



The EORTC QLU-C10D is a valid cancer-specific preference-based measure for cost-utility and health technology assessment in the Netherlands

Micha J. Pilz^{1,2} · Simon Seyringer³ · Lára R. Hallsson² · Andrew Bottomley⁴ · Femke Jansen^{5,6,7} · Madeleine T. King⁸ · Richard Norman⁹ · Marianne J. Rutten¹⁰ · Irma M. Verdonck-de Leeuw^{5,6,7,11} · Peter D. Siersema^{12,13} · Eva Maria Gamper^{1,3}

Received: 2 May 2023 / Accepted: 10 January 2024
© The Author(s) 2024

Abstract

Background Cost-utility analysis typically relies on preference-based measures (PBMs). While generic PBMs are widely used, disease-specific PBMs can capture aspects relevant for certain patient populations. Here the EORTC QLU-C10D, a cancer-specific PBM based on the QLQ-C30, is validated using Dutch trial data with the EQ-5D-3L as a generic comparator measure.

Methods We retrospectively analysed data from four Dutch randomised controlled trials (RCTs) comprising the EORTC QLQ-C30 and the EQ-5D-3L. Respective Dutch value sets were applied. Correlations between the instruments were calculated for domains and index scores. Bland–Altman plots and intra-class correlations (ICC) displayed agreement between the measures. Independent and paired *t*-tests, effect sizes and relative validity indices were used to determine the instruments' performance in detecting clinically known-group differences and health changes over time.

Results We analysed data from 602 cancer patients from four different trials. In overall, the EORTC QLU-C10D showed good relative validity with the EQ-5D-3L as a comparator (correlations of index scores $r=0.53$ – 0.75 , ICCs 0.686 – 0.808 , conceptually similar domains showed higher correlations than dissimilar domains). Most importantly, it detected 63% of expected clinical group differences and 50% of changes over time in patients undergoing treatment. Both instruments showed poor performance in survivors. Detection rate and measurement efficiency were clearly higher for the QLU-C10D than for the EQ-5D-3L.

Conclusions The Dutch EORTC QLU-C10D showed good comparative validity in patients undergoing treatment. Our results underline the benefit that can be achieved by using a cancer-specific PBM for generating health utilities for cancer patients from a measurement perspective.

Keywords Cancer-specific preference-based measure · EORTC QLU-C10D · EQ-5D-3L · Validity · Responsiveness

JEL Classification I180

Introduction

Health systems face increasing pressure to scrutinise health-care expenditures and allocate scarce resources in the most effective manner. Therefore, an increased utilisation of

cost-utility analysis (CUA) is observed, since CUA allows policy makers to directly compare the clinical benefits and economic costs of different healthcare interventions [1]. In CUAs, quality-adjusted life years (QALYs) commonly serve as an outcome parameter combining quality of life, expressed as health utilities between 1 for perfect health and 0 for death, and lifespan into a single metric [2]. QALYs constitute the most regarded single indicator in health economics [3]. To estimate health utilities, direct preference assessments and preference-based measures (PBMs) are two commonly applied methods [4]. Direct preference

The research in this article was done during Andrew Bottomley's tenure at EORTC. Andrew Bottomley is no longer affiliated with EORTC.

Extended author information available on the last page of the article

assessments such as time-trade-off, standard gamble or visual analogue scales are known to be resource-intensive and time-consuming [5]. Estimating health utilities utilising PBMs provides a convenient alternative and has therefore gained increasing attention [5]. PBMs are based on a health state classification system as well as utility decrements, usually derived from the general population, to determine health state values [6, 7]. Several generic PBMs, such as the Health Utility Index (HUI) [8], the EuroQol 5-Dimensions (3 level and 5 level version-EQ-5D-3L and EQ-5D-5L) [9, 10] and the Short Form Six-Dimensions (SF-6D) [6] have been developed. Generic PBMs include general and universally applicable domains, such as physical functioning, role functioning, emotional functioning, and pain. They can, therefore, be applied in a broad range of medical conditions and make results comparable across disease groups and patient populations [11]. Currently, generic PBMs are primarily used for CUA assessments in many countries [12].

However, generic instruments have been criticised as they might fail to capture important health-related quality of life (HRQOL) symptom and functioning domains [13] such as nausea, fatigue, appetite loss in cancer patients [7]. Therefore, any impact of treatments on these domains cannot be accounted for when performing CUAs when using generic PBMs to evaluate interventions for cancer. Due to these considerations, disease specific instruments have evolved, allowing the estimation of CUA for certain patient populations [11, 14–16].

The Multi-Attribute Utility in Cancer (MAUCa) Consortium [7, 17] and the European Organisation of Research and Treatment of Cancer (EORTC) [18–21] recently developed a disease specific PBM for the cancer patient population—the EORTC Quality of Life Utility-Core 10 dimensions (EORTC QLU-C10D). For its development, the structure and content of the widely used HRQOL questionnaire EORTC QLQ-C30 was utilised to identify the most relevant HRQOL domains for cancer patients (results reported elsewhere) [7]. The EORTC QLU-C10D has been designed as a scoring algorithm for the EORTC QLQ-C30, allowing the calculation of utility values from QLQ-C30 scores [17–25].

There is a wide discussion in the literature regarding the most appropriate PBM in certain settings (e.g. palliative care, elderly people) and conditions (e.g. cancer, chronic disease) [14, 26–28]. The issue of generic versus disease specific PBMs is of particular interest, with psychometric characteristics of sensitivity (detecting health status differences) and responsiveness (detecting health status changes over time) highlighted as particularly important [26, 29]. To date, several empirical studies have investigated the measurement properties of different PBMs [28–33], with some scrutinising the sensitivity and responsiveness in various patient populations and disease categories [32, 34, 35]. In some cases, the disease-specific PBM had superior

measurement attributes compared to the generic PBM [29, 32, 33, 35]. This methodological discussion is valuable as health utilities and CUAs are used by health care authorities for reimbursement decisions [36] and, therefore, determine which treatments will be reimbursed. Empirical evidence in a wide range of settings is essential to test theoretical arguments about the relative advantages of generic versus disease-specific PBMs, and the conditions under which these arguments hold or do not hold.

The aim of our study was to assess the validity of the EORTC QLU-C10D for use in the Dutch cancer patient population complying with the CONSORT quality criteria for health status questionnaires [37]. We evaluated the comparative clinical validity of the cancer specific EORTC QLU-C10D using the well-established EQ-5D-3L as the comparator measure. This includes the evaluation of floor and ceiling effects, correlations of index and domain scores, Bland–Altman plots, sensitivity for known-group differences between clinically defined groups, responsiveness (i.e. the ability to detect change in the health status), and relative efficiency (i.e. the statistical efficiency of the QLU-C10D to detect differences in health statuses in comparison to the EQ-5D-3L). We draw on data from four RCTs that investigated different healthcare interventions in the Dutch cancer patient population targeting HRQOL as secondary outcomes [38–41]. In all of them, EORTC QLQ-C30 as well as EQ-5D-3L data were collected, allowing the retrospective validation of the Dutch version of the EORTC QLU-C10D.

Methods

The instruments EORTC QLU-C10D

The EORTC QLU-C10D [7] is a recently developed cancer-specific PBM. It has been derived from the most widely used HRQOL questionnaire in cancer research—the EORTC QLQ-C30 [42]. Thirteen of its 30 items have been selected to form ten domains constituting the QLU-C10D health state classification system [7]. These domains are physical functioning, role functioning, social functioning, emotional functioning, pain, fatigue, sleep, appetite, nausea, and bowel problems. Each domain consists of four severity levels (“not at all”, “a little”, “moderate”, “very much”). Preference-based scoring algorithms have been developed for a range of countries [17–21], including recently the Netherlands [19]. Table 1 shows QLU-C10D domains and health state descriptions as well as conceptually similar counterpart domains of the EQ-5D-3L.

Table 1 QLU-C10D and EQ-5D domain and health state description analogies

Counterpart domains					
QLU-C10D			EQ-5D-3L		
Domain	Level	Health state descriptions	Domain	Level	Health state descriptions
Physical functioning	1	No trouble taking a long walk outside of the house	Mobility	1	No problems in walking about
	2	No trouble taking a short walk outside of the house, but at least a little trouble taking a long walk		2	Some problems in walking about
	3	At least a little trouble taking a short walk outside of the house, and at least a little trouble taking a long walk		3	Confined to bed
	4	Quite a bit or very much trouble taking a short walk outside the house			
Role functioning	1	Not at all limited in pursuing work or other daily activities	Usual activities (<i>e.g. work, study, housework, family or leisure activities</i>)	1	No problems with performing usual activities
	2	A little limited in pursuing work or other daily activities		2	Some problems with performing usual activities
	3	Quite a bit limited in pursuing work or other daily activities		3	Unable to perform usual activities
	4	Very much limited in pursuing work or other daily activities			
Social functioning	1	Physical condition or medical treatment interferes not at all with social or family life	Usual activities (<i>e.g. work, study, housework, family or leisure activities</i>)	1	No problems with performing usual activities
	2	Physical condition or medical treatment interferes a little with social or family life		2	Some problems with performing usual activities
	3	Physical condition or medical treatment interferes quite a bit with social or family life		3	Unable to perform usual activities
	4	Physical condition or medical treatment interferes very much with social or family life			
Emotional functioning	1	not at all feeling depressed	Anxiety/depression	1	Not anxious or depressed
	2	feeling a little depressed		2	Moderately anxious or depressed
	3	feeling quite a bit depressed		3	Extremely anxious or depressed
	4	feeling very much depressed			
Pain	1	no pain	Pain/discomfort	1	No pain or discomfort
	2	a little pain		2	Moderate pain or discomfort
	3	quite a bit pain		3	Extreme pain or discomfort
	4	very much pain			
<i>Domains without counterpart</i>					
<i>QLUC10D</i>			<i>EQ-5D</i>		
Fatigue	1	not at all tired	Self-Care	1	No problems with self-care
	2	a little pain		2	Some problems washing or dressing
	3	quite a bit pain		3	Unable to wash or dress
	4	very much pain			
Sleep disturbance	1	no trouble sleeping			
	2	a little trouble sleeping			
	3	quite a bit trouble sleeping			
	4	very much trouble sleeping			

