Journal of Plastic, Reconstructive & Aesthetic Surgery (2019) 72, 52–61

JPRAS
An International Journal of Surgical Reconstruction
www.JPRASurg.com

"The validity of the EQ-5D-5L in measuring quality of life benefits of breast reconstruction"☆

Casimir A.E. Kouwenberg[a,b,*], **Leonieke W. Kranenburg**[b], **Martijn S. Visser**[b], **Jan J. Busschbach**[b], **Marc A.M. Mureau**[a]

[a] *Department of Plastic and Reconstructive Surgery, Erasmus MC Cancer Institute, University Medical Center Rotterdam, PO Box 2040, Rotterdam 3000 CA, The Netherlands*
[b] *Department of Psychiatry, Section Medical Psychology and Psychotherapy, Erasmus MC, University Medical Center Rotterdam, PO Box 2040, Rotterdam 3000 CA, The Netherlands*

Received 14 March 2018; accepted 25 August 2018

KEYWORDS
Breast reconstruction;
EQ-5D;
Outcome assessment (health care);
Quality of life;
Mastectomy;
Cost-benefit analysis

Summary *Background:* The EuroQol EQ-5D-5L instrument is the most widely used quality of life (QoL) measure in health economic evaluations. It is unclear whether such a generic instrument is valid enough to estimate the benefits of breast reconstruction (BR), given the specific changes observed in QoL after BR. Hence, we aimed to evaluate the validity of the EQ-5D-5L in patients who had undergone postmastectomy BR.
Methods: In a 10-year cross-sectional cohort study, 463 mastectomy patients completed an online survey: 202 patients with autologous-BR (A-BR), 103 with implant-based-BR (I-BR), and 158 without BR (MAS). The results were used to evaluate the psychometric performance of the EQ-5D-5L with respect to the ceiling effect and to known-group, convergent, and discriminant validity, by comparing it with the Breast-Q, the cancer-specific (EORTC-QLQ-C30), and breast cancer-specific (EORTC-QLQ-BR23) questionnaires.
Results: The EQ-5D-5L was able to discriminate between patients with and without complications, MAS with or without BR and MAS versus the general population. It was, however, not able to discriminate between *A-BR* vs. *I-BR* as well as *BR* vs. *general population*. It is not clear whether this was due to the insensitivity of the instrument, insufficient sample sizes, or because there were no actual differences in QoL between these groups. Good convergent and discriminant validity of both the EQ-5D-5L and its individual dimensions were demonstrated. Additional support for the instrument's validity was revealed by moderate correla-

☆ Presented at the Fall meeting of the Netherlands Society for Plastic Surgery (NVPC), 8 October 2016, Rotterdam, The Netherlands.

* Corresponding author at: Department of Plastic and Reconstructive Surgery, Erasmus MC Cancer Institute, University Medical Center Rotterdam, PO Box 2040, Rotterdam 3000 CA, The Netherlands.
E-mail address: kouwenberg@gmail.com (C.A.E. Kouwenberg).

https://doi.org/10.1016/j.bjps.2018.08.023

tions between the generic EQ-5D-5L and specific QoL aspects of BR such as sexuality and body image.

Conclusions: The results of this study support the validity of the EQ-5D-5L as an outcome measure in health economic evaluations of BR.

Introduction

Health-care budgets are under substantial strain because of increasing health care costs. Society and insurance companies are progressively confronted with difficult choices about which medical interventions are to be reimbursed. Because elective procedures such as breast reconstruction (BR) are not life saving but primarily aimed at improving quality of life (QoL), they may be among the first medical interventions to be critically reviewed. Difficult decisions about which interventions should be reimbursed can be made only when it is possible to reliably compare different medical interventions. The formal way to do this is to perform a cost-effectiveness evaluation that makes use of appropriate measures such as the EQ-5D-5L.

The EQ-5D-5L is a widely used generic health-related QoL instrument, designed to measure the most important aspects of health over a broad spectrum of health conditions and diseases[1]. The instrument is especially used in health economic appraisals where comparisons are made between different therapeutic areas when deciding on the allocation of resources. In such comparisons, the core value of an intervention for the patients needs to be evaluated, that is, the effect of an intervention both on survival *and* QoL. In economic appraisals, survival *and* QoL are combined in the quality-adjusted life year (QALY). QALYs allow, for example, BR to be compared to an intervention under a condition such as diabetes. The QALY is the preferred outcome measure according to various guidelines for health economic evaluations from national reimbursement agencies such as NICE in the U.K.[2,3] The EQ-5D-5L is specifically designed to measure the Q in QALY and is notably different from other questionnaires used to measure QoL as it provides "preference-weighted quality of life scores" (utilities), on the basis of values of the general public. These utilities are needed for the calculation of QALYs. The EQ-5D is the most widely used questionnaire in health economic evaluations and is the preferred questionnaire of many national reimbursement agencies.[2,4] It is therefore important to evaluate whether the EQ-5D-5L is a valid instrument for the evaluation of QoL of BR patients.

Given the requirement to use appropriate and valid QALY estimates and the increased importance of health economic evaluations that provide comparable outcome measures, it is relevant for the field of BR surgery to know whether the generic EQ-5D-5L is a valid instrument to measure the specific benefits of BR. The present study aimed to evaluate the validity of the EQ-5D-5L for BR following mastectomy for breast cancer in a large cohort of patients with a long follow-up.

