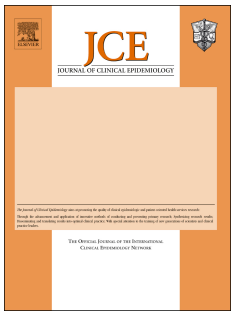
Accepted Manuscript



Psychometric development of the Upper Limb Lymphoedema Questionnaire demonstrated the patient-reported outcome measure to be a robust measure for breast cancer-related lymphoedema

Angela E. Williams, Frances Rapport, Ian T. Russell, Hayley A. Hutchings

PII: S0895-4356(17)30446-8

DOI: 10.1016/j.jclinepi.2018.04.014

Reference: JCE 9641

To appear in: *Journal of Clinical Epidemiology*

Received Date: 28 April 2017

Revised Date: 26 March 2018

Accepted Date: 13 April 2018

Please cite this article as: Williams AE, Rapport F, Russell IT, Hutchings HA, Psychometric development of the Upper Limb Lymphoedema Questionnaire demonstrated the patient-reported outcome measure to be a robust measure for breast cancer-related lymphoedema, *Journal of Clinical Epidemiology* (2018), doi: 10.1016/j.jclinepi.2018.04.014.

Author names and affiliations:

Angela E. Williams[a1], Frances Rapport[b], Ian T. Russell[a], Hayley A. Hutchings[a]

[a]Swansea University Medical School
Swansea University
Singleton Park
Swansea
SA2 8PP
UK

[b]Professor of Health Implementation Science
Macquarie University
Australian Institute of Health Innovation
75 Talavera Rd
North Ryde
Sydney, NSW
Australia

Email address of each author:

Angela E. Williams	awilliams42@sky.com
Frances Rapport	frapport@mq.edu.au
Ian T Russell	I.T.Russell@Swansea.ac.uk
Hayley A. Hutchings	H.A.Hutchings@Swansea.ac.uk

Corresponding author:

Angela E. Williams
[1]Permanent postal address for correspondence: 17 Eagle Lane, Dullingham, Newmarket, Cambridge, Cambridgeshire, CB8 9UZ
Email: awilliams42@sky.com
Phone: +44 (0)7950205570

(PROM) to assess health-related quality of life in patients with breast cancer-related upper limb lymphedema (ULL).

Study Design and Setting: We developed and validated the Upper Limb Lymphedema Quality of Life (ULLQoL) scale in two stages: devising the items and pre-testing with patients and clinicians; longitudinal validation to test its psychometric properties - underlying dimensions, internal consistency, test-retest reliability, construct validity, and responsiveness. Patients with ULL were recruited from two out-patient clinics.

Results: We derived the ULLQoL scale from a pool of 98 items generated by patients. After further consultation we produced the draft ULLQoL scale. For validation 103 patients with ULL completed the draft scale and two generic health measures: SF-36 and ED-5D-3L. Psychometric analysis identified two components, physical and emotional well-being, with good internal consistency and test-retest reliability. Significant correlations with SF-36, EQ-5D-3L, and percentage excess limb volume confirmed construct validity. The ULLQoL scale showed good responsiveness to change reported by lymphedema patients and moderate to large effect sizes.

Conclusion: The 14-item ULLQoL scale is a robust upper limb lymphedema-specific measure that is feasible and valid to use in both the clinical and research settings.

Keywords: quality of life, lymphedema, development, psychometric analysis, patient-reported outcome measure, breast cancer.

Running title: Measurement of breast cancer-related upper limb lymphedema quality of life

Abstract word count: 199 words

excess tissue oedema being the predominant feature during the early stages [2]. The limb is soft and pitting, and the excess tissue oedema may reduce when the limb is elevated [2]. As the condition advances, it can be differentiated from other types of oedema because the limb becomes firm, non-pitting and fibrotic, with deepened natural skinfolds.

There are primary and secondary causes of lymphedema, but in developed countries, lymphedema is usually secondary to treatment for cancer [3]. The focus of this study is upper limb lymphedema (ULL), the predominant cause of which is treatment for breast cancer. Lymphedema is a potential lifelong concern for women who survive breast cancer; estimates suggest "that about 20% of women will develop arm lymphedema after breast cancer" [4] (p. 500). Considerable morbidity is associated with ULL, as reported by studies of the lived experience of lymphedema, including impact on physical function, social function, emotional well-being and body-image [5-9].

Given the potential effect of ULL on patients' quality of life, 'Best Practice for the Management of Lymphedema' suggested that the assessment of patients should extend beyond the location and extent of limb swelling and include assessment of the 'morbidity impact', that is psychosocial status, mobility and function [3]. We know of eight lymphedema-specific measures of Health-related Quality of Life (HRQOL). Four of these are not specific to ULL, one is specific to cancer-related lymphedema, one is specific to breast-cancer-related lymphedema and two are specific to ULL (Table 1). However, none of these measures meet all of Terwee et al.'s (2007) quality criteria [10] and the COSMIN methodological checklist [11]. Indeed, four of the measures failed to meet any of these quality criteria, and methodological quality was generally rated as 'poor' or 'fair'.