Table 1 (continued)

Counterpart domains			Counterpart domains		
QLU-C10D			EQ-5D-3L		
Domain	Level	Health state descriptions	Domain	Level	Health state descriptions
Appetite loss	1	not at all lacking appetite			
	2	a little lacking appetite			
	3	quite a bit lacking appetite			
	4	very much lacking appetite			
Nausea	1	not at all feeling nauseated			
	2	a little feeling nauseated			
	3	quite a bit feeling nauseated			
	4	very much feeling nauseated			
Bowel problems	1	no constipation or diarrhoea			
	2	a little constipation or diarrhoea			
	3	quite a bit constipation or diarrhoea			
	4	very much constipation or diarrhoea			

EQ-5D-3L

The EQ-5D-3L was developed in 1990 as a simple, standardised, generic HRQOL questionnaire that can also be applied as a generic PBM [43]. Its five domains are mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The EQ-5D-3L has three response levels for each domain (“no problems”, “some problems”, “severe problems/unable to”). Value sets are available for a range of countries [44–47]; the Dutch value sets were published in 2006 [44]. Table 1 shows EQ-5D-3L domains and health state descriptions as well as theoretical counterpart domains of the QLU-C10D.

Data description

In this retrospective analysis, data from four multicentre RCTs were used, which assessed HRQOL in a broad range of cancer patients and interventions in the Netherlands. To be eligible, patients had to have completed the EORTC QLQ-C30 and the EQ-5D-3L at the same measurement time points. The first RCT is the SIREC study ($n=209$; 199 eligible for our analysis), which was conducted in nine Dutch hospitals and investigated the effect of stent insertion vs. brachytherapy on dysphagia relief in patients with oesophagus cancer [39]. The second RCT examined the effect of a Stepped Care (SC) ($n=156$; 147 patients eligible for our analysis) approach of psychosocial interventions versus care-as-usual on psychological distress and HRQOL in head and neck cancer and lung cancer patients who were not in active anti-tumor treatment [40]. The third RCT investigated the effect of meaning-centred group psychotherapy

(MCGP-CS) vs. supportive group psychotherapy (SGP) vs. care-as-usual on personal meaning and HRQOL in ($n=170$; 168 cancer survivors eligible for our analysis) Dutch cancer survivors [41]. The fourth RCT was conducted as a multicentre study in eight Dutch gynaecological centres, investigating the effects of diagnostic laparoscopy on treatment decision making, such as applying primary cytoreductive surgery (LapOvCa), in 201 patients with advanced ovarian cancer (78 patients eligible for our analysis) [38]. All trials assessed the EORTC QLQ-C30 and the EQ-5D-3L as secondary endpoints. Patient data were included in the current analysis if EORTC QLU-C10D and EQ-5D-3L longitudinal data was available. Details regarding the study design and sampling procedures of the RCTs have been reported previously [38–41].

Statistical analysis

Data from all four RCTs were analysed separately as pooling of data was not appropriate due to the differing study designs. Sample characteristics were presented as absolute and relative frequencies, means, and standard deviations. Utility scoring was conducted using the Dutch value sets for the EORTC QLU-C10D and the EQ-5D-3L [19, 44].

To investigate the clinical validity, ceiling and floor effects for the PBMs were calculated as the percentage reaching maximum/minimum health utility scores for each instrument. According to Terwee et al. [37], substantial floor or ceiling effects are present if more than 15% report the lowest or highest score possible, respectively. Furthermore, Spearman correlation coefficients were calculated to evaluate the construct and content validity by correlating the

utility value index scores and domain pairs considering coefficients of 0.1–0.3 as weak, 0.4–0.6 as moderate, and 0.7–0.9 as strong [48]. On the domain level, some cancer-specific aspects from the QLU-C10D conceptually differ from the generic EQ-5D domains (displayed in Table 1) and were therefore expected to show low correlations (divergent validity). For domains with conceptually similar content (EQ-5D mobility and QLU-C10D physical functioning, EQ-5D usual activities and QLU-C10D role- and social functioning, EQ-5D pain/discomfort and QLU-C10D pain, EQ-5D anxiety/depression and QLU-C10D emotional functioning), moderate to strong correlations were expected (convergent validity). Strong correlations were expected between the index scores of the two PBMs. As measures of agreement between the QLU-C10D and the comparator measure, intraclass correlations (ICCs) of index scores were calculated. To further scrutinize the agreement between the scores across the measurement ranges, Bland–Altman-Plots were created. These were used to display systematic (dis)agreements between the measures.

Sensitivity and responsiveness were investigated by scrutinising the PBMs ability to discriminate between clinical known-groups and between time points for which health states were expected to differ, respectively. Sensitivity was investigated in cross-sectional groups at baseline (for the “dysphagia improved” and “retreatment” groups in the SIREC trial and all groups in the LapOvCa trial, follow-up data was used). The definition of clinical known-groups was performed according to what was known from the literature, complemented by the expert opinion of co-author clinicians. That is, we considered what could be expected with regard to health or HRQOL differences based on the results from the respective included trials [38–41], groups based on WHO performance status [49], type of cancer treatment [50–52], working status [53], histology [54], or tumour stage [55]. Selection of variables for known-group comparisons and the expected direction of utility differences are reported in Table 2.

Differences of index scores between groups (sensitivity) and over time (responsiveness) were statistically tested using *T*-tests (independent *t*-tests for sensitivity and repeated measures *t*-tests for responsiveness). Minimal important differences for the EQ-5D-3L [56] were used to contextualise the magnitude of differences and change. Relative validity (comparative validity) indices were calculated using a three-fold approach; by calculating effect sizes, the Relative Efficiency, and the Responsiveness Index. Cohen’s *D* and standardised response mean (SRM) were used as effect size measures; Cohen’s *D* was used to estimate effect sizes between groups (sensitivity) while the SRM was used as effect size measure for change over time (responsiveness). Cohen’s *D* were classified as small (0.2–0.49), moderate (0.5–0.79) and large (≥ 0.8) [57]. We used the same thresholds for SRM. The Responsiveness Index

(RI) was calculated for each instrument by dividing the change over time by the SD of the control condition following the methodology applied by King et al. [58]. The RI was originally developed to test an instruments’ ability to assess change between a treatment and a (stable) intervention arm [59]. When testing the RI also for further clinical variables, we additionally calculated the RI for other clinical subgroups which deviated over time (e.g., improved/persistent dysphagia, (non-) residual disease, recipient of formal care yes/no, etc.). In addition, when both instruments detected statistically significant effects, the instruments’ relative efficiency (RE) was assessed as the ratio of squared *t* values $\frac{(t\text{-statisticEORTCQLU-C10D})^2}{(t\text{-statisticEQ-5D-3L})^2}$, whereby a RE > 1 indicates a higher efficiency for the QLU-C10D and RE < 1 a higher efficiency for the EQ-5D, respectively [58, 60]. To further investigate the relative validity of responsiveness for the instruments, the difference of RI was calculated by subtracting resulting RI indexes. A positive RI difference indicates better responsiveness of the EORTC QLU-C10D, while a negative RI difference indicates superior responsiveness for the EQ-5D-3L.

To adjust for multiple testing and to keep the alpha error $\leq 5\%$, Bonferroni correction was applied (p-values corrected according to the number of tests performed per data set). All statistical analyses were performed using IBM SPSS 21 [61].

Results

Sample characteristics

The included patient data across the four RCTs represented a diverse sample with respect to sociodemographic and clinical characteristics. The mean age across the analysed studies ranged from 57.1 years for the MCGP-CS trial to 68.7 years for the SIREC trial. The proportion of male participants per study ranged from 0% for the LapOvCa trial up to 77.5% for the SIREC trial. Patients had mostly been treated with chemotherapy, surgery or radiotherapy. Cancer sites differed across the trials (SIREC: oesophagus, LapOvCa: ovaries, SC: head and neck, and lung, MCGP-CS: heterogeneous sample, including patients with breast and colon cancer). Patients were in active anti-tumour treatment in two trials (SIREC, LapOvCa) and were survivors or at post treatment (majority > 7 months) in the other two (MCGP-CS and SC). Further details are provided in Table 2.