Methods

Patient recruitment

Data were gathered using a cross-sectional online survey sent to patients who in the last 10 years had been treated for breast cancer at Erasmus MC Cancer Institute in Rotterdam, the Netherlands. There were three cohorts of patients who had undergone mastectomy for breast cancer: autologous BR (A-BR), implant-based BR (I-BR), and women who had not undergone BR (MAS). Patients were identified using the hospital's reimbursement administrative system with specific codes for the respective procedures. Patients were sent an invitation letter through mail requesting participation in an online survey. Patients who were not proficient in Dutch or who had developed a distant metastasis were excluded. We considered including patients with a distant metastasis, but a large proportion of this patient group communicated that they did not wish to participate in research on this specific topic. Because the inclusion of this sub-population was not necessary with respect to the aim of the study, these patients were excluded on ethical grounds. Respondents filled out an online informed consent form and a series of self-administered questionnaires. Nonresponders were contacted 3 weeks later by telephone and asked to consider participating. The Medical Ethics Committee of the Erasmus MC approved the study (MEC-2015-273).[8]

Measures

EuroQol-5D-5L

The EQ-5D-5L is a standardized measure of health status designed to be a simple and generic measure of health-related QoL that can be used in clinical trials and economic evaluations of health care interventions.[9] It has a 5-dimension, 5-level descriptive system, covering the following dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. It describes $5^5 = 3125$ unique health states, which all have a utility value known from previous valuation studies. This utility value is anchored at two points, a value of 1 indicates "perfect health" and 0 indicates "death." In accordance with economic theory and health economic appraisal guidelines, we used the EQ-5D-5L societal utility (value) set specific to the study country, in this case the Netherlands, to score the questionnaire and obtain utility values for our sample.[2,10,11]

Breast-Q

The Breast-Q is a validated patient-reported outcome questionnaire that is widely used in the field of breast surgery.[12]

The modules specifically developed for BR and mastectomy were used if applicable. The following five domains of the Breast-Q were used in the current study: (1) physical well-being, (2) psychosocial well-being, (3) sexual well-being, (4) satisfaction with breasts, and (5) satisfaction with the overall outcome. The Breast-Q comes with an official score algorithm in the form of the "Q-score application." This application was used to transform the questionnaire responses to the respective modules on a 0–100-point scale where a higher score indicates a better outcome on the scales.

EORTC QLQ-C30 and QLQ-BR23

The EORTC questionnaires are measures for evaluating health-related QoL of patients with cancer which were designed for use in clinical trials.[13] In the present study, the cancer-specific QLQ-C30 and the breast cancer-specific QLQ-BR23 questionnaires were used. Both have been validated and are widely used in oncology and oncologic surgery patients.[13-15] The EORTC questionnaires consist of various scales where higher scale scores represent higher response levels. This means that a high score on one of the functional or QoL scales represents a high level of functioning or QoL, respectively. In contrast, a high score on one of the symptom scales indicates a high level of problems.[13,15]

Statistical methods

Characteristics of the study population were analyzed using descriptive statistics. The construct validity of the EQ-5D-5L in women who had undergone postmastectomy BR for breast cancer was evaluated. Construct validity is defined as the degree to which an instrument measures what it was intended to measure.[16] The construct validity of the EQ-5D-5L was tested by its correlation with other QoL instruments with known validity for BR and the ability of the EQ-5D-5L to discriminate between various relevant patient groups and outcomes. Three specific forms of construct validity were evaluated.

Distribution of the EQ-5D-5L health profiles

The distribution of the responses to the different EQ-5D-5L dimensions and the combination of these responses (the health profiles) within individual patients were assessed. This provides insight into the sensitivity of the instrument in terms of variance in scores on the dimensions and of the number of profiles. It also provides insight about a potential ceiling effect. This ceiling effect refers to the common observation that a large number of patients report "no problems" on any of the EQ-5D dimensions. This is often considered a psychometric problem, because it may imply an insufficiently sensitive questionnaire.[16] Hence, we investigated whether BR patients with a perfect health score on EQ-5D-5L also showed very good health scores on the Breast-Q well-being dimensions. To do this, an aggregated mean score of the Breast-Q "psychosocial", "chest and upper body," and "abdomen" well-being scores was calculated for each BR patient.

Known-group validity

The evaluation of known-group validity is based on the idea that distinctively different groups should score differently on the measure(s) or instrument(s) under evaluation, in this case the EQ-5D-5L. Known-group comparisons and hypotheses about the expected effects were formulated beforehand and were based on the literature and clinical experience. Patients who had not received BR after mastectomy, had experienced a complication, had received radiotherapy, or were of an older age were hypothesized to have a (relatively) lower QoL and therefore a lower EQ-5D-5L score. Patients with A-BR were hypothesized to have a higher QoL and higher EQ-5D-5L scores/values than I-BR patients.[17,18] The EQ-5D-5L outcomes for BR patients were not expected to significantly differ from the Dutch general population reference data. To test this hypothesis, we used the raw data from the official Dutch EQ-5D-5L valuation study. This is a large representative study with 1000 respondents (505 females) from the general public, which is now used as the mandatory reference study for EQ-5D-5L in health economic evaluations in the Netherlands.[2,10] A skewed EQ-5D-5L distribution score was expected, as the EQ-5D-5L is a generic QoL questionnaire, and most patients were expected to have few side effects by comparison with the impact of BR. Given the expected skewed distribution of outcomes, group comparisons were performed using the nonparametric Wilcoxon rank-sum (for two-group comparisons) and Kruskal–Wallis equality-of-populations rank tests (>2 groups).

We performed propensity score matching to control for differences in pretreatment patient characteristics in group comparisons directly related to the treatment modality by using the PSMATCH3 module of SPSS.[19,20] Three consecutive matching procedures were performed. First, the A-BR and I-BR cohorts were matched on pretreatment clinical and sociodemographic characteristics (Table 1). Subsequently, MAS was matched with the combined matched BR cohort on clinical characteristics, because sociodemographic characteristics were not available for all patients in this cohort. Finally, the Dutch general population reference sample was age and sex matched to the combined matched BR and MAS cohorts.