The most robust measure specific to ULL when assessed against Terwee et al.'s (2007) [10] quality criteria and the COSMIN methodological checklist [11] is the ULL27, developed in French and sponsored by the French Lymphology Society. Although detailed information is available of the development of the ULL27 to support its content validity, psychometric analysis has only been reported from an interim analysis of the validation study [12], with no continuation of the work to provide the full analysis of this initial validation study. The development of the 38-item LYMQOL-arm has been reported along with information on its measurement properties [13]. Though the 38-item

The remaining incompleteness of the initial psychometric validation studies for the ULL27 and the LYMQOL-arm means that there is still a need for a well-validated tool to assess quality of life in patients with ULL. The need to reflect patients' experiences and views led us to develop a patient-reported outcome measure (PROM) specific to the assessment of HRQoL associated with ULL and to assess its psychometric measurement properties, including an estimate of the minimal clinical difference. We referred to the person's perception of well-being as it specifically and directly relates to their health as our definition of HRQoL in the development process [15].

2. Materials and methods

We followed the approach set out by Streiner et al. [16] for the development and validation of health measurement scales.

2.1 Devising the items and pre-testing

To derive the items for our ULL PROM we undertook a literature review and consulted a convenience sample of patients with ULL who were attending routine out-patient appointments at Sir Michael Sobell House Lymphedema Clinic in Oxford. All patients approached during the week agreed to participate in the study. We asked patients to list the five most important everyday activities that had been affected by ULL in the last month.

We grouped the activities into concepts of interest and relevance, and noted the number of activities in each concept. Items considered irrelevant to quality of life associated with ULL were discarded. We organised the remaining items into an initial scale (version 0) using a 5-point, adjectival rating scale which enabled patients to rate the effect of ULL on each item over the previous two weeks. To pre-test the scale we asked a different convenience sample of patients with ULL and 25 lymphedema specialist therapists to complete this initial version of the scale, and to comment on face and content validity, wording and acceptability. This feedback enabled us to update the scale to version 1 for the validation study.

2.2 Validation study

The validation study comprised a longitudinal study a test-retest study after two weeks. We worked with patients from two established out-patient Lymphedema Clinics in Cambridge. Patients aged 18

in the study.

At the initial visit we asked these consented participants to complete version 1 of the ULLQoL scale, a socio- demographic form, and two generic health outcome measures - the SF-36 [17] and EQ-5D-3L [18], and measured their limb volume. To assess the test-retest reliability of the ULLQoL scale we sent a subsample of participants a second questionnaire two weeks after their initial visit. We asked patients to complete and return all three questionnaires, and assess whether they judged that their condition had improved, remained the same, or deteriorated since their initial visit. We collected further data from participants at an outpatient visit three to six later, which was used to assess responsiveness. We asked them to complete all of the questionnaires again and comment on whether their condition had changed since their initial visit. We measured their limb volume again.

We completed a formal pilot lasting ten weeks to check recruitment rates. A retrospective review of the clinic diary suggested that one could recruit about two patients a week. So we expected to recruit 20 patients after ten weeks, and to complete recruitment after about one year. We also checked the appropriateness and feasibility of this design with the staff, and made a preliminary check whether they were using the full range of responses of the newly devised PROM.

Sample size

The literature recommends that for validating PROMs a sample size of between 50 and 100 patients is required [10, 16]. Nunnally [19] suggested a ratio of five to ten patients per item. Based on this, we aimed to recruit at least 100 patients over a one-year period for the purpose.

Ethical Considerations

All management staff at the two Lymphedema Clinics agreed to participate and to allow their patients' to participate in the study. When generating items, we sat down with each patient and explained the purpose of the task to the patients and invited them to give oral consent to take part which we then documented. We told patients that they were under no obligation to participate and that refusal would not affect their treatment within the clinic. The Cambridge Research Ethics Committee (LREC 97/165) gave a favourable ethical opinion on the validation study. We gave all potential participants a Patient Information Sheet describing the purpose of the study, what their

2.3 Statistical analysis

We analysed the data using the Statistical Package for Social Sciences (SPSS version 22).

2.3.1 Item selection and assessment of underlying dimensions

We considered items for inclusion in the ULLQoL scale by assessing their item completion rates and frequency of endorsement, and whether they performed consistently. We considered individual items for rejection if more than 5% of their data were missing [16]. We set the target endorsement rate at between 20% and 80%, and considered rejecting items outside the range, suggesting risk of 'floor' or 'ceiling' effects [16]. We used item-total correlations below 0.2 to identify items inconsistent with the others, and therefore candidates to be discarded [20].