Ceiling and floor effects

Ceiling effects were below the 15% threshold for the EORTC QLU-C10D (0.7% in SC to 7.7% in LapOvCa) and exceeded the threshold of 15% for the EQ-5D-3L in two trials (20.8% in MCGP-CS and 15.4% in LapOvCa). Floor effects for both

Table 2 Sociodemographic data and treatment information across four RCTs, including variables used for known group comparisons*

	SIREC (N=209)	SC (N=147)	MCGP-CS (N=168)	LapOvCa (N=78)**
Mean age (SD)	68.7 (11.6)	62.8 (9.3)	57.1 (10.2)	62.4 (9.3)
Sex				
Male	162 (77.5%)	83 (58.9%)	29 (17.3%)	0 (0.0%)
Female	47 (22.5%)	58 (41.1%)	139 (82.7%)	78 (100.0%)
Living arrangements				
Married/living together	-	96 (68.1%)	133 (79.1%)	-
Unmarried/divorced/ widowed	-	45 (31.9%)	33 (19.6%)	-
Other	-	-	4 (2.3%)	-
Paid work				
No	-	98 (69.5%)	77 (45.8%)	-
Yes	-	43 (30.5%)	88 (52.4%)	-
Missing	-	-	3 (1.8%)	-
Variables for known group comparisons*	Treatment stent < brachytherapy	Treatment Stepped Care > Care as usual	Treatment Meaning-Centred Group Psychotherapy ~ Supportive Group Psychotherapy ~ Care as usual	Futile Laparotomy no > yes
	WHO-PS 0–2 > 3–4	Paid work (working age < 66yrs) yes > no	Primary tumour breast ~ colon	Formal care no > yes
	Histology Squamous cell vs. Adeno Cell	Time since treatment less than 7 months < more than 7 months	Treatment chemoradiation ~ radiotherapy	Informal care no > yes
	Previous CTX no > yes	Treatment Unimodal > Multimodal	Hormonal therapy received yes < no	Fig0 stage < IIIc > IIIc/IV
	Improved dysphagia at day 30 yes > no	Tumour stage local > advanced		WHO PS 1 > 2–4
	Persistent dysphagia at day 30 yes < no			Histology malign < benign
	Retreatment after 30 days yes > no			Residual disease; none > +1 cm
Tumour location				
Oesophag	179 (85.6%)	–	–	–
Oesophag. Junct	30 (14.4%)	–	–	–
Lip/oral cavity/oropharynx	–	69 (48.9%)	–	–
Hypopharynx/larynx	–	36 (26.2%)	–	–
Other head and neck cancers	–	27 (18.4%)	–	–
Lung	–	9 (6.4%)	–	–
Breast	–	–	112 (66.7%)	–
Colon	–	–	36 (21.4%)	–
Ovaria	–	–	–	78 (100%)
Other	–	–	20 (11.9%)	–
Tumour treatment				
Surgery	–	33 (23.4%)	167 (99.4%)	60 (76.9%)
Radiotherapy	–	34 (24.1%)	95 (56.5%)	–
Chemotherapy	179 (85.6%)	24 (17.0%)	95 (56.5%)	8 (10.3%)
Hormonal therapy	–	–	80 (47.6%)	–
Surgery and radiotherapy	–	38 (27.0%)	–	–
Surgery and chemoradiation	–	8 (5.7%)	–	–
Surgery and chemotherapy	–	4 (2.8%)	–	–

Table 2 (continued)

	SIREC (N=209)	SC (N=147)	MCGP-CS (N=168)	LapOvCa (N=78)**
Brachytherapy	101 (48.3%)	–	–	–
Stent	108 (51.7%)	–	–	–
WHO performance status at baseline (t0)				
0—Fully active	79 (37.8%)	–	–	–
1—Restricted in physically strenuous activity	66 (31.6%)	–	–	45 (57.69%)
2—unable to carry out any work activities	39 (18.7%)	–	–	32 (41.03%)
3—confined to bed or chair more than 50% of waking hours	19 (9.1%)	–	–	1 (1.28%)
4—completely disabled	3 (1.4%)	–	–	–
Missing	3 (1.4%)	–	–	–
Tumor stage				
UICC I–II	–	60 (42.6%)	–	–
UICC III–IV	–	72 (51.1%)	–	–
Missing	–	9 (6.4%)	–	–
FIGO stage				
FIGO I–IIIb	–	–	–	25 (32.1%)
FIGO IIIc–IV	–	–	–	53 (68.0%)
Mean Months since last treatment				
<7 months	–	48 (34.0%)	–	–
7–12 months	–	24 (17.0%)	–	–
>12 months	–	69 (48.9%)	–	–

*Variables for known-group comparisons were selected based on availability in the trial data sets; the direction of expected differences is indicated by < (utilities of group 1 is expected to be smaller than utilities of group 2), > (utilities of group 1 is expected to be greater than utilities of group 2), or ~ ((utilities of group 1 and 2 are expected to be similar)

**As the LapOvCa trial had a high missing HRQOL data rate, potential systematic bias was assessed with independent t-tests of clinical variables between included (n=78) versus excluded patients (n=123); there were no statistically significant differences (p-values ranged from 0.082 to 0.909)

Abbreviations: *SIREC* stent or Intraluminal Radiotherapy for inoperable Esophageal Cancer, *SC* stepped care, *MCGP-CS* Meaning Centred Group Psychotherapy for Cancer Survivors, *LapOvCa* laparoscopy to predict the result of primary cytoreductive surgery in advanced ovarian cancer patients, *UICC* Union for International Cancer Control, *FIGO* federation Internationale de Gynecologie et d'Obstetrique, *SD* standard deviation

instruments were marginal throughout (≤1.3%). For details, see Table 3.

Correlation between measures – convergent and divergent validity

Overall, Spearman’s correlation coefficient between the EQ-5D-3L and the QLU-C10D index scores were moderate to strong ranging from 0.534 to 0.749. Conceptually, similar domains showed mostly moderate correlations. For the pairs Physical Functioning—Mobility, Role Functioning—Usual Activities, and Social Functioning—Usual Activities the statistically significant correlation coefficients ranged from 0.432 to 0.711. For the domain pair Pain—Pain/Discomfort, the statistically significant

correlations ranged from 0.335 to 0.697. The Emotional Functioning—Anxiety/Depression domain pair showed statistically significant moderate correlation only in the

Table 3 Ceiling and Floor Effect for the EQ-5D-3L and the QLU-C10D

	EQ-5D-3L		QLU-C10D	
	Ceiling effect	Floor effect	Ceiling effect	Floor effect
SIREC	12.3%	0.0%	4.3%	0.0%
Stepped Care	8.5%	0.0%	0.7%	0.0%
MCGP—CS	20.8%	0.0%	4.8%	0.0%
LapOvCa	15.4%	1.3%	7.7%	1.3%

SC trial (0.352, $p \leq 0.001$); in the other three trials, correlations were weak (0.119–0.178). Conceptually dissimilar domains mostly showed weak correlations. Only for the domain pair Physical Functioning—Usual Activities, a statistically significant correlation coefficient exceeded 0.5 points in the SIREC and LapOvCa trial, which is classified as moderate correlation. Further details are provided in Table 4.

Agreement between measures

The ICCs between the index scores of the QLU-C10D and the EQ-5D ranged from 0.686 up to 0.808. According to Cicchetti [62], these ICCs can be classified as good to excellent (see Table 5). Bland–Altman plots (Fig. 1) indicated that the QLU-C10D produced consistently higher utility values compared to the EQ-5D-3L. This systematic measurement difference ranged from 0.005 for the LapOvCa trial to 0.055 for the SC trial. In addition, the Bland–Altman-plots showed some systematic bias with score differences becoming smaller towards the upper end of the measurement scale. The distribution and span of the scores differences are presented as the level of agreements (LOA), whereby it is defined that 95% of the observed score difference lies within the LOA. The LOA ranged from 0.648 for the MCGP-CS trial up to 0.859 for the SIREC trial (see Fig. 1).

Sensitivity and responsiveness analysis

Sensitivity

The results of the known-group comparisons can be seen in detail in Table S1. The EORTC QLU-C10D and the EQ-5D-3L both detected statistically significant differences of index scores in the same three groups (WHO performance status 0–2 vs. 3–4, persistent dysphagia at day 30 yes vs. no, retreatment needed yes vs. no) with p -values ≤ 0.05 and REs twice in favour of the QLU-C10D (see Fig. 2). The QLU-C10D detected three additional differences not detected by the EQ-5D-3L (histology malignant vs. benign in ovarian cancer, formal care needed yes vs. no, FIGO stage < IIIc vs IIIc/IV) and the EQ-5D-3L found two additional difference that the QLU-C10D did not detect (histology in oesophagus cancer, informal care needed yes vs. no). All significant utility score differences were in the expected directions (e.g., patients with higher WHO-PS had lower utility values; patients with higher FIGO staging had lower utility values; patients with benign histology had higher utility values). The results indicate a good sensitivity for the EORTC QLU-C10D to detect differences in health states. When compared to the EQ-5D-3L the EORTC QLU-C10D detected these

cross-sectional health state differences with a higher efficiency in 5 from 8 (62.5%) comparisons (Fig. 2).

Responsiveness

The EORTC QLU-C10D detected four statistically significant changes of index scores for overall changes over time (SC: t0 to t1, $p < 0.013$), three of which were not detected by the EQ-5D-3L (SIREC: t0 to t1; $p < 0.001$, SC: t0 to t5, LapOvCa: t2 to t3) (Table S2). Where both instruments measured differences with statistical significance (SC: t0 to t1), the QLU-C10D had a higher relative efficiency in detecting this difference. (Table S2 and Fig. 2). All test for overall change showed a score change in the expected direction, e.g. lower follow-up scores for the SIREC where health states were expected to deteriorate over time, and higher follow-up utility values for the LapOvCa trial (after treatment completion) and the SC trial where health states were expected to improve over time, indicating good responsiveness for the QLU-C10D.

Analyses of changes over time within (treatment) groups (see Table S3) showed that the EORTC QLU-C10D detected six groups with statistically significant change, while the EQ-5D-3L was able to identify one groups with statistically significant changes over time. Most of these statistically significant change scores were congruent with the hypothesised direction, e.g., utility scores deteriorated for patients with stent treatment, or with non-improved and persistent dysphagia in the SIREC trial, and improvement for malignant patients and patients relying on formal care after treatment completion in the LapOvCa trial. Additionally, the SRM was calculated for (treatment) group comparisons. There was no group comparison with a high SRM, four SRMs of the EORTC QLU-C10D and one SRM of the EQ-5D-3L were classified as moderate, respectively. In the comparison where both instruments detected statistically significant changes, the EORTC QLU-C10D showed a higher RE compared to the EQ-5D-3L (see Fig. 2).