Convergent and discriminant validity

Convergent validity is based on the idea that items or scales that measure a similar concept should be strongly correlated to each other, whereas other items or scales that measure concepts that are unrelated should have a weak correlation to one another, which indicates discriminant validity. The convergent and discriminant associations were hypothesized beforehand and were assessed using nonparametric Spearman rank correlation coefficients. The following criteria for correlation strength, as formulated by Cohen, were utilized: weak for $0.1 \leq rs < 0.3$, moderate for $0.3 \leq rs < 0.5$, and strong for $rs \geq 0.5$.[21] For statistical testing, two-sided p-values ≤ 0.05 were considered statistically significant. Statistical analyses were performed using IBM® SPSS Statistics version 24 for Mac OSX.

Table 1. Sociodemographic and clinical characteristics of patient samples.

	Unmatched cohort				Matched cohorts			
	A-BR	I-BR	MAS	D-GP	A-BR	I-BR	MAS	D-GP
N	202	103	158	505	67	67	134	268
Age, median (SD)	55a (9.28)	53a,c (12.22)	63b (11.94)	47a,b (17)	55a (9.49)	55a (11.63)	61b (10.97)	58a,b (11)
Year of BC diagnosis, median (SD)	2008a (5.23)	2007a (6.64)	2003b (7.95)	N/A	2007a (6.27)	2006a,b (7.26)	2003b (7.85)	N/A
Year of mastectomy, median (SD)	2008a (4.84)	2007a (6.31)	2006b (6.35)	N/A	2008a (6.19)	2007a (6.87)	2007a (5.74)	N/A
Year of first BR, median (SD)	2011a (4.29)	2009a (7.18)	2010a,b (11.38)	N/A	2010a (5.77)	2009a (6.66)	2014a (2.13)	N/A
Year of last BR, median (SD)	2013a (2.89)	2012a (3.51)	2015a (1.15)	N/A	2013a (2.78)	2012a (4.21)	2015a (1.15)	N/A
Laterality mastectomy								
Unilateral	179a (88.6%)	48b (46.6%)	71a (83.5%)	N/A	46a (68.7%)	47a (70.1%)	64a (85.3%)	N/A
Bilateral	23a (11.4%)	55b (53.4%)	14a (16.5%)	N/A	21a (31.3%)	20a (29.9%)	64a (85.3%)	N/A
Reconstruction status								
Unilateral BR	175a (86.6%)	46b (44.7%)	1c (1.2%)	N/A	44a (65.7%)	45a (67.2%)	0 (0.0%)	N/A
Bilateral BR	26a (12.9%)	56b (54.4%)	0 (0.0%)	N/A	22a (32.8%)	22a (32.8%)	0 (0.0%)	N/A
Previously had a BR	1a (0.5%)	1a (1.0%)	9b (10.6%)	N/A	1a (1.5%)	0 (0.0%)	8b (10.7%)	N/A
Never had BR and does not want BR	0 (0.0%)	0 (0.0%)	70a (82.4%)	N/A	0 (0.0%)	0 (0.0%)	61a (81.3%)	N/A
Never had BR but wants BR	0 (0.0%)	0 (0.0%)	5a (5.9%)	N/A	0 (0.0%)	0 (0.0%)	5a (6.7%)	N/A
Patient-reported complications								
None	106a (52.5%)	46a (44.7%)	6b (3.8%)	N/A	37a (55.2%)	34a (50.7%)	4b (3.0%)	N/A
Yes	80a (39.6%)	47a (45.6%)	7b (4.5%)	N/A	24a (35.8%)	26a (38.8%)	7b (5.3%)	N/A
N/A	16a (7.9%)	10a (9.7%)	144b (91.7%)	N/A	6a (9%)	7a (10.4%)	122b (91.7%)	N/A
Breast Cancer Recurrence								
No recurrence	183a (90.6%)	82b (79.6%)	122b (77.2%)	N/A	57a (85.1%)	55a (82.1%)	101a (75.4%)	N/A
Local recurrence	10a (5.0%)	13b (12.6%)	21b (13.3%)	N/A	7a (10.4%)	8a (11.9%)	19a (14.2%)	N/A
Distant recurrence	9a (4.5%)	8a (7.8%)	15a (9.5%)	N/A	3a (4.5%)	4a (6.0%)	14a (10.4%)	N/A
Chemotherapy								
Yes	139a (68.8%)	52b (50.5%)	98a,b (62.0%)	N/A	40a (59.7%)	33a (49.3%)	82a (61.2%)	N/A
No	63a (31.2%)	51b (49.5%)	60a,b (38.0%)	N/A	27a (40.3%)	34a (50.7%)	52a (38.8%)	N/A
Radiotherapy								
Yes	74a (36.6%)	27a (26.2%)	84b (53.2%)	N/A	16a (23.9%)	19a (28.4%)	64b (47.8%)	N/A
No	128a (63.4%)	76a (73.8%)	74b (46.8%)	N/A	51a (76.1%)	48a (71.6%)	70b (52.2%)	N/A

(*continued on next page*)

Table 1. (*continued*)