We used principal component analysis (PCA) to assess the underlying dimensions of the ULLQoL scale. To check the suitability of the data for PCA, we initially undertook an unrotated exploratory PCA. We examined the Kaiser-Meyer-Olkin Index [21] and Bartlett's test of sphericity [22] using a threshold of 0.5 to explore whether the sample size was suitable [23]. In preliminary analysis we required that eigenvalues of important components should clearly exceed 1.0 [24]. We also suggested that factor loadings should exceed 0.4 to retain items in components [16]. Having determined the components, our next step was to improve these by rotation. We chose to use an oblique rotation of the factors to allow for the possibility that components were related rather than orthogonal to each other. An assessment of the face validity of items and components also contributed to the identification of the underlying principal components; items and components needed to be a recognisable aspect of a patient's health and well-being, the naming of the components being subjective.

2.3.2 Calculation of scores for ULLQoL scale

We imputed missing items by the lowest available score, thus underestimating the item score [16]. We calculated total scores and dimension scores for the ULLQoL scale by giving equal weight to each item and, summing the item scores. Thus the total score for the ULLQoL scale ranged from 0 to 56, with lower scores showing better QoL. As the physical well-being dimension comprised 9 items, scores potentially ranged from 0 to 36; as the emotional well-being dimension comprised 5 items, scores potentially ranged from 0 to 20.

evidence for a clinical tool [19]. We assessed test-retest reliability by analysing change in the ULLQoL scale scores of participants who reported that their lymphedema had not changed over time [26]. We chose an interval of two weeks to minimise a participant's ability to recall their previous response while maximising the chance that their health had indeed remained stable. We assessed reliability using the intra-class correlation coefficient, and used an alpha of 0.7 as the threshold for reliability.

We assessed construct validity of the ULLQoL scale by the strength and direction of the correlation with other attributes, both similar and dissimilar [16]. We categorised correlations greater than 0.7 as showing strong relationships; those between 0.5 and 0.7 as moderate relationships; those between 0.3 and 0.5 as weak relationships; and those less than 0.3 as very weak relationships [27]. From our clinical experience of ULL, we predicted moderate correlations between the physical well-being and emotional well-being dimensions of the ULLQoL scale; but very weak correlations between both dimensions of the ULLQoL scale and limb volume, the traditional measure of severity of lymphedema. It is not just the size of the swollen arm that could interfere with physical and emotional well-being, but also should stiffness, pain and swollen fingers and hand. Considering the content of the generic PROMs we also predicted moderate correlations between the ULLQoL scale physical well-being dimension and the EQ-5D-3L utility score, the EQ-5D thermometer and the physical well-being dimension of the SF-36; but weak correlation with the mental well-being dimension of the SF-36. We predicted weak correlations between the emotional well-being dimension of the ULLQoL scale and the EQ-5D-3L utility score, and the mental well-being dimension of the SF-36; but even weaker correlation with the physical component of the SF-36.

We assessed the responsiveness of the ULLQoL scale, that is its ability to detect change in patients' well-being, by examine correlation of changes scores as well as providing the effect size statistic using the Modified Standardised Response Mean (MSRM) for the ULLQoL. The MSRM is calculated by dividing the mean change score by the standard deviation of change in patients who report the lymphedema to be stable [28]. Following the criteria of Cohen [29] we categorised changes of the order of one-fifth of a standard deviation as small, those about one-half a standard deviation as moderate, and those about four fifths of a standard deviation as large. We conducted an initial estimate of the minimal important change (MIC) calculating the one-half standard deviation (SD).

female patients, aged between 48 and 72 years old and diagnosed with breast cancer-related lymphedema ULL, to suggest relevant items. They offered a total of 98 everyday activities affected by ULL. Two items were discarded as we considered them to be irrelevant to upper limb lymphedema-specific HRQoL - “worry about cancer returning” and “financial worries” both of which relate to cancer rather than lymphedema. Both of the items were also only identified by one patient. With the remaining 96 items we went one-by-one and grouped similar concepts. We derived 12 distinct concepts and the number of activities in each concept ranged from 3 to 18 activities. We then converted the concepts into questions about respondents’ experiences: eight starting: " During the past two weeks to what extent _ _ _ ?" and four starting: "During the past two weeks how much of the time _ _ _ ?" (ULLQoL version 0). We considered a recall interval of two-weeks was appropriate to measure the concept of HRQoL and for the condition [30]. Furthermore, we considered the target group to be capable of recalling the past two weeks in relation to the item content. For response categories we adopted a 5-point scale: for the eight magnitude questions – “not at all”, “a little bit”, “moderately”, “quite a bit” and “extremely”; and for the four frequency questions: “none of the time”, “some of the time”, “a good bit of the time”, “most of the time” and “all of the time”.