For the responsiveness analysis for change over time between (treatment) groups, we performed 14 between-group analyses. For these 14 comparisons, the EORTC QLU-C10D was able to identify two statistically significant differences between groups—change over time differences, while the EQ-5D-3L identified one statistically significant difference for change over time between (treatment) groups. Again, the observed statistically significant differences for change over time between groups are in line with the expected direction, showing higher utility values for patients with improved/non-persistent dysphagia in the SIREC-trial. Where applicable, RE was in favour of the EORTC QLU-C10D (see Fig. 2). As an additional parameter of responsiveness, the RI showed that the EORTC QLU-C10D showed a higher responsiveness (sensitivity to change) in eight of

Table 4 Spearman Correlations of index values and utilities per domain

QLU-C10D	SIREC				SC				MCGP-CS				LapOvCa							
	EQ5D-3L	EQ5D-3L	EQ5D-3L	EQ5D-3L	EQ5D-3L	EQ5D-3L	EQ5D-3L	EQ5D-3L	EQ5D-3L	EQ5D-3L	EQ5D-3L	EQ5D-3L	EQ5D-3L	EQ5D-3L	EQ5D-3L					
Index	Index	Index	Index	Index	Index	Index	Index	Index	Index	Index	Index	Index	Index	Index	Index					
Index	0.640**	0.657**	0.657**	0.657**	0.534**	0.534**	0.534**	0.534**	0.534**	0.534**	0.534**	0.534**	0.534**	0.534**	0.749**					
Domains	MO	SC	PD	AD	UA	MO	SC	PD	AD	UA	MO	SC	PD	AD	UA					
PF	0.623*	0.441*	0.068	0.139	0.558*	0.590*	0.348*	0.291*	0.192	0.341*	0.574*	0.137	0.370*	0.248*	0.430*	0.504*	0.352*	0.320*	0.007	0.526*
RF	0.456*	0.473*	0.292*	0.272*	0.688*	0.417*	0.266*	0.332*	0.288*	0.483*	0.286*	0.227*	0.324*	0.170	0.532*	0.205	0.270	0.224	0.033	0.711*
SF	0.325*	0.403*	0.318*	0.228*	0.495*	0.321*	0.234*	0.372*	0.269*	0.524*	0.304*	0.114	0.079	0.267*	0.432*	0.144	0.235	0.384*	0.225	0.568*
EF	0.127	0.298*	0.120	0.178	0.248*	0.085	0.048	0.194	0.352*	0.222*	-0.063	-0.021	-0.016	0.162	0.015	0.162	-0.03	0.085	0.119	0.047
PA	0.079	0.095	0.501*	0.274*	0.235*	0.220*	0.170	0.414*	0.182	0.129	0.445*	0.077	0.335*	0.174	0.246*	0.372*	-0.038	0.459*	0.126	0.305*
FA	0.308*	0.316*	0.292*	0.235*	0.483*	0.214	0.061	0.365*	0.373*	0.393*	0.146	0.112	0.245*	0.175	0.299*	0.318*	0.350*	0.290	0.165	0.531*
SL	0.051	0.120	0.072	0.223*	0.217*	0.196	0.019	0.174	0.099	0.051	0.104	-0.004	0.139	0.121	0.075	0.082	-0.037	0.230	0.197	0.170
AP	0.068	0.141	0.263*	0.235*	0.164	0.154	0.154	0.315*	0.277*	0.305*	0.076	0.081	0.063	0.114	0.095	0.225	0.189	0.393*	0.344*	0.347*
NA	0.108	0.117	0.167	0.114	0.260*	0.049	0.083	0.304*	0.152	0.232*	0.217*	0.086	0.287*	0.055	0.224*	0.123	0.116	0.198	0.053	0.125
BO	0.228*	0.206*	0.162	0.117	0.193*	0.247*	-0.056	0.210	0.172	0.173	0.226*	-0.009	0.154	0.074	0.227*	0.221	0.072	0.21	0.240	0.040

Correlation coefficients of theoretically corresponding domains are formatted **bold**

EQ-5D: MO mobility, SC self care, PD pain and discomfort, AD anxiety and depression, UA usual activities, QLU-C10D: PF physical functioning, RF role functioning, SF social functioning, EF emotional functioning, PA pain, FA fatigue, SL sleep, AP appetite, NA nausea, BO bowel problems

* significant after Bonferroni correction at p < 0.05

Table 5 Intra-class correlation from the EQ-5D-3L and the EORTC QLU-C10D across four RCTs

	ICC ^a	CI ^b	<i>p</i> value
SIREC	0.808	0.745–0.856	<0.001
SC	0.740	0.633–0.815	<0.001
MCGP-CS	0.686	0.573–0.769	<0.001
LapOvCa	0.751	0.609–0.842	<0.001

^aIntra-class correlation^bCI: 95% Confidence Interval

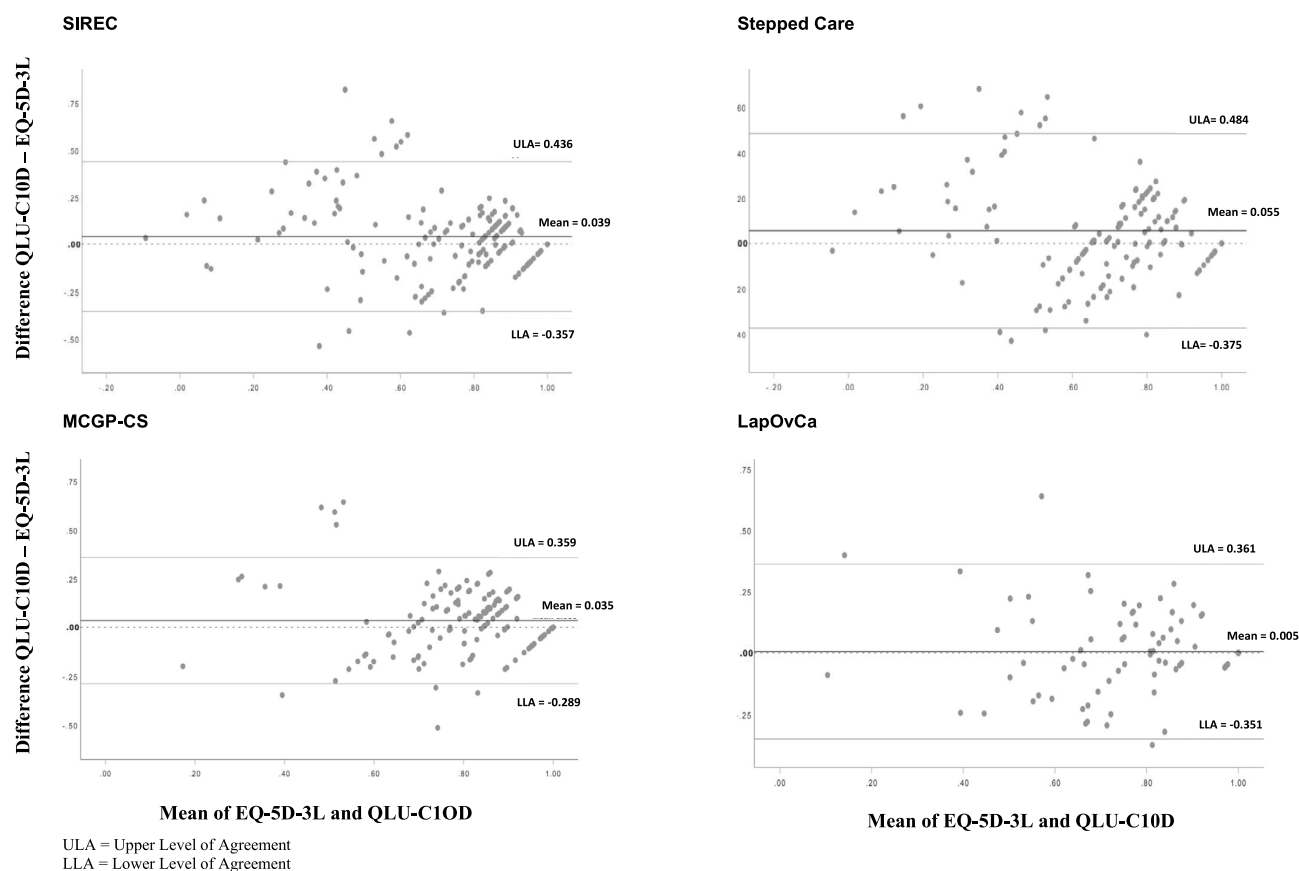
these comparisons, indicating a higher measurement precision for health state changes over time for the EORTC QLU-C10D. For further details, please see table S3 and Fig. 2.

Discussion

The EORTC QLU-C10D is a cancer-specific utility instrument developed as a scoring algorithm for the widely used cancer-specific quality of life questionnaire EORTC QLQ-C30. It aims to support health economic evaluations in

cancer patient populations [7]. The instrument is now in the final clinical validation process, to which this and other studies [31–33, 63] contribute. The EQ-5D-3L is a well-established and widely used PBM, making it a useful standard comparator against which to assess the clinical validity of the EORTC QLU-C10D. As our analyses have shown, the QLU-C10D measures five generic constructs similar to those included in the EQ-5D, and an additional five domains related to symptoms commonly experienced by cancer patients. Additional to the content differences, the preference weights (i.e. utility decrements), also contribute to score differences of the two PBMs. Below we discuss the various aspects of validity we assessed in this study and interpret these in the context of these two key differences [31, 32, 63] between the QLU-C10D and the EQ-5D.

