Conceptualisation, development and validation of T-QoL© (Teenagers' Quality of Life): a patient-focused measure to assess quality of life of adolescents with skin diseases

M. K. A. Basra[1,2], M. S. Salek[3], D. Fenech[2], A. Y. Finlay[2]

[1]Dermatology Department, Frimley Park Hospital, Frimley Health NHS Foundation Trust, Frimley, Surrey, UK

[2]Department of Dermatology and Wound Healing, Division of Infection and Immunity, Cardiff University School of Medicine, Cardiff, UK

[3] School of Life and Medical Sciences, University of Hertfordshire, Hatfield, UK.

Short title: Teenagers' Quality of Life questionnaire

Word count after revisions: 5734

Number of Figures: 5

Number of Tables: 4

Appendices: 1

Number of Supplementary Files: 7

Key words: adolescents, teenagers, quality of life, skin diseases, dermatology, young adults, impact, patient-reported outcome

ORCID numbers

A Y Finlay: 0000 0003 2143 1646

M S Salek: 0000 0002 4612 5699

Address for correspondence

Dr Mohammad K. A. Basra

Dermatology Department, Frimley Park Hospital,

Frimley, GU16 7UJ, Surrey, United Kingdom

Tel: +447915049726

E-mail: drkhurshid69@hotmail.com

Funding: No external funding for this study; the study was funded by Cardiff University Dermatology Department quality of life research fund.

Disclosures: MB, SS and AYF are joint copyright owners of T-QoL. AYF is joint copyright owner of the DLQI and CDLQI: Cardiff University and AYF receive royalties.

What's already known about this topic?

- Skin disease can affect quality of life (QoL) of teenagers in a variety of different ways, some being unique to this age group
- Dermatology-specific QoL measures exist for adults and children
- Skindex-Teen, has been the only (21-item) QoL measure for adolescents

What does this study add?

- T-QoL is a new age-specific measure to assess QoL of teenagers suffering from different skin diseases
- Items were generated using direct input from target population
- T-QoL was developed and psychometrically analysed using both traditional classical test theory and modern item response theory models

Abstract

Aim:

To develop and validate a dermatology-specific quality of life (QoL) instrument for adolescents with skin diseases.

Methods:

Qualitative semi-structured interviews were conducted with adolescents with skin disease to gain in-depth understanding of how skin diseases affect their QoL. A prototype instrument based on the themes identified from content analysis of interviews was tested in several stages, using Classical Test Theory (CTT) and Item Response Theory (IRT) models to develop this new tool and conduct its psychometric evaluation.

Results:

Thirty-three QoL issues were identified from semi-structured interviews with 50 adolescents. A questionnaire based on items derived from content analysis of interviews was subjected to Rasch analysis: factor analysis identified three domains, therefore not supporting the validity of T-QoL as a unidimensional measure. Psychometric evaluation of the final 18-item questionnaire was carried out in a cohort of 203 adolescents. Convergent validity was demonstrated by significant correlation with Skindex-Teen and CDLQI or DLQI. The T-QoL showed excellent internal consistency reliability: Cronbach's α=0.89 for total scale score and 0.85, 0.60, and 0.74 respectively for domains 1, 2 and 3. Test-retest reliability was high in stable subjects. T-QoL showed sensitivity to change in two sub-groups of patients who indicated change in their self-assessed disease severity.

Conclusion:

Built on rich qualitative data from patients, the T-QoL is a simple and valid tool to quantify the impact of skin disease on adolescents' QoL; it could be used as an outcome measure in both clinical practice and clinical research.

Introduction

There are several specific aspects of quality of life (QoL) affected by skin disease in adolescence, highlighting the importance of having an age-specific QoL tool.[1] The Dermatology Life Quality Index (DLQI),[2,3] designed for adults aged over16 years, contains some items irrelevant to younger adolescents. The Children's Dermatology Life Quality Index (CDLQI),[4,5] designed for children aged 4-16 years, is not focussed on specific teenage issues. The overlap of CDLQI and DLQI scores highlights the need for an adolescent specific tool.[6] Self-image, influence of peers and new friendships are of paramount importance to teenagers,[1,7] aspects under-represented in existing dermatology-specific tools. The impact of skin diseases is predominantly based on adolescents' subjective perceptions of their condition, rather than on objective clinical measures.[8] Other QoL measures focussed on the concerns of adolescents include acne specific measures, as acne primarily affects this age group.[9,10] However when measuring QoL in adolescence, researchers have generally used measures designed to be used across childhood.[11,12] In a systematic review of the measurement of QoL in adolescents with psoriasis, measures used were a general pediatric measure, PedsQL 4 or the CDLQI.[13]

Skindex-Teen,[14] has two domains: physical symptoms and psychosocial functioning. The items were derived from the pre-existing adult tool Skindex, literature review and feedback from experts. Unfortunately, there was no direct input from adolescent patients at the time of item generation, though the questions were later reviewed by a patient panel. The aims of this study were to use information directly from adolescents, to give a comprehensive insight into the impact of skin diseases on their QoL in order to develop and psychometrically validate an adolescent-specific QoL instrument.