	Unmatched cohort				Matched cohorts			
	A-BR	I-BR	MAS	D-GP	A-BR	I-BR	MAS	D-GP
N	202	103	158	505	67	67	134	268
Hormone therapy								
Currently undergoing treatment	49a (24.3%)	18a (17.5%)	58b (36.7%)	N/A	13a (19.4%)	11a (16.4%)	40a (29.9%)	N/A
Treated	66a (32.7%)	21a,b (20.4%)	22b (13.9%)	N/A	20a (29.9%)	14a (20.9%)	21a (15.7%)	N/A
Not treated	87a (43.1%)	64b (62.1%)	78a,b (49.4%)	N/A	34a (50.7%)	42a (62.7%)	73a (54.5%)	N/A
Employment status								
Yes, outdoor	117a (60.3%)	61a (64.2%)	N/A	N/A	43a (68.3%)	41a (67.2%)	N/A	N/A
Yes, inhome	19a (9.8%)	8a (8.4%)	N/A	N/A	3a (4.8%)	6a (9.8%)	N/A	N/A
No	58a (29.9%)	26a (27.4%)	N/A	N/A	17a (27%)	14a (23.0%)	N/A	N/A
Participation in social activities								
Rarely	16a (8.2%)	8a (8.2%)	N/A	N/A	5a (7.8%)	4a (6.5%)	N/A	N/A
Average	112a (57.1%)	49a (50.0%)	N/A	N/A	35a (54.7%)	28a (45.2%)	N/A	N/A
Often	68a (34.7%)	41a (41.8%)	N/A	N/A	24a (37.5%)	30a (48.4%)	N/A	N/A
Living arrangement								
1 person household	30a (15.5%)	19a (19.2%)	N/A	N/A	14a (22.2%)	14a (22.2%)	N/A	N/A
Multiperson household	164a (84.5%)	80a (80.8%)	N/A	N/A	49a (77.8%)	49a (77.8%)	N/A	N/A
Children in household								
Yes	103a (52.8%)	50a (51.0%)	N/A	N/A	36a (56.3%)	27a (42.9%)	N/A	N/A
No	92a (47.2%)	48a (49.0%)	N/A	N/A	28a (43.8%)	36a (57.1%)	N/A	N/A
Education								
Elementary school	0 (0.0%)	2a (2.1%)	N/A	N/A	0 (0.0%)	2a (3.2%)	N/A	N/A
Lower-level professional schooling	31a (16.3%)	10a (10.3%)	N/A	N/A	4a (6.5%)	6a (9.5%)	N/A	N/A
Mid-level high school	44a (23.2%)	19a (19.6%)	N/A	N/A	15a (24.2%)	13a (20.6%)	N/A	N/A
Mid-level professional schooling	41a (21.6%)	21a (21.6%)	N/A	N/A	14a (22.6%)	12a (19.0%)	N/A	N/A
Upper-level high school	23a (12.1%)	16a (16.5%)	N/A	N/A	6a (9.7%)	9a (14.3%)	N/A	N/A
Higher-level professional schooling	35a (18.4%)	22a (22.7%)	N/A	N/A	15a (24.2%)	15a (23.8%)	N/A	N/A
Academic schooling	16a (8.4%)	7a (7.2%)	N/A	N/A	8a (12.9%)	6a (9.5%)	N/A	N/A

Values in the same row and sub-table not sharing the same subscript are significantly different at $p < .05$ in the two-sided test of equality for column proportions. Cells with no subscript are not included in the test. BR, breast reconstruction; A-BR, autologous BR; I-BR, implant BR; MAS, mastectomy not followed by BR; and D-GP, Dutch age- and sex-matched reference population. A-BR, I-BR, and MAS cohorts were propensity score matched. D-GP cohort was age and sex matched (Versteegh, 2016).

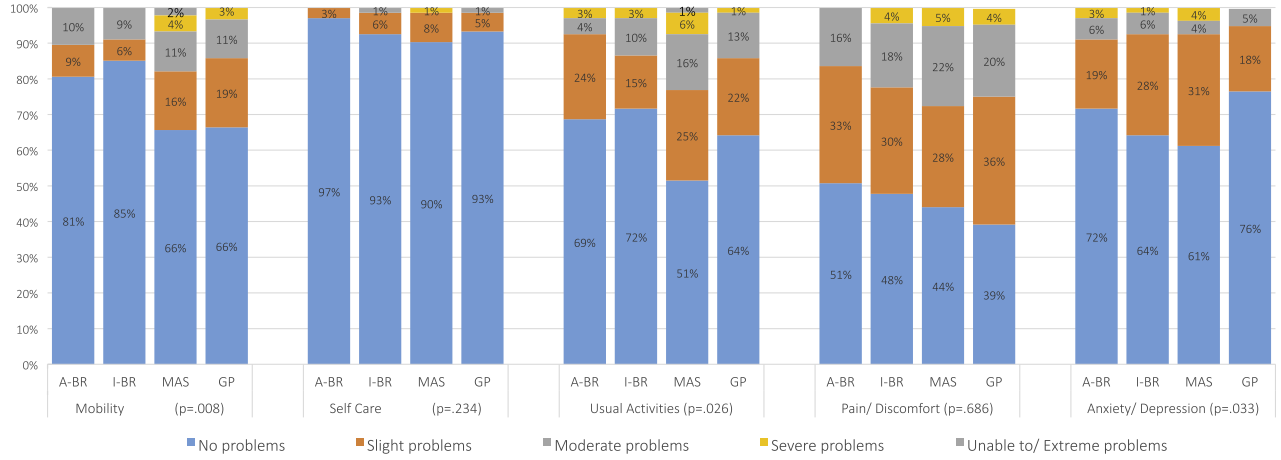

Figure 1. EQ-5D-5L responses in matched samples of A-BR ($n = 67$), I-BR ($n = 67$), and MAS patients ($n = 134$) and Dutch general population ($n = 268$). All comparisons were tested by the Kruskal-Wallis equality-of-populations rank test. A-BR, autologous breast reconstruction; I-BR, implant breast reconstruction; MAS, mastectomy not followed by breast reconstruction; and D-GP, Dutch age- and sex-matched reference general population. A-BR, I-BR, and MAS cohorts were propensity score matched. D-GP cohort was age and sex matched (Versteegh, 2016).