The QLU-C10D's criterion validity relative to the EQ-5D-3L is supported by adequate correlations between the QLU-C10D's and the EQ-5D's index scores and conceptually similar domain pairs, although correlations between counterpart domains were overall somewhat weaker than expected. There was a striking difference between expected and actual correlations for the Emotional

**Fig. 1** Bland–Altman Plots displaying the agreement between the EORTC QLU-C10D and the EQ-5D-3L index scores

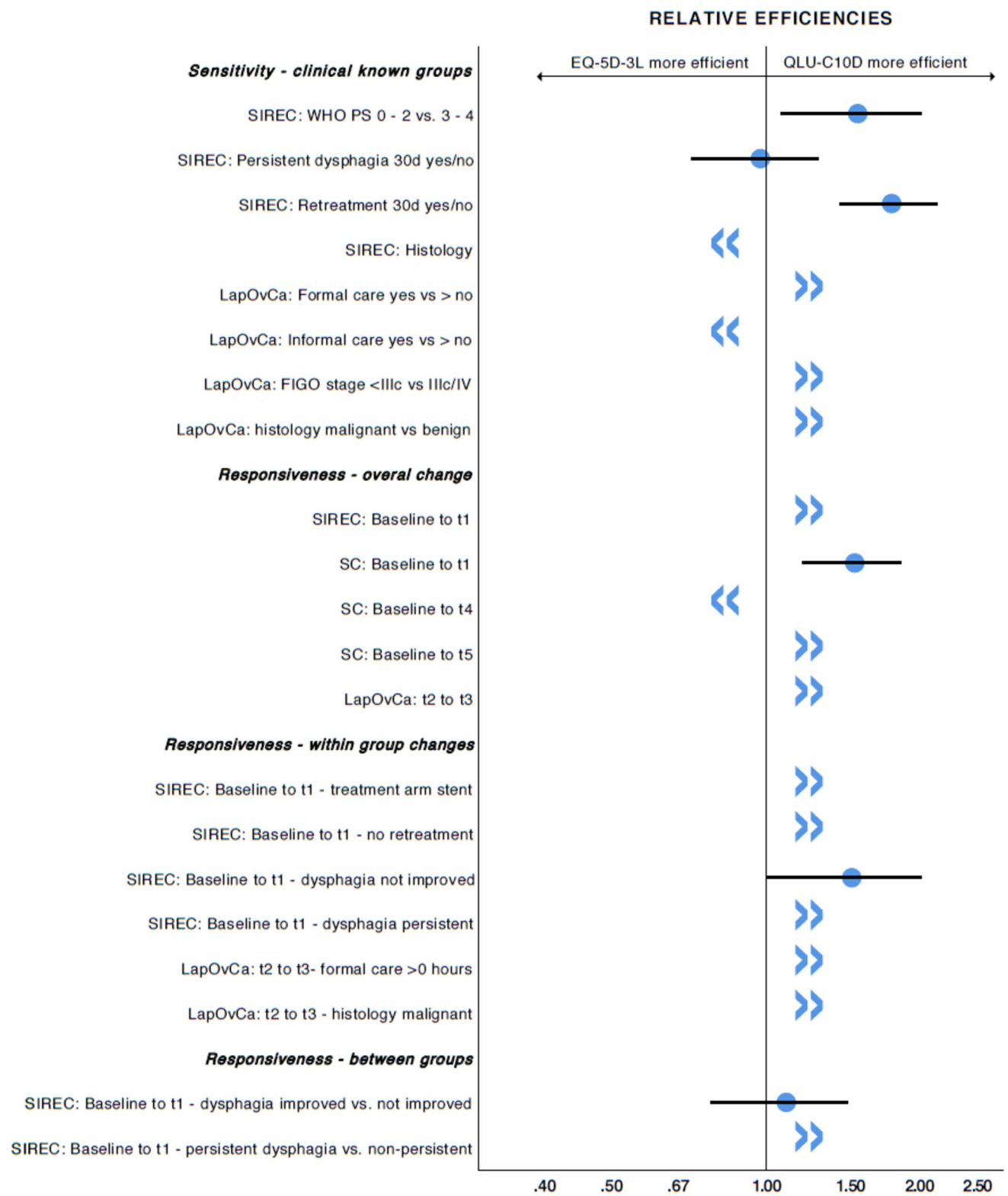


Fig. 2 Relative Efficiencies and detection of clinical group differences/change over time. >> The QLU-C10D detected a statistically significant difference where the EQ-5D-3L did not. << The EQ-5D-3L detected a statistically significant difference where the QLU-

C10D did not. *Blue filled circle* represent relative efficiency with confidence interval (only calculated if both instruments found statistically significant differences/changes)

Functioning domain of the QLU-C10D (which assesses feeling depressed) and the Anxiety/Depression domain of the EQ-5D-3L (which assesses feeling anxious or depressed). However, data from an international observational study in patients with myelodysplastic syndrome indicated that these two scales also behaved differently using the Italian tariffs [32]. As hypothesised in that paper, this may be because the EQ-5D domain includes anxiety which very well can be present without feelings of depression, while the QLU-C10D domain captures only depression, not anxiety. Aside from differences in the way the instruments ask about emotional burden, the impact of emotional functioning in the Dutch EORTC QLU-C10D value set is smaller than for other countries [17–25], and is clearly lower than the Dutch EQ-5D-3L anxiety/depression utility decrement [44]. Hence, the rather low correlation for the emotional domain may also be influenced by our use of the Dutch QLU-C10D value set where only the highest response category (highest emotional burden) has a utility decrement. It is unclear whether this expresses an informative preference or is the result of a translation issue of the word ‘depression’ in the QLQ-C30; if the latter, then the impact of emotional functioning on Dutch QLU-C10D utility may be underestimated using the current Dutch versions of the QLQ-C30 and QLU-C10D value set.

The moderate correlation of the theoretically distant domains pair Mobility and Usual Activities in the SIREC trial and LapOvCa trial might be explained by the palliative setting in the SIREC trial [64] and the advanced disease stages in the LapOvCa trial [38].

The ceiling effects for the EORTC QLU-C10D (0.7–7.7%) were clearly lower than for the EQ-5D-3L (8.5–20.8%) across the four trials. BA plots indicated that the EORTC QLU-C10D produced systemically higher utility values than the EQ-5D-3L in all four studies. The level of agreements showed a maximal range from -0.375 to 0.484 in the SC trial, exceeding the range of any observed SD of utility values in that trial. Still, the mean difference of the scores does not exceed the minimal important difference of the EQ-5D-3L in Norwegian glioma patients [56], which we used as a crude measure to evaluate the magnitude of utility score differences. However, the BA plots also showed at least some systematic bias in all four studies, indicating that the differences between QLU-C10D scores and EQ-5D3L scores vary across the utility measurement scale. The difference in measurement precision towards the upper end of the scales (existing ceiling effects for the EQ-5D-3L) potentially contributes to the systematic difference of utility scores derived from the two instruments. Therefore, utility scores between the instruments are not interchangeable.

Most importantly for clinical validity, the QLU-C10D’s sensitivity and responsiveness were assessed. We found statistically significant differences/changes in the hypothesised direction, indicating a good construct validity and

sensitivity to expected differences/changes of utility scores. There was an agreement between the QLU-C10D and EQ-5D-3L regarding the detection of expected differences. In overall, the QLU-C10D had an advantage with measurement efficiency which showed by either being the only instrument picking up a hypothesised effect or being more efficient in direct comparison with the EQ-5D-3L. This was especially true for responsiveness analyses. Higher efficiency would translate to smaller required sample sizes of future clinical studies assessing utilities. Similar results were obtained when analysing the responsiveness within and between clinical subgroups across different assessment time points, whereby the EORTC QLU-C10D showed continuously higher responsiveness within and between group comparisons over time. Shaw et al. [33] reported similar findings of improved sensitivity and responsiveness for the UK version of the EORTC QLU-C10D in patients receiving Nivolumab. Additionally, Bulamu et al. reported a superior responsiveness of the EORTC QLU-C10D over the EQ-5D-3L in patients undergoing esophagectomy.

The ability to discriminate was limited in some hypothesised known-group and responsiveness analyses. In a range of analyses neither of the PBM instruments detected a statistically significant difference/change. The included RCTs had HRQOL as secondary outcomes and therefore might be underpowered for HRQOL analyses. It can be surmised that some known-group differences went undetected. A number of effects were found despite underpowered subgroup analyses and, therefore, allow drawing solid conclusions on the applicability of the Dutch QLU-C10D as these would not disappear with higher power.

