is the number of possible categories (or levels in this study), and $P_i = n_i/N$ is the proportion of observations in the ith category ($i = 1,...,C$), where n_i is the observed number of scores (responses) in category i and N is the total sample size. In case of an even distribution (i.e., if all levels are evenly filled), the optimal amount of information is captured and the Shannon entropy has reached its upper limit (H'max), as represented by the formula: H'max $= \log_2 C$, which amounts to 1.58 on the EQ-5D-3L, and to 2.32 on the EQ-5D-5L. H'max increases as the number of levels increases. Nevertheless, the empirical informativity H' will increase only if the newly added categories are actually used. The Shannon entropy combines the number of categories defined by a system, as a measure for the extent to which the information is evenly spread over the categories. The information efficiency reflects the evenness of a distribution, regardless of the number of levels. The information efficiency measure $J' = H'/H'$max describes the use of a system (H'), given its potential (H'max). The Shannon entropy H' can be therefore considered an expression of the absolute informativity of a system, whereas the information efficiency J' expresses only the relative informativity of a system, regardless of the number of categories [4].

Validity

Convergent validity of the EQ-5D-5L and EQ-5D-3L was examined by comparing the EQ-VAS score, ECOG performance status, and EORTC QLQ-C30 subscales within each dimension. We assumed that each dimension in the both EQ-5D instruments would be more highly correlated to the related subscales than to other subscales in the EORTC QLQ-C30 (e.g., mobility in the EQ-5D is more likely to correlate with physical function than with emotional function in the EORTC QLQ-C30). We also hypothesized that the correlation between the EQ-5D-5L and other measures would be similar to or higher than the correlation between the EQ-5D-3L and other measures. These assumptions were examined by the Spearman's rank correlation coefficient. Fisher's T transformations were utilized to determine whether correlations between EQ-5D versions and other instruments differed significantly [19].

For known-group construct validity, both EQ-5D-3L and EQ-5D-5L indexes were calculated by performance status, age-group, and VAS score quartile. We assumed

that both EQ-5D-3L and EQ-5D-5L indexes would be lower in groups with higher ECOG performance score, in older than in younger patients, and in higher VAS score group than low VAS score group.

Test–retest reliability

The agreement of each dimension on the EQ-5D instruments was evaluated by kappa and weighted kappa statistics. The larger the number of scale categories, the greater the potential for disagreement. Therefore, we calculated weighted kappa statistics in parallel, which uses weights to quantify the relative difference between categories.

We applied Fleiss's standards for strength of agreement for the kappa values, as follows: <0.4 = poor, 0.4–0.75 = fair to good, >0.75 = excellent [20]. Test–retest reliability of both EQ-5D indexes was evaluated by the intraclass correlation coefficient (ICC, two-way random effects, absolute agreement).

All statistical analyses were performed using the SAS software version 9.1 (SPSS Inc., Chicago, IL), and the differences were considered statistically significant if P-value was less than 0.05.

Results

Subjects

A total 2,316 visits for chemotherapy were recorded during the research period, except repeated visits. A research nurse asked the patients to participate in the study. If they consented to join the study, they were asked to fill out the questionnaire. In total, 901 questionnaires were collected. Out of 901, four cases were duplicates, three had missing or duplicate answers on the EQ-5D-5L or EQ-5D-3L, and one case missed the identification number. Thus, our final analysis set consisted of 893 patients. The final response rate was 38.5% (893/2316). The difference in the distribution of gender and age-group was not statistically different between candidates and responders ($P = 0.054$ and $P = 0.10$, respectively, data not shown). The mean age of the subjects was 53.0 (SD $\pm$ 11.2) years, with 56.8% women. These subjects had 30 different types of cancer, the most frequent being breast (32.9%) and colorectal (20.0%) cancer.