To pre-test the ULLQoL scale (version 0) the first five patients who attended a routine lymphedema out-patient appointment all agreed to participate, as well as 16 of 25 (64%) lymphedema specialist therapists. The patients confirmed face and content validity and made no recommendations for changes to item wording, finding all of the items acceptable. Unaided completion of version 0, including reading the instructions took between two and four and a half minutes. The lymphedema specialist therapists suggested the addition of an item to capture relationship difficulties with their partners to complete content validity as they reported that this arises in conversation with patients in routine clinical practice and affects their HRQoL. An item asking ‘to what extent did your swollen arm interfere with your sexuality’ was added to the PROM. The nurses also made two comments on item wording: they suggested adding the word ‘swollen’ to ‘arm’ for clarification; and, as patients are advised not to carry heavy objects, they suggested that the wording of the item on lifting should focus on the carrying of light objects. Following this pre-testing we revised the ULLQoL scale and took version 1, a 13-item scale was taken into the validation study.

visit after three to six months for responsiveness.

In a pilot lasting ten weeks, we recruited 14 patients of the target of 20 patients. Interviews with the staff confirmed that the tested processes were suitable for data collection. A review of the completed 13-item ULLQoL scale (version 1) suggested that the wording of the item on sexuality may be inadequate to describe the concept in mind. So we changed its wording to replace the word 'sexuality' to 'femininity/masculinity' aiming to more accurately reflect the concept under consideration by this question. We also responded to feedback from participants and clinical staff by adding a 'frequency' question to capture the general effect of lymphedema on mood. We used the revised 14-item scale (ULLQoL version 2) for the remainder of the study.

Of the 103 recruited patients, 100 (97%) were female, with a mean age of 60 years (SD 13). Their ages spanned seven decades (Table 2). In 28 patients (27%) the swollen arm was less than 5% larger than the unaffected arm; in 37 patients (36%) the swollen arm was between 5% and 20% larger than the unaffected arm; in 29 patients (28%), the excess limb volume was between 21 and 40%; and for nine patients (9%) it was greater than 40%. The majority of patients had developed ULL secondary to breast cancer (95). . Other causes of ULL were non-Hodgkin's lymphoma (1 patient), primary lymphoedema (1 patient), and secondary to infection (2 patients). There were missing data for 4 of the patients. Fifty-five patients (53%) had lymphedema of their dominant arm. The mean time between onset of ULL and recruitment into the study was 37 months (SD 12).

3.3 Statistical analysis

3.3.1 Item selection and assessment of underlying dimensions

The 103 patients completed more than 95% of ULLQoL items (Table 3). Apart from the 14 pilot participants who necessarily completed version 1 rather than version 2 at baseline, one of the 103 participants did not complete three items (Q4, Q5, Q7) at baseline as they missed the back of the double-sided sheet; and four participants did not complete Q9 at baseline for unknown reasons. At the second visit all 103 patients completed version 2. In total we needed to impute the lowest score for only 35 items across 206 questionnaires. All items showed the full range of responses with no floor or ceiling effects (Table 3). The item-total correlation coefficients for the 14 items ranged

Initial analysis confirmed that the data were suitable for PCA, with KMO Index of 0.83 and Bartlett's test significant at the 0.1% level. The unrotated exploratory analysis generated three components explaining 62% of the total variance. Some items loaded clearly on one component; but others had smaller loadings on more than one component. The scree-plot showed a clear elbow after two components, with components 1 and 2 explaining 54% of the total variance. Furthermore, the third component was uninterpretable, and only added 8% to the total variance explained. Hence, we opted for two components, within which most items loaded clearly.

We then proceeded to oblique rotation, which allows for correlated components. Table 4 shows the resulting component loadings, the eigenvalues for each component and the variation explained by each component. Component 1 comprises items about 'physical well-being' aspects of ULL, while Component 2 addresses 'emotional well-being'. The loadings exceeded 0.4 for all but two items: (9) "Do you feel down because of swollen arm?"; and (12) "How much time does your swollen arm cause pain or discomfort?" which had similar borderline loadings on both components.

To allocate items 9 and 12, we observed the slightly higher item-total correlations when Q9 was in the emotional Component 2 and Q12 was in the physical Component 1 (Table 5), which also seems to improve face validity. All the resulting item-total correlation coefficients exceeded the threshold of 0.4 for removal. Thus, we recommend that Component 1 comprise nine physical well-being items and recommended Component 2 comprise five emotional well-being items.

3.3.2 Calculation of scores of ULLQoL scale

Total scores for the ULLQoL scale ranged from 0 to 49 with a mean score of 18.7 and standard deviation (SD) of 10.7. Though the distribution of scores was positively skewed at 0.50, the discrepancy was not substantial (Table 6).