Variations in psychometric properties across diverse populations, disease- and treatment groups are no surprise [65] and it is important to know the instruments characteristics for a specific target population. In our analyses, we attribute the differences in psychometric performances to the distinct clinical contexts (e.g. inpatients vs. outpatients), the trial population (actively treated cancer patients vs. cancer survivors; different cancer sites), and the interventions (medical vs. psychosocial). For example, in the MCGP-CS trial, both PBMs showed poor performance. As in the original MCGP-CS publications [41, 66], there was no statistically significant difference/change reported for EORTC QLQ-C30 and EQ-5D-3L data; our findings align with this. This suggests that in cancer survivors both instruments need to be used with caution until their fitness for this purpose is explicitly evaluated. An additional possible explanation for different findings across trials might be the nature of intervention across the trials (psycho-social interventions for the MCGP-CS and SC trial vs. somatic interventions for the SIREC and LapOvCa trial) and the conceptual scale design of the PBMs. Although the psychological aspects of health (Anxiety/Depression) is one of the five domains

from the EQ-5D-3L, a study has shown that the EQ-5D-3L does not sufficiently portray the influence of mental and social health [67]. Similarly, the Emotional Functioning domain, asking for depressive mood and constituting one of the ten EORTC QLU-C10D domains, was shown to have a low impact on health utility values in Dutch cancer survivors [68]. Therefore, it can be argued that the instruments' abilities to capture the effects of emotional aspects on health utility values might be limited to some extent which could be reflected in lower sensitivity in samples where these are predominant problems. The topic of generic and disease-specific PBMs has been discussed for several years among health economists [11, 14–16, 69]. While a multitude of disease-specific instruments were developed, the National Institute for Health and Care Excellence suggests using the EQ-5D-3L as the preferred measure of health-related quality of life [70]. In the Netherlands, the use of the EQ-5D-5L is recommended, although additional PBMs may be justified to supplement evidence from the EQ-5D [71]. However, this generic approach in health economic evaluations may be reconsidered when targeting specific patient populations [27]. The EQ-5D-3L has been shown to produce valid results in a cancer patient population, yet specific conditions were identified in which it appears to be limited, and in which the EORTC QLQ-C30 seemed more adequate [15]. In addition, it is an extra questionnaire which should be filled out by the patient. The good construct and content validity of the EORTC QLU-C10D [32] in combination with its backwards compatibility with the EORTC QLQ-C30 are both favourable prerequisites for the EORTC QLU-C10D to become adopted in health economic evaluation schemes. This study contributes to this discussion on using generic or disease-specific instruments and suggests this topic needs to be explored further. Selecting the appropriate instrument is important as health utilities and the economic evaluations are increasingly used by health care authorities for decision which treatments to reimburse [36].

Limitations

A limitation of this study is the relatively small sample size for some of the known-group comparisons. This limited statistical power such that some real differences may have gone undetected. Another limitation of this analysis is its retrospective nature; therefore, we were not able to a priori define known-group comparisons and had to rely on variables available in the data sets from the primary studies. Nonetheless, the heterogeneous nature of the interventions and clinical contexts allowed us to perform known-group analysis across patient groups with various characteristics. This validation study of the EORTC QLU-C10D relied on comparison with data obtained using the EQ-5D-3L. Using

the newer version of the EuroQol measurement system, the EQ-5D-5L [10], as comparator in the Dutch cancer patient population was not possible due to the retrospective design of this study. The two additional levels of the newer version of the EQ-5D have improved its precision relative to the EQ-5D-3L [72]. However, it is not only the number of response category that affects the granularity of the measurements but also the content, and the additional disease-specific domains of the QLU-C10D are relevant here. Bulamu et al. [73] and Pan et al. [63] have compared aspects of validity and the RE of the EORTC QLU-C10D versus the EQ-5D-5L in gastric, and oesophagus cancer patients, whereby a superior relative validity was reported for the EQ-5D-5L in comparison to the findings of the 3-level version in this manuscript [63]. Further comparisons of the QLU-C10D and versus the EQ-5D-5L in other patient groups and clinical contexts are warranted to assess the generalisability of these findings. A final limitation was that we were not able to calculate and compare QALYs from the QLU-C10D and the EQ-5D as we did not have access to existing cost-effectiveness models for the four RCTs in this study. This had been done four times before [33, 74–76], exclusively showing lower incremental QALYs for the EORTC QLU-C10D compared to the EuroQOL measurement system. Where possible, future comparisons of the QLU-C10D with a generic PBM should include comparison of QALYs yielded, as this is of great relevance when choosing PBMs to conduct CUA.

Conclusion

Using data of four Dutch RCTs in this retrospective analysis showed good psychometric properties and clinical validity of the EORTC QLU-C10D compared to the EQ-5D-3L in the Dutch cancer patient population. Our findings show promising results for the further use of the EORTC QLU-C10D when facilitating CUA for cancer patients in the Netherlands. Importantly, in the Netherlands, health technology assessments have been applied since the early 1980s, mostly for drug reimbursement decisions [71]. Additionally, it is mentioned that in the area of specialist medical care, the nature of complexity and availability of quality data poses a limitation to health economic evaluations in the Dutch context [77, 78]. It, therefore, appears that the existence of huge data registries such as PROFILES [68] containing EORTC QLQ-C30 data and the backward compatibility of the EORTC QLU-C10D with its parent instrument, bear great opportunities for maturing health economic evaluations in the Netherlands.

Acknowledgements We would like to acknowledge the funding agencies for the Stepped Care trial (SC) and the Meaning Centered Group Psychotherapy trial (MCGP). SC: Supported by the Netherlands

Organization for Health Research and Development. Grant No. 300020012 MCGP: Supported by the Dutch Cancer Society. Grant No. 4864.

Authors contributions All authors have read and approved the manuscript.

Funding Open access funding provided by University of Innsbruck and Medical University of Innsbruck. This work was supported by funding from the EORTC Quality of Life Group: EORTC grant (#12/2016). The EORTC Quality of Life Group business model involves charges for commercial companies using EORTC instruments. Academic use of EORTC instruments is free of charge.

Data availability Data may be available upon reasonable request from the corresponding author.

Declarations

Conflict of interests The authors have no relevant financial or non-financial interests to disclose.

Ethical approval Ethical approval was obtained by the local ethical committee of the Medical University of Innsbruck (#20170531-1585).

Consent Patient consent sought in the original trials, covers the retrospective analysis of the trial data.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Neumann, P.J., Thorat, T., Shi, J., Saret, C.J., Cohen, J.T.: The changing face of the cost-utility literature, 1990–2012. *Value Health* **18**(2), 271–277 (2015)
2. Roudijk, B., Donders, A.R.T., Stalmeier, P.F.M.: Setting dead at zero: applying scale properties to the QALY model. *Med. Decis. Making* **38**(6), 627–634 (2016)
3. Joish, V.N.O.G.M.: Cost-utility analysis and quality adjusted life years. *J. Pain Palliat. Care Pharmacother.* **19**(1), 57–61 (2005)
4. Shamdas, M., Bassilious, K., Murray, P.I.: Health-related quality of life in patients with uveitis. *Br. J. Ophthalmol.* **103**(9), 1284–1288 (2019)
5. McDonough, C.M.T., Tosteson, A.N.A.: Measuring preferences for cost-utility analysis: how choice of method may influence decision-making. *Pharmacoeconomics* **25**(2), 93–106 (2007)
6. Brazier, J., Usherwood, T., Harper, R., Thomas, K.: Deriving a preference-based single index from the UK SF-36 Health Survey. *J. Clin. Epidemiol.* **51**(11), 1115–1128 (1998)
7. King, M.T., Costa, D.S., Aaronson, N.K., Brazier, J.E., Cella, D.F., Fayers, P.M., Grimison, P., Janda, M., Kemmler, G., Norman, R., Pickard, A.S., Rowen, D., Velikova, G., Young, T.A., Viney, R.: QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30. *Qual. Life Res.* **25**(3), 625–636 (2016)
8. Feeny, D., Furlong, W., Boyle, M., Torrance, G.W.: Multi-attribute health status classification systems. *Health Utilities Index. Pharmacoeconomics* **7**(6), 490–502 (1995)
9. Rabin, R., de Charro, F.: EQ-5D: a measure of health status from the EuroQol Group. *Ann. Med.* **33**(5), 337–343 (2001)
10. Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Bonse, G., Badia, X.: Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual. Life Res.* **20**(10), 1727–1736 (2011)
11. Rowen, D., Brazier, J., Ara, R., Azzabi, Z.I.: The role of condition-specific preference-based measures in health technology assessment. *Pharmacoeconomics* **35**(Suppl. 1), 33–41 (2017)
12. Drummond, M.S.M.: Common methodological flaws in economic evaluations. *Med. Care* **43**(7 Suppl.), 5–14 (2005)
13. Neumann, P.J., Goldie, S.J., Weinstein, M.C.: Preference-based measures in economic evaluation in health care. *Annu. Rev. Public Health. Rev. Public Health* **21**, 587–611 (2000)
14. Goodwin, E., Green, C.: A systematic review of the literature on the development of condition-specific preference-based measures of health. *Appl. Health Econ. Health Policy* **14**, 161–183 (2016)
15. Longworth, L., Yang, Y., Young, T., Mulhern, B., Hernández Alava, M., Mukuria, C., Rowen, D., Tosh, J., Tsuchiya, A., Evans, P., DevianeeKeetharuth, A., Brazier, J.: Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technol. Assess.* **18**(9), 201–224 (2014)
16. Versteegh, M.M., Leunis, A., Uyl-de Groot, C.A., Stolk, E.A.: Condition-specific preference-based measures: benefit or burden? *Value Health* **15**(3), 504–513 (2012)
17. King, M.T., Viney, R., Simon Pickard, A., Rowen, D., Aaronson, N.K., Brazier, J.E., Cella, D., Costa, D.S.J., Fayers, P.M., Kemmler, G., McTaggart-Cowan, H., Mercieca-Bebber, R., Peacock, S., Street, D.J., Young, T.A., Norman, R.: Australian utility weights for the EORTC QLU-C10D, a multi-attribute utility instrument derived from the cancer-specific quality of life questionnaire, EORTC QLQ-C30. *Pharmacoeconomics* **36**(2), 225–238 (2018)
18. Gamper, E.M., King, M.T., Norman, R., Efficace, F., Cottone, F., Holzner, B., Kemmler, G.: EORTC QLU-C10D value sets for Austria, Italy, and Poland. *Qual. Life Res.* **29**(9), 2485–2495 (2020)
19. Jansen, F., Verdonck-de Leeuw, I.M., Gamper, E., Norman, R., Holzner, B., King, M., Kemmler, G.: Dutch utility weights for the EORTC cancer-specific utility instrument: the Dutch EORTC QLU-C10D. *Qual. Life Res.* **30**(7), 2009–2019 (2021)
20. Kemmler, G., Gamper, E., Nerich, V., Norman, R., Viney, R., Holzner, B., King, M.: German value sets for the EORTC QLU-C10D, a cancer-specific utility instrument based on the EORTC QLQ-C30. *Qual. Life Res.* **28**(12), 3197–3211 (2019)
21. Finch, A.P., Gamper, E., Norman, R., Viney, R., Holzner, B., King, M., Kemmler, G.: Estimation of an EORTC QLU-C10 value set for Spain using a discrete choice experiment. *Pharmacoeconomics* **39**(9), 1085–1098 (2021)
22. McTaggart-Cowan, H., King, M.T., Norman, R., Costa, D.S.J., Pickard, A.S., Regier, D.A., Viney, R., Peacock, S.J.: The EORTC QLU-C10D: The Canadian valuation study and algorithm to derive cancer-specific utilities from the EORTC QLQ-C30. *MDM Policy Pract.* **4**(1), 238146831984253 (2019)
23. Nerich, V., Gamper, E.M., Norman, R., King, M., Holzner, B., Viney, R., Kemmler, G.: French value-set of the QLU-C10D, a