In the second survey, 81 out of 250 patients responded to questionnaires by mail, but three questionnaires were not usable because of missing data. Therefore, responses from 78 subjects were used to analyze reliability. Their mean age was 53.9 (SD $\pm$ 10.9) years, and 56.4% were women (Table 1). The mean time interval between the initial and follow-up surveys was 11.5 days (IQR 6–15 days).

Table 1 General characteristics of the study subjects

Characteristics	First survey N (%)	Second survey N (%)
Age (mean ± SD), years	53.0 (11.2)	53.9 (10.9)
Gender		
Male	386 (43.2)	34 (43.6)
Female	507 (56.8)	44 (56.4)
Performance status (ECOG)		
0	785 (87.9)	72 (92.3)
1	86 (9.6)	4 (5.1)
2+	22 (2.5)	2 (2.7)
Type of cancer		
Breast	291 (32.9)	24 (31.2)
Colorectal	121 (13.7)	19 (24.7)
Lung	98 (11.1)	12 (15.6)
Stomach	89 (10.1)	3 (3.9)
Pancreas	39 (4.4)	4 (5.2)
Others	190 (21.5)	15 (19.5)
History of surgery		
Yes	343 (38.4)	26 (33.3)
Duration from initial diagnosis		
Less than 1 year	435 (53.6)	37 (50.7)
1 year to less than 2 years	169 (20.8)	12 (16.4)
2 years to less than 3 years	111 (13.7)	14 (19.2)
3 years to less than 4 years	44 (5.4)	3 (4.1)
4 years and more	52 (6.4)	7 (9.6)

Response redistribution

Table 2 shows the proportions of the EQ-5D-3L and EQ-5D-5L response pairs within each 3L response level, and the mean and median VAS values for each subgroup with consistent responses. The mean and median VAS values tended to decrease when the 3L-5L response pairs in each dimension increased from $3L_1$ to $5L_1$ (subjects who selected level 1 in EQ-5D-3L and level 1 in EQ-5D-5L) to $3L_3$–$5L_5$ (subjects who selected level 3 in EQ-5D-3L and level 5 in EQ-5D-5L), that is, from the most to the least healthy subgroup. There was substantial partitioning of level 2 in the EQ-5D-3L, and the majority of level 3 responses in all dimensions of the EQ-5D-3L moved to level 4 of the EQ-5D-5L. The proportion of inconsistency in dimensions ranged from 2.4% for anxiety/depression to 4.5% for usual activities. The average size of inconsistency was highest (1.21) for self-care and lowest (1.06) for pain/discomfort (Table 3).

Informativity

We found that 150 respondents (16.8%) on the EQ-5D-3L and 87 (9.7%) on the EQ-5D-5L reported no problems on

Dimension	3L	5L	N	% by 3L level	VAS	
					Mean	Median
Mobility	1	1	473	83.1	71.4	75
		2	96	16.9	65.1	70
	2	2	179	61.7	60.2	60
		3	79	27.2	53.5	50
		4	32	11.0	41.1	40
	3	4	1	100.0	10.0	10
Self-care	1	1	692	93.8	68.4	70
		2	46	6.2	53.5	50
	2	2	84	67.7	57.0	60
		3	35	28.2	46.8	50
		4	5	4.0	49.0	50
	3	4	2	66.7	50.0	50
		5	1	33.3	0.0	0
Usual activities	1	1	271	71.7	78.4	80
		2	107	28.3	68.7	70
	2	2	298	64.6	60.5	60
		3	133	28.9	55.3	50
		4	30	6.5	51.1	50
	3	4	8	57.1	35.0	35
		5	6	42.9	15.8	17.5
Pain/discomfort	1	1	228	72.4	76.4	80
		2	87	27.6	73.0	76
	2	2	384	74.0	63.4	70
		3	111	21.4	53.7	50
		4	24	4.6	50.2	50
	3	4	22	84.6	42.5	40
		5	4	15.4	63.8	60
Anxiety/depression	1	1	309	80.5	75.9	80
		2	75	19.5	66.5	70
	2	2	344	72.9	61.8	60
		3	117	24.8	52.4	50
		4	11	2.3	56.8	60
	3	4	13	81.3	37.3	40
		5	3	18.8	16.7	20