3.3.3 Psychometric properties of the ULLQoL scale

The internal consistency of the ULLQoL scale achieved Cronbach's alpha of 0.87, well above our target of 0.7. The two components also achieved good internal consistency: alpha was 0.87 for physical well-being and 0.77 for emotional well-being. The test-retest reliability of the ULLQoL scale

The ULLQoL scale showed good construct validity with all eight predicted correlations with other PROMs achieving statistical significance and all correlations in the predicted direction. (Table7). Overall, the observed correlation between the two dimensions of the ULLQoL and the EQ-5D-3L utility score was as predicted, but the correlation between the ULLQoL physical dimension and the EuroQoL thermometer were smaller than predicted. The predicted correlations between the dimensions of the ULLQoL and the SF-36 PCS and MCS were all observed, except the correlation between the ULLQoL physical dimension and the MCS, which was larger than expected. The observed correlation for the two dimensions of the ULLQoL with % excess limb volume was weak as predicted.

Of the 85 participants who completed the lymphoedema health transition question as well as the baseline and final questionnaires, 42 (49%) perceived their lymphedema as improved, and 33 (39%) perceived it as little changed, while ten (12%) perceived it as worse. Table 8 shows that the first group reported clearly improved ULLQoL scores (functional and emotional well-being) while the third group reported that all mean scores deteriorated. Furthermore, these changes reflected analogous changes in limb volume, with a decrease in limb volume when the patients reported improvement on the lymphoedema transition question, no change in limb volume when patients reported no change on the transition question, and increase in limb volume when they reported deterioration on the lymphedema transition question. All tests for linearity were significant at the 1% level. Table 9 shows moderate to large effect sizes for the responsiveness of the ULLQoL scale. The one-half SD was 5.35 suggesting the minimal important change was in this range.

4. Discussion

The final version 2 of the ULLQoL scale showed good item completion, wide ranges of responses and consistency between items so we retained all 14 items. Only 35 items of data were missing and these were imputed with the lowest item value, giving rise to a very small potential bias. A variety of methods for handling missing data and imputation alternatives have been used to manage missing quality of life data [31-33]. All methods of imputation will introduce bias and arguably none of them are completely satisfactory [16]. Imputation of the lowest score is an acceptable method [16] that would have introduced a very small bias towards the better quality of life which resonates with our clinical experience and findings of qualitative research [34]. Nonetheless, the authors consider that

Psychometric analysis identified two believable dimensions –physical well-being (9 items) and emotional well-being (5 items). Given the dearth of literature at the time of instrument development, the scale structure of the ULLQoL relied on the findings of exploratory PCA and a definition of HRQoL as opposed to setting up a hypothesis-driven factor structure to be tested against the data. Ideally, this first psychometric evidence needs confirmatory analysis on another cohort of patients with ULL, to explore the factor structure further and specifically investigate the heterogeneity of the physical well-being dimension. Using the findings of in-depth interviews, our understanding of the impact of ULL on HRQoL was extended [34]. The finding that symptoms of upper limb lymphoedema had an array of physical, emotional and social impacts on HRQoL supports the item content validity and scale structure of the ULLQoL [34]. In addition, the labelling of the physical well-being dimension may be reviewed in future validation studies to ensure there is no problem with the naming given its use in other PROMs such as the SF-36, with additional qualitative research to validate the items and structure of the ULLQoL.

The resulting upper limb lymphedema-specific QoL measure showed excellent internal consistency and test-retest reliability, and good construct validity, with the generic SF-36 and EQ-5D-3L as constructs. Nevertheless, we recommend further analysis of the test-retest reliability using an objective measure for stability, and further larger studies of construct validity, particularly against more specific constructs, for example, a clinician's global assessment of well-being, or severity of lymphedema using the International Society of Lymphology's staging system [3]. The vast majority of the patients in the validation study had developed ULL secondary to breast cancer. The authors consider that the ULLQoL may be appropriate for a broader target group with ULL, but this would have to be tested in further validation studies.

Before our definitive analysis we predicted only very weak correlation between the physical and emotional dimensions of the ULLQoL scale and limb volume, the usual measures of severity of lymphedema. Both Passik et al. [35] and Woods [5] had reported that limb volume did not influence patients' emotional well-being. Fortunately, we observed the predicted very weak correlation. Though the assessment of limb size is important in clinical assessment of ULL, the ULLQoL scale now provides complementary, holistic assessment that takes full account of the wide range of morbidity.

swollen limb. Future investigation of the minimal clinically important differences on the ULLQoL scale would enhance this finding.