Results

Sociodemographic and clinical characteristics

The original unmatched cohorts consisted of 202 A-BR, 103 I-BR, and 158 MAS patients and showed a relatively large imbalance in age (with a disproportionate proportion older than 70 years in the MAS cohort) and laterality of the mastectomy, reconstruction status, breast cancer recurrence, chemotherapy, and hormone therapy. Table 1 shows the sociodemographic and clinical characteristics of both the unmatched and matched cohorts of A-BR and I-BR patients in addition to two reference cohorts that were used, a MAS cohort and an age- and sex-matched sample of the general population. The matching procedures resulted in a largely balanced cohort, with no statistically significant differences between the A-BR and I-BR cohort on the matched pretreatment patient characteristics.

Distribution of the EQ-5D-5L health profiles

Figure 1 illustrates the distribution of responses to the individual dimensions for all samples. One-to-one comparisons between the matched BR cohorts (i.e., A-BR and I-BR) did not show statistically significant differences on any of the EQ-5D-5L dimensions. Comparisons between the BR and Dutch general population (D-GP) cohorts on the individual dimensions also showed no statistically significant differences. Finally, comparisons between the BR and MAS cohorts showed substantial differences, which were statistically significant for the "mobility" and "usual activities" dimensions (both $p < .001$).

Table 2 depicts the most frequently occurring EQ-5D-5L health profiles in the different cohorts and allows the evaluation of a potential ceiling effect. The unmatched cohorts are presented in this table because comparisons are made within and not between patient groups. In total, 69 unique health profiles were reported in the BR cohort. Thirty-one percent of A-BR patients and 35% of I-BR patients reported no problems on any of the five dimensions (health profile, 11111), making it the most frequent health profile in these cohorts, similar to that of the D-GP. This ceiling effect was less pronounced among MAS patients where 24.7% reported perfect health in the unmatched cohort. To further explore this ceiling effect, an aggregated mean score of the Breast-Q "psychosocial", "chest and upper body," and "abdomen" well-being scores was calculated for each BR patient. We found that over three quarters of BR patients with an EQ-5D-5L score of 1 had an aggregated Breast-Q score of 80 or higher compared to only 22% of BR patients who had an EQ-5D-5L score lower than 1. This suggests that the ceiling effect on the EQ-5D-5L score represented patients who

Table 2. Most frequently occurring EQ-5D-5L health profiles in unmatched BR cohorts.

Health profile	A-BR	I-BR	MAS	D-GP
11111	63 (31.2%)	36 (35%)	39 (24.7%)	179 (35.4%)
11121	25 (12.4%)	9 (8.7%)	10 (6.3%)	63 (12.5%)
11112	12 (5.9%)	8 (7.8%)	9 (5.7%)	19 (3.8%)
11221	12 (5.9%)	5 (4.4%)	6 (3.8%)	21 (4.2%)
11122	11 (5.4%)	8 (7.8%)	5 (3.2%)	15 (3%)
11131	7 (3.5%)	3 (2.9%)	3 (1.9%)	7 (1.4%)
11222	3 (1.5%)	3 (2.9%)	4 (2.5%)	5 (1%)
21121	3 (1.5%)	2 (1.9%)	2 (1.3%)	5 (1.6%)

Health profile denoting the respective level of the following dimensions in the order: Mobility, Self-Care, Usual Activities, Pain/Discomfort, and Anxiety/Depression. 1 "no problems" up to 5 "severe problems/unable to." All health profiles that occurred five or more times in the unmatched BR cohort are listed. BR, breast reconstruction; A-BR, autologous BR; I-BR, implant BR; MAS, mastectomy not followed by BR; and D-GP, Dutch age- and sex-matched reference population.

Table 3. Known-group comparisons.

		EQ-5D-5L		
*Age**	Count	Mean	Standard deviation	*p*-value
< 50	387	0.867	0.17	0.87
50-60	244	0.8390	0.18	
60-70	222	0.831	0.18	
> 70	115	0.786	0.20	
Cohorts				
BR	305	0.844	0.18	
BR (matched)	134	0.863	0.16	
A-BR	202	0.840	0.18	
A-BR (matched)	67	0.872	0.14	
I-BR	103	0.851	0.17	
I-BR (matched)	67	0.853	0.18	
MAS	158	0.792	0.20	
MAS (matched)	134	0.798	0.20	
D-GP	268	0.841	0.16	
A-BR vs. I-BR				0.89 (matched 0.70)
BR vs. MAS				0.00 (matched 0.00)
BR vs. GP				0.49 (matched 0.15)
MAS vs. GP				0.01 (matched 0.03)
All groups				0.00 (matched 0.00)
*Patient-reported complications**				0.00
None	152	0.872	0.16	
Yes	127	0.806	0.19	
*Reconstruction status**				0.56
Unilateral BR	222	0.836	0.19	
Bilateral BR	82	0.859	0.15	
*Radiotherapy**				0.06
Yes	101	0.857	0.19	
No	204	0.837	0.17	

All comparisons were tested using the Kruskal–Wallis equality-of-populations rank test. Matched, propensity score matched; BR, breast reconstruction; A-BR, autologous BR; I-BR, implant BR; MAS, mastectomy not followed by BR; and D-GP, Dutch age- and sex-matched reference population (Versteegh, 2016). The group comparisons marked with * were performed on the unmatched BR cohort. The outcome values in this table are based on a sample from an academic hospital, may not be representative for the BR population as a whole, and are solely illustrative of the ability of the EQ-5D-5L to detect differences between relevant groups. Hence, these outcomes should not be used as the EQ-5D-5L reference values in scientific studies.

indeed experienced very few or no problems with regard to their BR-related well-being and consequently did not necessarily represent insensitivity of EQ-5D-5L to BR-related QoL problems.