- cancer-specific utility measure derived from the QLQ-C30. *Appl. Health Econ. Health Policy* **19**(2), 191–202 (2021)
24. Norman, R., Mercieca-Bebber, R., Rowen, D., Brazier, J.E., Cella, D., Pickard, A.S., Street, D.J., Viney, R., Revicki, D., King, M.T.: European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group and the MAUCA Consortium. U.K. utility weights for the EORTC QLU-C10D. *Health Econ.* **28**(12), 1385–1401 (2019)
 25. Revicki, D.A., King, M.T., Viney, R., Pickard, A.S., Mercieca-Bebber, R., Shaw, J.W., Müller, F., Norman, R.: United States Utility Algorithm for the EORTC QLU-C10D, a multiattribute utility instrument based on a cancer-specific quality-of-life instrument. *Med. Decis. Making* **41**(4), 485–501 (2021)
 26. Richardson, J., Iezzi, A., Khan, M.A., Chen, G., Maxwell, A.: Measuring the sensitivity and construct validity of 6 utility instruments in 7 disease areas. *Med. Decis. Making* **36**(2), 147–159 (2016)
 27. Brazier, J.E., Rowen, D., Mavranzouli, I., Tsuchiya, A., Young, T., Yang, Y., Barkham, M., Ibbotson, R.: Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol. Assess.* **32**(16), 101–114 (2012)
 28. Pettitt, D.A., Raza, S.: The limitations of QALY: a literature review. *J. Stem Cell Res. Ther.* (2016). <https://doi.org/10.4172/2157-7633.1000334>
 29. Marra, C.A., Woolcott, J.C., Kopec, J.A., Shojania, K., Offer, R., Brazier, J.E., Esdaile, J.M., Anis, A.H.: A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Soc Sci Med* **60**(7), 1571–1582 (2005)
 30. Zeng, X., Sui, M., Liu, B., Yang, H., Liu, R., Tan, R.L., Xu, J., Zheng, E., Yang, J., Liu, C., Huang, W., Yu, H., Luo, N.: Measurement properties of the EQ-5D-5L and EQ-5D-3L in six commonly diagnosed cancers. *Patient* **14**(2), 209–222 (2021)
 31. Bulamu, N.B., Vissapragada, R., Chen, G., Ratcliffe, J., Mudge, L.A., Smithers, B.M., Isenring, E.A., Smith, L., Jamieson, G.G., Watson, D.I.: Responsiveness and convergent validity of QLU-C10D and EQ-5D-3L in assessing short-term quality of life following esophagectomy. *Health Qual. Life Outcomes* **19**(1), 233 (2021)
 32. Gamper, E.M., Cottone, F., Sommer, K., Norman, R., King, M., Breccia, M., Caocci, G., Patriarca, A., Palumbo, G.A., Stauder, R., Niscola, P., Platzbecker, U., Caers, J., Vignetti, M., Efficace, F.: The EORTC QLU-C10D was more efficient in detecting clinical known group differences in myelodysplastic syndromes than the EQ-5D-3L. *J. Clin. Epidemiol.* **137**, 31–44 (2021)
 33. Shaw, J.W., Bennett, B., Trigg, A., DeRosa, M., Taylor, F., Kiff, C., Ntais, D., Noon, K., King, M.T., Cocks, K.: A comparison of generic and condition-specific preference-based measures using data from nivolumab trials: EQ-5D-3L, mapping to the EQ-5D-5L, and European Organisation for Research and Treatment of Cancer Quality of Life Utility Measure-Core 10 Dimensions. *Value Health* **24**(11), 1651–1659 (2021)
 34. Mehdipour, A., Beauchamp, M.K., Wald, J., Peters, N., Kuspinar, A.: Measurement properties of preference-based measures for economic evaluation in COPD: a systematic review. *Qual. Life Res.* **29**(11), 2875–2885 (2020)
 35. Krahn, M., Bremner, K.E., Tomlinson, G., Ritvo, P., Irvine, J., Naglie, G.: Responsiveness of disease-specific and generic utility instruments in prostate cancer patients. *Qual. Life Res.* **16**(3), 509–522 (2007)
 36. Xie, F., Zoratti, M., Chan, K., et al.: Toward a centralized, systematic approach to the identification, appraisal, and use of health state utility values for reimbursement decision making: introducing the health utility book (HUB). *Med. Decis. Making* **39**(4), 371–379 (2019)
 37. Terwee, C.B., Bot, S.D., de Boer, M.R., van der Windt, D.A., Knol, D.L., Dekker, J., Bouter, L.M., de Vet, H.C.: Quality criteria were proposed for measurement properties of health status questionnaires. *J. Clin. Epidemiol.* **60**(1), 34–42 (2007)
 38. Rutten, M.J., van Meurs, H.S., van de Vrie, R., Gaarenstroom, K.N., Naaktgeboren, C.A., van Gorp, T., et al.: Laparoscopy to predict the result of primary cytoreductive surgery in patients with advanced ovarian cancer: a randomized controlled trial. *J. Clin. Oncol.* **35**(6), 613–621 (2017). <https://doi.org/10.1200/JCO.2016.69.2962>
 39. Homs, M.Y.V., Essink-Bot, M.-L., Borsboom, G.J.J.M., Steyerberg, E.W., Siersema, P.D.: Quality of life after palliative treatment for oesophageal carcinoma—a prospective comparison between stent placement and single dose brachytherapy. *Eur. J. Cancer* **40**(12), 1862–1871 (2004). <https://doi.org/10.1016/j.ejca.2004.04.021>
 40. Jansen, F., Krebber, A.M.H., Coupé, V.M.H., Cuijpers, P., de Bree, R., Becker-Commissaris, A., et al.: Cost-utility of stepped care targeting psychological distress in patients with head and neck or lung cancer. *J. Clin. Oncol.* **35**(3), 314–324 (2017). <https://doi.org/10.1200/JCO.2016.68.8739>
 41. van der Spek, N., Jansen, F., Holtmaat, K., Vos, J., Breitbart, W., van Uden-Kraan, C.F., et al.: Cost-utility analysis of meaning-centered group psychotherapy for cancer survivors. *Psychooncology* **27**(7), 1772–1779 (2018). <https://doi.org/10.1002/pon.4726>
 42. Giesinger, J.M., Efficace, F., Aaronson, N., Calvert, M., Kyte, D., Cottone, F., Cella, D., Gamper, E.M.: Past and current practice of patient-reported outcome measurement in randomized cancer clinical trials: a systematic review. *Value in Health* **24**(4), 585–591 (2021)
 43. EuroQol Group: EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* **16**(2), 199–208 (1990)
 44. Lamers, L.M., McDonnell, J., Stalmeier, P.F., Krabbe, P.F., Busschbach, J.J.: The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Econ.* **15**(10), 1121–1132 (2006)
 45. Omelyanovskiy, V., Musina, N., Ratushnyak, S., Bezdenezhnykh, T., Fediaeva, V., Roudijk, B., Purba, F.D.: Valuation of the EQ-5D-3L in Russia. *Qual. Life Res.* **30**(7), 1997–2007 (2021)
 46. Paveliu, M.S., Olariu, E., Caplescu, R., Oluboyede, Y., Niculescu-Aron, I.G., Ernu, S., Vale, L.: Estimating an EQ-5D-3L value set for Romania using time trade-off. *Int. J. Environ. Res. Public Health* **14**(18), 7415 (2021)
 47. EuroQol Group. EQ-5D-3L | Valuation: Overview of available standard value sets for EQ-5D-3L. [January 18, 2023]; Available from: <https://euroqol.org/eq-5d-instruments/eq-5d-3l-about/valuation/>.
 48. Akoglu, H.: User's guide to correlation coefficients. *Turk. J. Emerg. Med.* **18**(3), 91–93 (2018)
 49. Laird, B.J., Fallon, M., Hjermstad, M.J., Tuck, S., Kaasa, S., Klepstad, P., McMillan, D.C.: Quality of life in patients with advanced cancer: differential association with performance status and systemic inflammatory response. *J. Clin. Oncol.* **34**(23), 1769–1775 (2016)
 50. Prigerson, H.G., Bao, Y., Shah, M.A., Paulk, M.E., LeBlanc, T.W., Schneider, B.J., Garrido, M.M., Reid, M.C., Berlin, D.A., Adelson, K.B., Neugut, A.I., Maciejewski, P.K.: Chemotherapy use, performance status, and quality of life at the end of life. *JAMA Oncol.* **1**(6), 778–784 (2015)
 51. Zamel, O.N., Inocian, E.P., Alshehry, A.S., Tumala, R.B., Patalagsa, J.G., Alsaleh, K.A.: Quality of life among breast and colon cancer patients before and after first-cycle chemotherapy. *J. Holist. Nurs.* **39**(2), 116–125 (2021)