Table 2 Response redistribution from the EQ-5D-3L to the EQ-5D-5L, by dimension and by level in consistent responses

all dimensions. Eighty out of 150 respondents answered to the 11111 item of health state in EQ-5D-5L as well, but other 70 patients answered to other health states in EQ-5D-5L. The mean VAS score in the former group was 84.3 and that of the latter group was 78.1; the difference was significantly different ($P = 0.007$). In both EQ-5D instruments, the self-care dimension showed the highest ceiling effect, whereas pain/discomfort showed the lowest ceiling effect. The proportions of respondents reporting 'no problems' across dimensions decreased on the EQ-5D-5L, compared with the EQ-5D-3L. The mobility domain showed the most reduction from 65.1% in EQ-5D-3L to 54.8% in EQ-5D-5L (Table 4). The difference in the ceiling effect between EQ-5D-3L and EQ-5D-5L was statistically significant in all domains, except self-care.

Table 5 shows the informativity results of the EQ-5D-3L and EQ-5D-5L. The EQ-5D-5L consistently showed higher informative quantity (Shannon entropy) than the EQ-5D-3L, with an average difference of 0.48. Information efficiency (J') in EQ-5D-5L was improved in the mobility and usual activity domains, whereas it was declined in the self-care and anxiety/depression domains, compared with EQ-5D-3L. The percentage gain of information efficiency ranged from −9.1% (self-care) to 8.5% (mobility).

Table 3 Inconsistent responses distributed from the EQ-5D-3L to the EQ-5D-5L

Dimension	Inconsistencies N (%)	Average size of inconsistency
Mobility	33 (3.7)	1.09
Self-care	28 (3.1)	1.21
Usual activities	40 (4.5)	1.08
Pain/discomfort	33 (3.7)	1.06
Anxiety/depression	21 (2.4)	1.14

Table 4 Proportion of 'no problem' responses on EQ-5D-3L and EQ-5D-5L

Dimension	EQ-5D-3L N (%)	EQ-5D-5L N (%)	P value[a]
Mobility	581 (65.1%)	489 (54.8%)	<0.001
Self-care	742 (83.1%)	714 (80.0%)	<0.001
Usual activities	386 (43.2%)	297 (33.3%)	<0.001
Pain/discomfort	323 (33.3%)	244 (27.3%)	<0.001
Anxiety/depression	391 (43.8%)	320 (35.8%)	<0.001
Full health (11111)	150 (16.8%)	87 (9.7%)	<0.001

[a] McNemar test

Validity

Table 6 shows correlations by dimension between EQ-5D, ECOG performance status, and EORTC QLQ-C30 subscales. The correlations between EQ-VAS and EQ-5D-5L responses tended to be slightly stronger than correlations between EQ-VAS and EQ-5D-3L responses, across all dimensions. Similarly, correlations between the EQ-5D-5L and EORTC QLQ-C30 subscales were slightly stronger than those between the EQ-5D-3L and EORTC QLQ-C30 subscales. None of these differences, however, was statistically significant. Spearman's rank correlation coefficients between the EQ-5D-3L and EQ-5D-5L ranged from 0.70 for usual activities to 0.77 for anxiety/depression. The Pearson's correlations between VAS and EQ-5D indexes were 0.52 for the EQ-5D-3L and 0.55 for the EQ-5D-5L (data not shown). In both EQ-5D instruments, mobility was

more highly correlated with physical function than with other subscales on the EORTC QLQ-C30, whereas anxiety/depression was more highly correlated with emotional function of the EORTC QLQ-C30.

Aspects of the known-group construct validity of EQ-5D-3L and EQ-5D-5L indexes by ECOG performance status, age-group, and VAS score quartiles are shown in Table 7. Both EQ-5D indexes had similar values and tended to decline as ECOG scores increased, age increased, and VAS score decreased. The VAS quartile variable explained 20.9 and 25.3% of the EQ-5D-3L index and EQ-5D-5L index, respectively, and the explanatory power increased approximately 5% in both indexes when continuous VAS scale was applied.