In conclusion, the first evidence of psychometric properties for the ULLQoL scale suggest the measure is a valid, reliable, consistent and responsive PROM for assessing HRQoL in breast cancer-related lymphoedema. It is the most robust measure for ULL as assessed by Terwee's quality criteria [10] and the COSMIN checklist [11]. It is short, takes less than five minutes to complete, and is easy to use in clinical practice and research. We therefore recommend that 'Best Practice for the Management of Lymphedema' [3] use the ULLQoL to improve clinical assessment. Furthermore, the regular collection of ULLQoL data has the potential to improve patient-clinician communication and even patients' HRQoL. To this end, the ULLQoL scale is freely available for use in clinical practice or research.

Acknowledgements:

The authors thank: Dr Andrew Garratt for advice on devising a PROM; the staff of the two Lymphedema Clinics, for their help in recruiting patients; and the patients who participated, without whom the study would not have been possible.

Funding:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

International Congress of Lymphology working group discussions and recommendations, Adelaide, Australia, August, 1985. Lymphology 1985;18:175-80.
[3] Lymphedema Framework. Best Practice for the Management of Lymphedema. International consensus. London: MEP Ltd,2006.
[4] DiSipio, T, Rye S, Newman B, Hayes S. Incidence of unilateral arm lymphedema after breast cancer: a systematic review and meta-analysis. Lancet Oncology. 2013;14:500-15.
[5] Woods M. Patients' perceptions of breast-cancer-related lymphedema. European Journal of Cancer Care. 1993;2:125-8.
[6] Johansson K, Holmstrom H, Nilsson I, Ingvar C, Albertsson M, Ekdahl C. Breast cancer patients' experiences of lymphedema. Scandinavian Journal of Caring Sciences. 2003;17:35-42.
[7] Williams AE, Moffatt CJ, Franks PJ. A phenomenological study of the lived experiences of people with lymphedema. International Journal of Palliative Nursing. 2004;10:279-86.
[8] Towers A, Carnevale FA, Baker ME. The psychosocial effects of cancer-related lymphedema. Journal of Palliative Care. 2008;24:134-43.
[9] Fu MR, Rosedale M. Breast cancer survivor's experiences of lymphedema-related symptoms. Journal of Pain and Symptom Management. 2009;38:849-59.
[10] Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, et al. Quality criteria proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology. 2007; 60:34-42.
[11] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of Life Research. 2010; 19:539-49.
[12] Launois R, Alliot F. Quality of Life Scale in Upper Limb Lymphedema – a validation study. Lymphology. 2000;33:226-74.
[13] Keeley VL, Veigas D, Crooks S, Locke J, Forrow H. The development of a condition-specific quality of life measure for lymphedema (LYMQOL). European Journal of Lymphology; 2004;12:36.
[14] Keeley VL, Crooks S, Locke J, Veigas D, Riches K, Hilliam R. A quality of life measure for limb lymphedema (LYMQOL). Journal of Lymphedema. 2010;5:26-37.
[15] Guyatt GH, Feeny DH, Patrick DL. Measuring Health-related Quality of Life. Annals of Internal Medicine. 1993;118:622-9.
[16] Streiner DL, Norman GR, Cairney J. Health measurement scales. Oxford, Oxford University Press, 2015.
[17] Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36). Medical Care. 1992;30:473-83.
[18] EuroQol Group. EuroQol – a new facility for the measurement of health-related quality of life. Health Policy. 1990;16:199-208
[19] Nunnally JC. Psychometric theory. New York, New York:McGraw-Hill; 1978
[20] Garratt AM, Macdonald M, Ruta DA, Russell IT, Buckingham JK, Krukowski ZH. Towards measurement of outcome for patients with varicose veins. Quality of Health Care. 1993;2:5-10.
[21] Kaiser HF. A second-generation little jiffy. Psychometrika. 1970;35:401-15.
[22] Bartlett MS. A note on the multiplying factors for various chi square approximations. Journal of Royal Statistical Society 16(Series B). 1954;296-298.
[23] Williams B, Brown T, Onsman A. Exploratory factor analysis: a five-step guide for novices. Journal of Emergency Primary Health Care. 2010;8-Article 990399.
[24] Joliffe IT, Morgan BJT. Principal component analysis and exploratory factor analysis. Statistical Methods in Medical Research. 1992;1:69-95.