Known-group validity

The above findings were based on the dimensions and profiles of the EQ-5D-5L, attributes common to QoL questionnaires. A key feature of the EQ-5D-5L is the utility score, which can be used in economic evaluations. The results based on this utility score are presented in an overview of the known-group comparisons in Table 3. Contrary to our hypothesis, no statistically significant differences were found between the A-BR and I-BR patient groups using the EQ-5D-5L utility scores. As hypothesized, BR yielded a statistically significant better QoL than MAS. Patients with breast cancer who had undergone BR did not show a statistically significant different QoL compared to those who had undergone D-GP. Patients who had experienced a complication following BR reported a statistically significant lower mean QoL than those who had not experienced a complication.

Convergent and discriminant validity

Table 4 shows the correlations of moderate strength or higher between the QoL of EQ-5D-5L and its individual dimensions on the one hand and the Breast-Q scales and EORTC scales for the unmatched BR cohorts on the other. Predefined hypotheses about the convergence and divergence of correlations were used to assess validity and are presented in Table 4.

Most scales relevant to BR surgery showed at least moderate correlation with the EQ-5D-5L. A notable exception to this finding was the Breast-Q dimensions that measured patient satisfaction with either the breast or outcome, because they showed weak correlations with the EQ-5D-5 L ($r = 0.345$ and $r = 0.327$, respectively).

Table 4. Convergent and discriminant validity between EQ-5D-5L dimensions, EQ-5D-5L, and other quality of life measures.

EuroQol-5D-5L Dimensions and Scores:	Mobility	Self-Care	Usual Activities	Pain/ Discomfort	Anxiety/ Depression	EQ-5D-5L NL
Breast-Q condition-specific QoL measure						
Psychosocial well-being	*-.288***	-.250***	-.335***	*-.390***	**-.501*****	**.524*****
Sexual well-being	*-.223***	-.249***	-.249***	-.240***	*-.417*****	*.401*****
Physical well-being: Chest and upper body	-.204***	-.249***	*-.409*****	**-.561*****	*-.266*****	**.516*****
Physical well-being: Abdomen	-.322***	-.191***	*-.427*****	*-.474*****	*-.228*****	*.484*****
EORTC QLQ-C30 cancer-specific QoL measure						
Global health status/QoL	-.332***	-.291***	*-.443*****	*-.464*****	*-.414*****	**.553*****
Physical function	**-.555*****	-.274***	**-.613*****	**-.505*****	*-.269*****	**.599*****
Role function	*-.428*****	-.266***	**-.690*****	**-.606*****	*-.276*****	**.634*****
Emotional function	*-.188*****	-.237***	*-.378*****	*-.382*****	**-.614*****	**.547*****
Cognitive function	-.213***	-.250***	*-.352*****	*-.353*****	*-.339*****	*.414*****
Social function	-.290***	-.246***	*-.465*****	*-.383*****	*-.389*****	*.497*****
Fatigue	.329***	.250***	**.530*****	**.522*****	*.384*****	**-.595*****
Pain	*.460*****	.294***	**.627*****	**.744*****	*.279*****	**-.704*****
Insomnia	.290***	.133*	*.398*****	*.412*****	.308***	*-.465*****
Appetite loss	.168***	.220***	.336***	.301***	.317***	*-.375*****
EORTC QLQ-B23 breast cancer-specific QoL measure						
Body image	*-.222*****	-.218***	-.313***	-.299***	*-.395*****	*.430*****
Future perspective	-.199***	-.239***	-.295***	-.288***	*-.485*****	*.442*****
Systemic therapy	.262***	.248***	*.400*****	*.384*****	.287***	*-.458*****
Breast symptoms	*.231*****	.235***	*.355*****	*.478*****	.188***	*-.435*****
Arm symptoms	.235***	.284***	*.490*****	*.499*****	.190***	*-.485*****

Only scales with at least one correlation of moderate strength (0.35-0.50) or higher are shown. Correlations that were hypothesized to show a convergent correlation are highlighted in dark gray with a fine border. Correlations that were hypothesized to show a discriminant correlation are highlighted in light gray with a thick border. Correlations of moderate strength (0.35-0.50) are shown in italics, and strong correlations (>0.50) are in bold. ** Correlation is significant at the 0.01 level (two-tailed). * Correlation is significant at the 0.05 level (two-tailed).

Only scales with at least one correlation of moderate strength (0.35-0.50) or higher are shown. Correlations that were hypothesized to show a convergent correlation are highlighted in dark gray with a fine border. Correlations that were hypothesized to show a discriminant correlation are highlighted in light gray with a thick border. Correlations of moderate strength (0.35-0.50) are shown in italics, and strong correlations (> 0.50) are in bold. ** Correlation is significant at the 0.01 level (two-tailed). * Correlation is significant at the 0.05 level (two-tailed).

Discussion

This is the first study to evaluate the validity of the EQ-5D-5L in patients who received BR after having undergone mastectomy for the treatment of breast cancer. Evaluation of the validity of this outcome measure is important as the EQ-5D-5L is currently the preferred QoL outcome measure in cost-effectiveness evaluations that inform healthcare policymakers and reimbursement agencies. The EQ-5D-5L was able to discriminate between several, but not all, patient groups and outcomes. Good convergent and discriminant validity of both EQ-5D-5L and its individual dimensions were demonstrated. Furthermore, additional support for validity was revealed by moderate correlations between the generic EQ-5D-5L and specific QoL aspects of BR such as sexuality and body image.