Test–retest reliability

Agreements by kappa on both EQ-5D instruments were fair to good in 4 dimensions, but not for usual activities. Kappa statistics on the EQ-5D-5L varied from 0.36 to 0.64 across dimensions. Kappa statistics of the EQ-5D-5L were slightly lower than those of EQ-5D-3L, whereas the weighted kappa of the EQ-5D-5L tended to be slightly higher than that of EQ-5D-3L. The difference in kappa statistics between the two EQ-5D instruments was not statistically significant. ICCs of the EQ-5D-3L and EQ-5D-5L indices were 0.75 and 0.77, respectively (Table 8).

Discussion

We found that the three levels on the EQ-5D-3L were substantially redistributed among the five levels on the EQ-5D-5L, with the majority of level 3 on the EQ-5D-3L rearranged to level 4 on the EQ-5D-5L. The proportion of respondents reporting 'no problem' on the EQ-5D-5L ranged from 27.3% for pain to 80.0% for self-care. Full health (11111) was significantly decreased from 16.8% in EQ-5D-3L to 9.7% in EQ-5D-5L. Ceiling effects on the EQ-5D-5L were still present, but were considerably decreased compared with the EQ-5D-3L, except for the self-care dimension. Although the ceiling effect of EQ-5D-

Table 5 Shannon entropy (H') and information efficiency (J') for the EQ-5D-3L and EQ-5D-5L

	EQ-5D-3L		EQ-5D-5L	
	H' (95% CI)	J' (95% CI)	H' (95% CI)	J' (95% CI)
Mobility	0.93 (0.90–0.97)	0.59 (0.57–0.61)	1.50 (1.43–1.56)	0.64 (0.62–0.67)
Self-care	0.69 (0.62–0.76)	0.44 (0.39–0.48)	0.92 (0.83–1.01)	0.40 (0.36–0.44)
Usual activities	1.11 (0.07–0.16)	0.70 (0.67–0.73)	1.72 (1.66–1.79)	0.74 (0.72–0.77)
Pain/discomfort	1.14 (1.09–1.19)	0.72 (0.69–0.75)	1.67 (1.60–1.73)	0.72 (0.69–0.75)
Anxiety/depression	1.12 (1.07–1.16)	0.71 (0.68–0.73)	1.60 (1.54–1.66)	0.69 (0.66–0.72)

Table 6 Convergent validity of the EQ-5D-5L and EQ-5D-3L by VAS score, ECOG performance status, and QLQ-C30 scale

EQ-5D dimension	EQ-VAS	ECOG status	QLQ-C30 Physical functioning	QLQ-C30 Role functioning	QLQ-C30 Emotional functioning	QLQ-C30 Cognitive functioning	QLQ-C30 Social functioning	QLQ-C30 Pain Scales
Mobility								
3L	−0.32	0.39	−0.51	−0.39	−0.20	−0.30	−0.19	0.35
5L	−0.37	0.39	−0.55	−0.40	−0.22	−0.28	−0.17	0.35
Self-care								
3L	−0.24	0.43	−0.38	−0.34	−0.18	−0.22	−0.18	0.31
5L	−0.28	0.45	−0.41	−0.34	−0.22	−0.23	−0.20	0.33
Usual activities								
3L	−0.46	0.31	−0.51	−0.54	−0.33	−0.30	−0.36	0.42
5L	−0.48	0.38	−0.57	−0.58	−0.39	−0.34	−0.44	0.48
Pain/discomfort								
3L	−0.39	0.25	−0.39	−0.39	−0.35	−0.29	−0.29	0.58
5L	−0.41	0.26	−0.43	−0.38	−0.35	−0.35	−0.29	0.68
Anxiety/depression								
3L	−0.41	0.13	−0.31	−0.36	−0.58	−0.32	−0.34	0.28
5L	−0.44	0.18	−0.35	−0.40	−0.64	−0.37	−0.38	0.32