[27] Garratt AM, Hutchinson A, Russell I. The UK version of the Seattle Angina Questionnaire (SAQ-UK): Reliability, validity and responsiveness. Journal of Clinical Epidemiology. 2001;54:907-15.
[28] Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Diseases. 1987;40(2):171-8.
[29] Cohen J. Statistical power analysis for the behavioral sciences. New York, New York:Academic Press, 1977.
[30] Norquist JM, Girman C, Fehnel S, DeMuro-Mercon C, Santanello N. Choice of recall period for patient-reported outcome (PRO) measures: criteria for consideration. Quality of Life Research. 2012;21:1013-20.
[31] Fielding S, Ogbuagu A, Sivasubramaniam S, MacLennan G, Ramsay CR. Reporting and dealing with missing quality of life data in RCTs: has the picture changed in the last decade? Quality of Life Research. 2016;25:2977-83
[32] Rombach I, Rivero-Arias O, Gray AM, Jenkinson C, Burke O. The current practice of handling and reporting missing outcome data in eight widely used PROMs in RCT publications: a review of the current literature. Quality of Life Research. 2016;25:1613-23.
[33] Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, CONSORT PRO Group. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA. 2013;309(8):814-22.
[34] Williams AE. 2016, Understanding and measuring health-related quality of life associated with upper limb lymphoedema: A mixed method research approach. PhD thesis. University of Swansea, Swansea.
[35] Passik S, Newman M, Brennan M, Holland J. Psychiatric consultation for women undergoing rehabilitation for upper extremity lymphedema following breast cancer treatment. Journal of Pain and Symptom Management. 1993;8:226-33.

WCLS	Mirolo et al.	Cancer-related	1995	5	No named domain
FACT-B+4	Coster et al.	Breast cancer-related arm morbidity	2001	5	No named domain
ULL27	Launois et al.	Upper limb lymphoedema	2000	27	Physical Psychological Social
Weiss and Spray QoL	Weiss & Spray	All lymphoedema	2002	17	Physical Psychological Functional
LQOLI	Kristjanson (not published)	All lymphoedema	2004	188	Physical Emotional Social Practical
LYMQOL-arm	Keeley et al.	Upper limb lymphoedema	2004	38	Symptoms Body image & appearance Function Mood
FLQA-l	Augustin et al.	All lymphoedema	2005	92	Physical complaints Everyday life Social life Emotional well-being Stress due to management Satisfaction with different areas of life Satisfaction with work &household
LyQLI	Klernas et al.	All lymphoedema	2015	45	Physical Emotional Social Practical

Gender:	
Female	100 (97%)
Male	3 (3%)
Age:	
20-29 years	1 (1%)
30-39 years	3 (3%)
40-49 years	15 (15%)
50-59 years	34 (33%)
60-69 years	20 (19%)
70-79 years	21 (20%)
80 -89 years	9 (9%)
Mean age (SD)	60.5 years (13.0)
Range	23 to 86 years
Percentage excess limb volume	
<5%	28 (27%)
0 - 20%	37 (36%)
21 - 40%	29 (28%)
≥41%	9 (9%)
Cause of upper limb lymphoedema:	
Breast cancer	102 (99%)
Non-Hodgkin's lymphoma	1 (1%)

Q1. Interfere with normal work activities	0	24.3	2.9
Q2. Interfere with normal leisure activities	0	27.2	4.9
Q3. Interfere with self-care	0	46.6	1.0
Q4. Interfere with dressing and undressing	1 (0.97)	57.8	1.0
Q5. Interfere with sleep	1 (0.97)	34.3	2.0
Q6. Influences femininity or masculinity vs3 §	0	61.8	2.2
Q7. Interfere with lifting and carrying	1 (0.97)	10.8	15.7
Q8. Interfere with reaching high objects	0	23.3	14.6
Q9. Feel down because of swollen arm vs3§	4 (4.40)	35.9	1.0
Q10. Feel down because restricts clothing you would like to wear	0	44.7	8.7
Q11. How much time swollen arm stops you wearing clothing would like to wear	0	41.7	9.7
Q12. How much time swollen arm causes pain or discomfort	0	13.6	8.7
Q13. How much time feel aware of swollen arm	0	12.6	13.6
Q14. How much time feel self-conscious or embarrassed in public	0	55.3	6.8

§ (Q6 and Q9 only included in the 14-item version which was issued after the pilot period to 91 patients)

Items	Physical well-being	Emotional well-being
1. To what extent did your swollen arm interfere with your normal work activities?	0.818	
2. To what extent did your swollen arm interfere with your normal leisure activities?	0.739	
3. To what extent did your swollen arm interfere with your self-care?	0.852	
4. To what extent did your swollen arm interfere with dressing and undressing?	0.780	
5. To what extent did your swollen arm interfere with your sleep?	0.665	
6. To what extent did you feel your femininity or masculinity was influenced by your swollen arm?		0.664
7. To what extent did your swollen arm interfere with lifting or carrying light objects?	0.674	
8. To what extent did your swollen arm interfere with reaching high objects?	0.851	
9. To what extent did you feel down because of your swollen arm?	0.317	0.375
10. To what extent did you feel down because your swollen arm has restricted the clothing that you would like to wear?		0.664
11. How much of the time did your swollen arm stop you wearing clothing that you would like to wear?		0.869
12. How much of the time did your swollen arm cause you pain or discomfort?	0.348	0.327
13. How much of the time were you aware of your swollen arm?	0.403	
14. How much of the time did your swollen arm cause you to feel self-conscious or embarrassed in public?		0.870
Eigenvalue	5.5	2.0
Percentage of variance explained by component	39.3%	14.3%