Distribution and ceiling effect of the EQ-5D-5L health profiles

One aspect on which the discriminative ability of a measure is frequently judged is its ability to detect differences between a given sample/cohort and the general population.[16] However, there was a large resemblance in the distribution of the EQ-5D-5L responses to the EQ-5D-5L of patients who had received BR and that of the age- and sex-matched cohort of the Dutch general population, with no statistically significant differences. In general, these findings would limit the validity of the outcome measure. However, in the case of BR, this may not necessarily be the case. Given that the overall aim of BR is to restore the QoL of patients with breast cancer to a level comparable with that before they were afflicted by breast cancer and that women eligible for BR may represent a relatively healthy patient group, outcomes comparable to the general population could be expected and were, indeed, hypothesized in this study. Further analysis of the distribution of responses (Figure 1 and Table 3) showed that the EQ-5D-5L can detect statistically significant differences between BR and MAS patients on both the "mobility" and "usual activities" EQ-5D-5L dimensions, thus indicating the sensitivity of the instrument. In our BR cohort, a considerable ceiling effect was found, which can be considered a psychometric problem in terms of sensitivity.[16] The EQ-5D dimensions might not tap into the relevant dimensions of QoL following BR, benefits of BR might be undetected, and the (cost-) effectiveness of BR would thus be underestimated. However, a ceiling effect may only represent a problem if it meant that the instrument is insensitive to problems actually present in the sample at hand. We found that the EQ-5D-5L ceiling effect represented patients who did indeed experience very few or no problems with respect to their BR-related well-being. Hence, we believe that the ceiling effect does not present a major problem in calculating a valid cost-effectiveness ratio in economic evaluations of BR.

Known-group validity

A-BR vs. I-BR

Currently, the only utilities available that differentiate between different BR techniques were obtained through expert opinion interviews with plastic surgeons, generally considered an inappropriate method for eliciting such values.[5-7] In these studies, surgeons estimated that A-BR resulted in the highest utility (0.83) followed by I-BR (0.66) and mastectomy not followed by BR (0.63).[17,18] However, in the known-group comparison, no significant differences were found on the EQ-5D-5L between A-BR or I-BR in either the matched or unmatched cohort. Because previous utility studies used controversial methods, it is difficult to determine whether the EQ-5D-5L was unable to detect differences between A-BR and I-BR because there were no substantial differences in QoL or because of the lack of sensitivity of the measure. As the EQ-5D-5L also showed convergent associations with the reference measures on four out of five dimensions, we consider the former more likely.

BR vs. MAS

The EQ-5D-5L was able to discriminate between BR and MAS patients both in an uncorrected and a matched cohort. Patients who had received BR after their mastectomy had significantly better EQ-5D-5L scores than those who had not received BR. This result corresponds with the findings of previous studies, which also reported a better QoL of BR patients than that of MAS patients, but conflicts with other studies that show little to no difference.[22-27] It appears that more recent studies have been more successful in finding significant differences between both patient groups, especially when making use of the Breast-Q questionnaire. This may be due to improved surgical techniques, improved sensitivity of QoL instruments, or a significant difference between groups may after all not exist. A systematic review or a meta-analysis comparing the QoL of MAS vs. BR could help inform us of the true effect on the QoL of BR.

Complications vs. no complications after BR

As hypothesized, patients who had experienced complications following their BR had a poorer mean QoL assessed with the EQ-5D-5L.

Convergent and discriminant validity

The EQ-5D-5L dimensions "mobility," "usual activities," "pain/discomfort," and "anxiety/depression" showed strong correlations with the domains and scales of ~~a~~ similar concepts on both the BR-specific Breast-Q and EORTC cancer- and breast cancer-specific measures, which implies good *convergent validity* of the EQ-5D-5L for BR. These correlations were considerably higher than those with dissimilar dimensions, which indicates good *discriminant validity* of the EQ-5D-5L for BR. The EQ-5D-5L dimension "self-care" showed only weak correlations with the other measures, which is probably explained by a lack of variance: the vast majority of BR patients (94.3%) reported no problems on this dimension. This result was also seen in the age- and sex-matched Dutch general population (Figure 1).

There are two important aspects of QoL in relation to BR, assessed by the Breast-Q and EORTC-BR23, that are worth highlighting because the EQ-5D-5L is potentially insensitive to these features, namely "sexuality" and "body image."[28,29] Both had correlations of moderate strength with the "anxiety/depression" scale.

Conclusions

The EQ-5D-5L was able to discriminate between various relevant patient groups and outcomes. It was not able, however, to discriminate between A-BR vs. I-BR and BR vs. general population. Convergent and discriminant validity of both the individual EQ-5D-5L dimensions and of the EQ-5D-5L were demonstrated by strong correlations with measures employing similar concepts. Furthermore, the EQ-5D-5L showed correlations of moderate strength with QoL aspects important to BR patients: sexuality and body image. In conclusion, the EQ-5D-5L showed sufficient validity to be used as one of the primary outcome measures in the evaluation of QoL outcomes in patients who have undergone postmastectomy BR for breast cancer treatment. The next step will be to obtain representative EQ-5D-5L reference values for this patient population.

Conflict of interest statement

None

Funding

Prof. Dr. Busschbach is a member of the nonprofit EuroQoL Group and receives financial compensation for managerial activities for the group.