Spearman's correlation coefficients. All correlation P-values <0.01

Table 7 EQ-5D indexes by ECOG performance status, age-group, and VAS score

	n	EQ-5D-3L index Mean (95% CI)	R^{2*}	EQ-5D-5L index Mean (95% CI)	R^{2*}
ECOG status					
0	785	0.85 (0.84–0.86)	0.261	0.84 (0.83–0.85)	0.237
1	86	0.70 (0.67–0.73)		0.71 (0.69–0.73)	
2–3	22	0.48 (0.39–0.56)		0.58 (0.53–0.64)	
Age-group					
18–29 years	18	0.83 (0.75–0.90)	0.012	0.81 (0.74–0.88)	0.020
30–39 years	82	0.86 (0.83–0.88)		0.85 (0.83–0.87)	
40–49 years	226	0.83 (0.82–0.85)		0.84 (0.82–0.85)	
50–59 years	307	0.82 (0.81–0.84)		0.82 (0.81–0.83)	
60–69 years	203	0.81 (0.79–0.83)		0.81 (0.79–0.82)	
≥70 years	57	0.79 (0.74–0.84)		0.78 (0.75–0.82)	
VAS score					
0–50	295	0.75 (0.73–0.77)	0.209	0.76 (0.75–0.77)	0.253
51–70	259	0.81 (0.80–0.83)		0.81 (0.80–0.82)	
71–80	161	0.87 (0.86–0.88)		0.86 (0.85–0.87)	
81–100	178	0.92 (0.90–0.93)		0.91 (0.90–0.92)	

* R^2 by univariate regression analysis

3L in our research was lower than previously reported in other studies [21–23], the ceiling effect of EQ-5D-5L was more improved than that of EQ-3D-3L.

Not surprisingly, the Shannon entropy was higher for the EQ-5D-5L than for the EQ-5D-3L. The increased Shannon entropy suggested that the EQ-5D-5L was able to better describe various health states and that these expanded levels were empirically used by respondents. The average information efficiencies also improved slightly from 63.1%

in EQ-5D-3L to 63.8% in EQ-5D-5L; however, the effect of information efficiency by domain was diverse. It means that the extent of even distribution was enhanced in EQ-5D-5L, but its impact was different, depending on domain. Our findings, showing that the EQ-5D-5L had greater absolute informativity and lower ceiling effect than the EQ-5D-3L, are consistent with previous results [11, 13].

The proportion of inconsistencies among our respondents averaged 3.5%. This was higher than the average of

Table 8 Test–retest reliability on the EQ-5D-3L and EQ-5D-5L

Dimension	EQ-5D-3L		EQ-5D-5L	
	Kappa (95% CI)	Weighted kappa (95% CI)	Kappa (95% CI)	Weighted kappa (95% CI)
Mobility	0.55 (0.37–0.73)	0.56 (0.38–0.74)	0.54 (0.38–0.70)	0.65 (0.51–0.80)
Self-care	0.66 (0.46–0.86)	0.70 (0.52–0.88)	0.64 (0.46–0.83)	0.69 (0.52–0.87)
Usual activities	0.39 (0.19–0.59)	0.43 (0.23–0.62)	0.36 (0.21–0.52)	0.50 (0.36–0.65)
Pain/discomfort	0.52 (0.34–0.70)	0.55 (0.37–0.73)	0.55 (0.39–0.70)	0.59 (0.43–0.74)
Anxiety/depression	0.53 (0.36–0.70)	0.54 (0.38–0.71)	0.49 (0.33–0.65)	0.55 (0.41–0.69)
		Intraclass correlation coefficient		
EQ-5D index	0.75 (0.63–0.83)		0.77 (0.67–0.85)	