Items	Item-total correlation
Physical well-being component (including Q12)	
1. To what extent did your swollen arm interfere with your normal work activities?	0.71
2. To what extent did your swollen arm interfere with your normal leisure activities?	0.68
3. To what extent did your swollen arm interfere with your self-care?	0.73
4. To what extent did your swollen arm interfere with dressing and undressing?	0.68
5. To what extent did your swollen arm interfere with your sleep?	0.53
7. To what extent di your swollen arm interfere with lifting or carrying light objects?	0.57
8. To what extent did your swollen arm interfere with reaching high objects?	0.77
12. How much of the time did your swollen arm cause you pain or discomfort?	0.46
13. How much of the time were you aware of your swollen arm?	0.45
Emotional well-being component (including Q9)	
6. To what extent did you feel your femininity or masculinity was influenced by your swollen arm?	0.60
9. To what extent did you feel down because of your swollen arm?	0.46
10. To what extent did you feel down because your swollen arm has restricted the clothing that you would like to wear?	0.60
11. How much of the time did your swollen arm stop you wearing clothing that you would like to wear?	0.60
14. How much of the time did your swollen arm cause you to feel self-conscious or embarrassed in public?	0.59

ULLQoL scores	Mean score (SD)	Range	Skewness
ULLQoL physical well-being component scores	13.5 (7.7)	0 to 30[a]	+ 0.391
ULLQoL emotional well-being component scores	5.2 (4.8)	0 to 19[b]	+ 0.921
ULLQoL total scores	18.7 (10.7)	0 to 49[c]	+ 0.504

[a] Maximum range 0 to 36; [b] Maximum range 0 to 20; [c] Maximum range 0 to 56.

Table 7. Predicted and observed correlations between ULLQoL scale components and existing measures

		ULLQoL Physical well-being component	ULLQoL Emotional well-being component	EQ-5D-3L Utility score	EuroQoL thermometer	SF-36 PCS
ULLQoL Physical well-being component	**Predicted**		+ve moderate	-ve moderate	-ve moderate	-ve moder
	Observed		+ 0.45 (weak)	- 0.59 (moderate)	- 0.44 (weak)	- 0.5 (moder
ULLQoL Emotional well-being component	**Predicted**			-ve weak	-ve weak	-ve very w
	Observed			- 0.50 (weak)	- 0.46 (weak)	- 0.3 (wea

ULLQoL: low score shows better QoL. SF-36: high score shows better QoL. EQ-5D-3L: high score shows better QoL.
'+ve' indicates the change in score was in the same direction; '-ve' indicates that the change in score was in different dire
Correlations greater than 0.7 show strong relationships; between 0.5 and 0.7 show moderate relationships; between 0.3 and less than 0.3 show very weak relationships [26]

(standard deviation)	(n=42)	(n=33)	(n=10)	
ULLQoL physical well-being component [a]	-8.9 (19.0)	-3.3 (13.9)	8.4 (13.8)	4.58**
ULLQoL emotional well-being component [a]	-5.4 (17.7)	2.8 (18.2)	15.0 (27.7)	5.02**
% change in limb volume	-4.8 (11.8)	-0.7 (5.7)	11.7 (14.6)	10.39**

a Lower ULLQ scores show better quality of life.
** Statistically significant at 1% level.

	Better (n = 42)		Worse (n = 10)	
	Mean change (SD of change scores)	**MSRM**	**Mean change (SD of change scores)**	**MSRM**
ULLQoL physical well-being component [a]	-8.9 (19.0)	0.64	8.4 (13.8)	0.61
ULLQoL emotional well-being component [a]	-5.4 (17.7)	0.30	15.0 (27.7)	0.83
% change in limb volume	-4.8 (11.8)	0.84	11.7 (14.6)	2.06

MSRM: Modified Standardised Response Mean

[a] Lower ULLQoL scores show better QoL

Key finding:

- The 14-item ULLQoL scale is a valid, reliable, consistent and responsive PROM for assessing quality of life in upper limb lymphedema.

What this study adds to what was known?

- Other lymphedema-specific measures exist, but when assessed against Terwee et al.'s (2007) quality criteria and the COSMIN checklist (Mokkink et al., 2010) for methodological quality none was adjudged robust.
- The ULLQoL scale was developed and validated rigorously, and is short, taking less than five minutes to complete, and is easy to use in clinical practice and research.

What is the implication and what should change now?

- The ULLQoL scale has potential for use in clinical practice and research to assess QoL alongside limb volume.
- We recommend including the ULLQoL scale in 'Best Practice for the Management of Lymphedema'