52. Wu, A.K., Cooperberg, M.R., Sadetsky, N., Carroll, P.R.: Health related quality of life in patients treated with multimodal therapy for prostate cancer. *J. Urol.* **180**(6), 2415–2422 (2008)
53. Schmidt, M.E., Scherer, S., Wiskemann, J., Steindorf, K.: Return to work after breast cancer: the role of treatment-related side effects and potential impact on quality of life. *Eur. J. Cancer Care* **28**(4), e13051 (2019)
54. Lai, W.W., Lin, C.N., Chang, C.C., Wang, J.D.: Lifetime risks, expected years of life lost, and cost-per-life year of esophageal cancer in Taiwan. *Sci. Rep.* **10**(1), 3722 (2020)
55. LankrewAyalew, T., Wale, B.G., Haile, K.E., Zewudie, B.T., Feleke, M.G.: Health-related quality of life and associated factors among cancer patients in Ethiopia: Systematic review and meta-analysis. *PLoS One* **17**(12), e0277839 (2022)
56. Sagberg, L.M., Jakola, A.S., Solheim, O.: Quality of life assessed with EQ-5D in patients undergoing glioma surgery: what is the responsiveness and minimal clinically important difference? *Qual. Life Res.* **23**(5), 1427–1434 (2014)
57. Cohen J. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.).
58. King, M.T., Bell, M.L., Costa, D., Butow, P., Oh, B.: The Quality of Life Questionnaire Core 30 (QLQ-C30) and Functional Assessment of Cancer-General (FACT-G) differ in responsiveness, relative efficiency, and therefore required sample size. *J. Clin. Epidemiol.* **67**(1), 100–107 (2014). <https://doi.org/10.1016/j.jclinepi.2013.02.019>
59. Tuley, M.R., Mulrow, C.D., McMahan, C.A.: Estimating and testing an index of responsiveness and the relationship of the index to power. *J. Clin. Epidemiol.* **44**(4–5), 417–421 (1991). [https://doi.org/10.1016/0895-4356\(91\)90080-s](https://doi.org/10.1016/0895-4356(91)90080-s)
60. Liang, M.H., Larson, M.G., Cullen, K.E., Schwartz, J.A.: Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum.* **28**(5), 542–547 (1985). <https://doi.org/10.1002/art.1780280513>
61. Corp, I.B.M.: *SPSS Statistics for Windows*. Amnonk, New York (2012)
62. Cicchetti, D.: Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* **6**(4), 284–290 (1994)
63. Pan, C.W., He, J.Y., Zhu, Y.B., Zhao, C.H., Luo, N., Wang, P.: Comparison of EQ-5D-5L and EORTC QLU-C10D utilities in gastric cancer patients. *Eur. J. Health Econ.* (2022). <https://doi.org/10.1007/s10198-022-01523-0>
64. Homs, M.Y., Steyerberg, E.W., Eijkenboom, W.M., Tilanus, H.W., Stalpers, L.J., Bartelsman, J.F., van Lanschot, J.J., Wijrdeman, H.K., Mulder, C.J., Reinders, J.G., Boot, H., Aleman, B.M., Kuipers, E.J., Siersema, P.D.: Single-dose brachytherapy versus metal stent placement for the palliation of dysphagia from oesophageal cancer: multicentre randomised trial. *The Lancet* **364**(9444), 23–29 (2004)
65. Penton, H., Dayson, C., Hulme, C., Young, T.: An investigation of age-related differential item functioning in the EQ-5D-5L using item response theory and logistic regression. *Value in Health* **25**(9), 1566–1574 (2022)
66. van der Spek, N., Vos, J., van Uden-Kraan, C.F., Breitbart, W., Cuijpers, P., Holtmaat, K., Witte, B.I., Tollenaar, R.A.E.M., Verdonck-de Leeuw, I.M.: Efficacy of meaning-centered group psychotherapy for cancer survivors: a randomized controlled trial. *Psychol. Med.* **47**(11), 1990–2001 (2017)
67. Chen, G.O.J.A.: Filling the psycho-social gap in the EQ-5D: the empirical support for four bolt-on dimensions. *Qual. Life Res.* **29**(11), 3119–3129 (2020)
68. van Gelder, T., Mulhern, B., Schoormans, D., et al.: Assessing health-related quality of life in cancer survivors: factors impacting on EORTC QLU-C10D-derived utility values. *Qual. Life Res.* **29**, 1483–1494 (2020)
69. Lorgelly, P.K., Doble, B., Rowen, D., Brazier, J.: Cancer 2015 investigators. Condition-specific or generic preference-based measures in oncology? A comparison of the EORTC-8D and the EQ-5D-3L. *Qual. Life Res.* **26**(5), 1163–1176 (2017)
70. NICE. *Guide to the methods of technology appraisal 2013: Process and methods*. London; 2013.
71. Franken, M., Koopmanschap, M., Steenhoek, A.: Health economic evaluations in reimbursement decision making in the Netherlands: time to take it seriously? *Z. Evid. Fortbild. Qual. Gesundheitswes. Fortbild. Qual. Gesundheitswes.* **108**(7), 383–389 (2014)
72. Janssen, M.F., Bonsel, G.J., Luo, N.: Is EQ-5D-5L better than EQ-5D-3L? a head-to-head comparison of descriptive systems and value sets from seven countries. *Pharmacoeconomics* **36**(6), 675–697 (2018). <https://doi.org/10.1007/s40273-018-0623-8>
73. Bulamu, N.B., Chen, G., Ratcliffe, J., Schloite, A., Bright, T., Watson, D.I.: Health-related quality of life associated with Barrett's Esophagus and cancer. *World J. Surg.* **43**(6), 1544–1562 (2019)
74. Kim, H., Cook, G., Goodall, S., Liew, D.: Comparison of EQ-5D-3L with QLU-C10D in metastatic melanoma using cost-utility analysis. *Pharmacoeconomics Open* **5**(3), 459–467 (2021)
75. Soare, I.A., Leeuwenkamp, O., Longworth, L.: Estimation of health-related utilities for 177Lu-DOTATATE in GEP-NET patients using utilities mapped from EORTC QLQ-C30 to EQ-5D-3L and QLU-C10D utilities. *Pharmacoeconomics Open* **5**(4), 715–725 (2021)
76. Jansen F, Coupé VMH, Eerenstein SEJ, Cnossen IC, van Uden-Kraan CF, de Bree R, Doornaert P, Halmos GB, Hardillo JAU, van Hinte G, Honings J, Leemans CR, Verdonck-de Leeuw IM. Cost-utility and cost-effectiveness of a guided self-help head and neck exercise program for patients treated with total laryngectomy: Results of a multi-center randomized controlled trial. *Oral Oncol.* 2021;117.
77. Stolk, E.A., de Bont, A., van Halteren, A.R., Bijlmer, R.J., Poley, M.J.: Role of health technology assessment in shaping the benefits package in The Netherlands. *Expert Rev. Pharmacoecon. Outcomes Res. Pharmacoecon. Outcomes Res.* **9**(1), 85–94 (2009)
78. Garattini, L.P.A.: Dutch guidelines for economic evaluation: “from good to better” in theory but further away from pharmaceuticals in practice? *R Soc. Med.* **110**(3), 98–103 (2017)

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

Authors and Affiliations

Micha J. Pilz^{1,2} · Simon Seyringer³ · Lára R. Hallsson² · Andrew Bottomley⁴ · Femke Jansen^{5,6,7} · Madeleine T. King⁸ · Richard Norman⁹ · Marianne J. Rutten¹⁰ · Irma M. Verdonck-de Leeuw^{5,6,7,11} · Peter D. Siersema^{12,13} · Eva Maria Gamper^{1,3}

✉ Eva Maria Gamper
eva.gamper@tirol-kliniken.at

¹ University Hospital of Psychiatry II, Medical University Innsbruck, Innsbruck, Austria

² Institute of Public Health, Medical Decision Making and Health Technology Assessment, Department of Public Health, Health Services Research and Health Technology Assessment, UMIT TIROL - University for Health Sciences and Technology, Hall, I.T., Austria

³ Department of Nuclear Medicine, Medical University of Innsbruck, 6020 Innsbruck, Austria

⁴ Quality of Life Department, European Organisation for Research and Treatment of Cancer, Brussels, Belgium

⁵ Department Otolaryngology-Head and Neck Surgery, Amsterdam UMC Location Vrije Universiteit Amsterdam, De Boelelaan 1117, Amsterdam, The Netherlands

⁶ Cancer Center Amsterdam, Treatment and Quality of Life, Amsterdam, The Netherlands

⁷ Department Clinical, Neuro and Developmental Psychology, Vrije Universiteit Amsterdam, Van Der Boechorststraat 7-9, Amsterdam, The Netherlands

⁸ School of Psychology, University of Sydney, New South Wales, Australia

⁹ School of Population Health, Curtin University, Perth, WA, Australia

¹⁰ Center of Gynaecologic Oncology Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands

¹¹ Amsterdam Public Health, Mental Health, Amsterdam, The Netherlands

¹² Department of Gastroenterology and Hepatology, Radboud University Medical Center, Nijmegen, The Netherlands

¹³ Department of Gastroenterology and Hepatology, Erasmus MC/University Medical Center, Rotterdam, The Netherlands