1.1% observed in hypothetical situations when the subjects were familiar with the EQ-5D instrument [11], but lower than 4.3% inconsistent responses reported in a previous study [13]. We found that only one person responded inadequately on the EQ-5D-5L. In comparison, a study performed in Singapore excluded 30 (3.7%) out of 803 patients because of missing values [23] Similarly, the number of missing items from the Welsh study was 4 (0.33%) [24], which was higher than the values on both the EQ-5D-5L (0.02%) and EQ-5D-3L (0.13%) in our study. For convergent validity in Korean cancer patients, the EQ-5D-5L showed stronger correlations with cancer-specific instruments than the EQ-5D-3L; however, the difference was not statistically significant. The association between EQ-5D instruments and other measures showed similar results. For known-group construct validity, we observed decreases in the EQ-5D-5L index by performance status and age-group. The ECOG variable explained 26.1 and 23.7% of the EQ-5D-3L index and EQ-5D-5L index, respectively, when univariate regression analysis was applied. These findings supported previous favorable evidence regarding the validity of the EQ-5D-5L [1, 13]. For example, a study on cervical cancer patients in Taiwan showed similar construct validity [2].

When we assessed reliability, we observed variations in agreement across dimensions. Not surprisingly, when using kappa statistics, the reliability of EQ-5D-3L was slightly better than that of EQ-5D-5L. When using weighted kappa, however, the reliability of EQ-5D-5L was similar or better than that of EQ-5D-3L. The intraclass correlation coefficient was slightly better for the EQ-5D-5L than for the EQ-5D-3L. Compared with the former validation study of the Korean EQ-5D-3L, we found that agreement was slightly decreased, whereas ICC was almost the same [14]. Janssen et al. [11] showed that EQ-5D-5L had generally better inter-observer and test–retest reliability than the EQ-5D-3L. The Taiwan study reported that the ICC for the

EQ-5D-3L was 0.83 and the kappa values for the EQ-5D dimensions ranged from 0.54 to 0.73 [2]. Our values were slightly lower, but still acceptable. Our participants were receiving chemotherapy at the initial survey point, so their condition may have changed during the second survey. In the second survey, we did not collect other information about stability in the subjects' health. Therefore, interpretation of reliability in our study was limited.

This study had several limitations. We examined the redistribution properties of the EQ-5D-5L in cancer patients. We also analyzed additional datasets to examine redistribution properties by type of cancer. Both breast and colorectal cancer patients showed similar distributions of matching pairs. However, it may not be possible to generalize our findings to non-cancer patients, because the proportion of problems reported by cancer patients may differ from those in the general population and in patients with other chronic conditions [22, 25]. In addition, we used experimental interim value sets for the EQ-5D-5L. We intended to recalculate our results when the EQ-5D-5L valuation algorithm was formulated, but the algorithm was in the pre-final stage; therefore, minor change could be made by the EuroQoL group. In the supplementary analysis, when we used the crude summary score transforming 0–100 in both EQ-5D versions, the known-group validity showed similar trends in both instruments and the ICC for the crude summary score in EQ-5D-5L was higher than that of EQ-5D-3L. Further research is required to determine the psychometric properties of the EQ-5D-5L, in particular the ceiling effect, in the general population, and the responsiveness of the measure and the reliability of the EQ-5D-5L in stable cancer patients.

In conclusion, our findings showed that the EQ-5D-5L had greater informativity and lower rate in the ceiling effect than those values of the EQ-5D-3L. Furthermore, the EQ-5D-5L showed good construct validity and reasonable reliability. Therefore, considering these findings, the EQ-5D-5L may be preferable to the EQ-5D-3L.