References

1. Brazier J, Tsuchiya A. Preference-based condition-specific measures of health: what happens to cross programme comparability. *Health Econ* 2010;**19**(2):125–9.
2. National Health Care Institute (Zorginstituut Nederland Z. *Guideline for economic evaluations in healthcare* 2016. https://english.zorginstituutnederland.nl/publications/reports/2016/06/16/guideline-for-economic-evaluations-in-healthcare.
3. Excellence NIfC. *Guide to the methods of technology appraisal* 2013. https://www.nice.org.uk/process/pmg9/chapter/the-reference-case.
4. Brazier J, Longworth L. *NICE DSU technical support document 8: an introduction to the measurement and valuation of health for NICE submissions*. London: NICE Decision Support Unit; 2011.
5. Weatherly H, Drummond M, Claxton K, et al. Methods for assessing the cost-effectiveness of public health interventions: key challenges and recommendations. *Health Policy* 2009;**93**(2-3):85–92.
6. Briggs A, Sculpher M, Claxton K. *Decision modelling for health economic evaluation*. Oxford university press; 2006.
7. Sheckter CC, Matros E, Momeni A. Assessing value in breast reconstruction: a systematic review of cost-effectiveness studies. *J Plastic, Reconstr Aesthet Surg* 2018;**71**(3):353–65.
8. CCMO. *Your research: does it fall under the WMO* 2018. http://www.ccmo.nl/en/your-research-does-it-fall-under-the-wmo.
9. Oemar M, Janssen B, Rabin R, Oppe M, Herdman M. EQ-5D-5L user guide, version 2.0. Basic information on how to use the EQ-5D-5L instrument Rotterdam. *EQ-5D-5L user guide, version 2.0. Basic information on how to use the EQ-5D-5L instrument Rotterdam*, 28. EuroQol Group; 2013.
10. Versteegh MM, Vermeulen KM, Evers SM, de Wit GA, Prenger R, Stolk EA. Dutch tariff for the five-level version of EQ-5D. *Value Health* 2016;**19**(4):343–52.
11. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. Oxford university press; 2015.
12. Cohen WA, Mundy LR, Ballard TNS, et al. The BREAST-Q in surgical research: a review of the literature 2009–2015. *J Plastic, Reconstr Aesthet Surg* 2016;**69**(2):149–62.
13. Fayers PM, Aaronson NK, Bjordal K, Grønvold M, Curran D, Bottomley A. *EORTC QLQ-C30 scoring manual*; 2001. https://www.eortc.be/qol/files/SCManualQLQ-C30.pdf.
14. Chen CM, Cano SJ, Klassen AF, et al. Measuring quality of life in oncologic breast surgery: a systematic review of patient-reported outcome measures. *Breast J* 2010;**16**(6):587–597.
15. Aaronson NK, Ahmedzai S, Bergman B, et al. The European organization for research and treatment of cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993;**85**(5):365–376.
16. Peter M, Fayers DM. Scores and measurements: validity, reliability, sensitivity. *Quality of Life*. John Wiley & Sons, Ltd; 2015. p. 89–124.
17. Grover R, Padula WV, Van Vliet M, Ridgway EB. Comparing five alternative methods of breast reconstruction surgery: a cost-effectiveness analysis. *Plast Reconstr Surg* 2013;**132**(5):709e–723e.
18. Thoma A, Khuthaila D, Rockwell G, Veltri K. Cost-utility analysis comparing free and pedicled TRAM flap for breast reconstruction. *Microsurgery* 2003;**23**(4):287–95.
19. Thoemmes F. Propensity score matching in SPSS. 2012. arXiv: 1201.6385.
20. Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. *J Econ Surv* 2008;**22**(1):31–72.
21. Cohen J. *Statistical power analysis for the behavioral sciences*, Vol 2nd. New Jersey: Lawrence Erlbaum Associates; 1988.
22. Eltahir Y, Werners LL, Dreise MM, et al. Quality-of-life outcomes between mastectomy alone and breast reconstruction: comparison of patient-reported BREAST-Q and other health-related quality-of-life measures. *Plastic Reconstr Surg* 2013;**132**(2):201e–209e.
23. Jeevan R, Cromwell DA, Browne JP, et al. Findings of a national comparative audit of mastectomy and breast reconstruction surgery in England. *J Plastic, Reconstr Aesthet Surg* 2014;**67**(10):1333–44.
24. Howes BHL, Watson DI, Xu C, Fosh B, Canepa M, Dean NR. Quality of life following total mastectomy with and without reconstruction versus breast-conserving surgery for breast cancer: a case-controlled cohort study. *J Plastic, Reconstr Aesthet Surg* 2016;**69**(9):1184–91.
25. Lee C, Sunu C, Pignone M. Patient-reported outcomes of breast reconstruction after mastectomy: a systematic review. *J Am Coll Surg* 2009;**209**(1):123–33.
26. Ng SK, Hare RM, Kuang RJ, Smith KM, Brown BJ, Hunter-Smith DJ. Breast reconstruction post mastectomy: patient satisfaction and decision making. *Ann Plastic Surg* 2016;**76**(6):640–4.
27. Cortés-Flores AO, Morgan-Villela G, Zuloaga-Fernández del Valle CJ, et al. Quality of life among women treated for breast cancer: a survey of three procedures in Mexico. *Aesthet Plastic Surg* 2014;**38**(5):887–95.
28. Gopie JP, Hilhorst MT, Kleijne A, et al. Women's motives to opt for either implant or DIEP-flap breast reconstruction. *J Plast Reconstr Aesthet Surg* 2011;**64**(8):1062–7.
29. Fobair P, Stewart SL, Chang SB, D'Onofrio C, Banks PJ, Bloom JR. Body image and sexual problems in young women with breast cancer. *Psycho-Oncol* 2006;**15**(7):579–94.