References

1. Park, S. S., Yoon, Y. S., & Oh, S. W. (2011). Health-related quality of life in metabolic syndrome: The Korea National Health and Nutrition Examination Survey 2005. *Diabetes Research and Clinical Practice, 91*(3), 381–388.
2. Lang, H. C., Chuang, L., Shun, S. C., Hsieh, C. L., & Lan, C. F. (2010). Validation of EQ-5D in patients with cervical cancer in Taiwan. *Supportive Care in Cancer, 18*(10), 1279–1286.
3. Hoi le, V., Chuc, N. T., & Lindholm, L. (2010). Health-related quality of life, and its determinants, among older people in rural Vietnam. *BMC Public Health, 10*, 549.
4. 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–S117.
5. Brooks, R., Rabin, R. E., & de Charro, F. (2003). *The measurement and valuation of health status using EQ-5D: A European perspective*. Dordrecht: Kluwer Academic Publishers.
6. Campbell, H., Rivero-Arias, O., Johnston, K., Gray, A., Fairbank, J., & Frost, H. (2006). Responsiveness of objective, disease-specific, and generic outcome measures in patients with chronic low back pain: an assessment for improving, stable, and deteriorating patients. *Spine (Phila Pa 1976), 31*(7), 815–822.
7. Gusi, N., Tomas-Carus, P., Hakkinen, A., Hakkinen, K., & Ortega-Alonso, A. (2006). Exercise in waist-high warm water decreases pain and improves health-related quality of life and strength in the lower extremities in women with fibromyalgia. *Arthritis and Rheumatism, 55*(1), 66–73.
8. 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*, 1169–1189.
9. 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.
10. Cunillera, O., Tresserras, R., Rajmil, L., Vilagut, G., Brugulat, P., Herdman, M., et al. (2010). Discriminative capacity of the EQ-5D, SF-6D, and SF-12 as measures of health status in population health survey. *Quality of Life Research, 19*(6), 853–864.
11. 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.
12. Bharmal, M., & Thomas, J., I. I. I. (2006). Comparing the EQ-5D and the SF-6D descriptive systems to assess their ceiling effects in the US general population. *Value Health, 9*, 262–271.
13. 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.
14. Kim, M. H., Cho, Y. S., Uhm, W. S., Kim, S., & Bae, S. C. (2005). Cross-cultural adaptation and validation of the Korean version of the EQ-5D in patients with rheumatic diseases. *Quality of Life Research, 14*(5), 1401–1406.
15. Lee, Y. K., Nam, H. S., Chuang, L. H., et al. (2009). South Korean time trade-off values for EQ-5D health states: Modeling with observed values for 101 health states. *Value in Health, 12*(8), 1187–1193.
16. Aaronson, N. K., Cull, A. M., Kaasa, S., & Sprangers, M. A. (1996). The European organization for research and treatment of cancer modular approach to quality of life assessment in oncology: An update. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (2nd ed., pp. 179–189). Philadelphia: Lippincott-Raven.
17. Shannon, C. E., & Weaver, W. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379–423.
18. Dahl, F. A., & Østerås, N. (2010). Quantifying information content in survey data by entropy. *Entropy, 12*, 161–163.
19. Kleinbaum, D. G., Kupper, L. L., Nizam, A., & Muller, K. E. (2008). Testing for the equality of two correlations. In C. Crockett (Ed.), *Applied regression analysis and other multivariable methods* (4th ed., pp. 103–104). Belmont, CA: Thomson Brooks/Cole.
20. Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York, NY: Wiley, John and Sons, Inc.
21. Brazier, J., Jones, N., & Kind, P. (1993). Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. *Quality of Life Research, 2*(3), 169–180.
22. 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.
23. Gao, F., Ng, G. Y., Cheung, Y. B., et al. (2009). The Singaporean English and Chinese versions of the EQ-5D achieved measurement equivalence in cancer patients. *Journal of Clinical Epidemiology, 62*(2), 206–213.
24. Hughes, D. A. (2007). Feasibility, validity and reliability of the Welsh version of the EQ-5D health status questionnaire. *Quality of Life Research, 16*(8), 1419–1423.
25. Johnson, J. A., & Pickard, A. S. (2000). Comparison of the EQ-5D and SF-12 health surveys in a general population survey in Alberta, Canada. *Medical Care, 38*(1), 115–121.