Methods

This study, undertaken at the University Hospital of Wales, Cardiff from 2008-2012, was approved by Cardiff and Vale NHS Trust and South East Wales Local Research Ethics Committee. Written informed consent was obtained from participants aged 18-19 years, or from a parent (with assent) if aged <18 years. All information was kept confidential. Inclusion criteria included patients aged 12-19 years, able to understand and read English, suffering from a diagnosed skin disease, able to give assent or written informed consent and not suffering from concomitant illness affecting QoL. The exclusion criteria included patients with significant co-morbidities. In practice this meant diseases such as physical and mental

disabilities, malignancies, or severe chronic disease, such as severe rheumatoid arthritis. In all phases of this study, patients were recruited by consecutive sampling methodology, a convenience sampling approach.[15] The study was conducted in five stages: conceptualisation and development of a conceptual framework; qualitative interviews; item generation; scale refinement and item reduction; and psychometric evaluation of the final instrument (Figure 1).

Conceptual framework

The impacts of skin diseases on adolescents were identified[1,4,7,8] and selected for inclusion in the preliminary hypothesized conceptual framework. The primary impacts included impacts on functional behaviour (i.e. both physical and psychosocial), changing or modifying daily schedules (missing work or school), wearing makeup or clothes to conceal appearance concerns, difficulties concentrating, feeling angry or frustrated, impacts on relationships with family and friends, and difficulties playing sports. Figure 2 shows the hypothesized conceptual framework for quality of life impacts in adolescents with skin disease.

Qualitative interviews, item generation and comprehensiveness testing

Semi-structured interviews were conducted with teenage patients and transcribed verbatim. Patients were asked to describe ways their lives had been affected by their skin disease. This information was used to draft an instrument with 5-point response categories, “T-QoL”, which was pilot-tested in a new cohort of teenagers for cognitive debriefing and to test content and face validity. Qualitative feedback was given by study participants and the four experts concerning item relevance, language clarity, completeness and scaling.

The comprehensiveness of the tool was assessed qualitatively during the second stage of item generation while conducting the face and content validation of the draft questionnaire. Twenty adolescent study subjects were recruited who gave their opinion about the relevance, wording, and clarity of individual items, and "whether anything that they thought was relevant but was missing from the questionnaire". They were also asked whether there were any items that should be omitted or altered. Qualitative feedback was also sought from four dermatology staff members with experience in treating young adults with skin conditions concerning the format, content, relevance of questions, clarity, wording and response categories. These experts were also asked whether they thought any important aspects were not covered by the draft questionnaire. Items were removed, substituted and merged, resulting in the second T-QoL version.

Scale refinement and item reduction

The second T-QoL version was tested in a new patient cohort to refine the instrument by applying classical test theory (CTT) and item response theory (IRT) of Rasch modelling. Rasch analysis was carried out using RUMM 2030 software to assess overall model fit, item responses, individual item fit and differential item functioning (DIF). Factor analysis identified the factor structure underlying the instrument items. Further reduction of items and of response categories yielded the final T-QoL version (Appendix 1).

Psychometric evaluation of the final instrument

The psychometric properties of T-QoL were assessed including validity, reliability and sensitivity to change. Convergent and construct validity was assessed by asking subjects to complete the T-QoL and the DLQI (if 16-19 years), or CDLQI (if 12-16 years), Skindex-Teen and a global question (GQ) on self-assessed disease severity on 0-10 scale, with 0 indicating clear skin and 10 most severe disease. Spearman rank correlation was used to assess the correlation coefficient between T-QoL and the three questionnaires. Correlation coefficient represents an effect size that describes the strength of relationship between two variables and its value ranges from 0-1. These effect sizes are interpreted as follows: 0-0.19 = very weak; 0.2-0.39 = weak; 0.4-0.59 = moderate; 0.6-0.79 = strong; 0.80-1.0 = very strong correlation.

Cronbach's α was used to assess the internal consistency reliability. There are different reports with regard to acceptable level of Cronbach's α for measurement scales, ranging from 0.7-0.9.[16,17,18] A value of 0.7 has been recommended as acceptable.[16] On the other hand a value of α above 0.9 may indicate item redundancy. An intraclass correlation coefficient (ICC) value between 0.60 and 0.74 is considered good whereas a higher value than 0.75 constitutes excellent.[19] For any scale to be useful a minimum ICC of 0.6 is required.[20]

Construct validity of the instrument was assessed by testing a number of a priori hypotheses: that there will be high correlation between the T-QoL and an instrument with similar construct i.e. Skindex-Teen and moderate to high correlation with other dermatology-specific QoL instruments - the DLQI and CDLQI. We also hypothesised a low-moderate correlation between T-QoL and self-assessed disease severity assessed by GQ based on previous experience.[14,21,22]

To assess test-retest reliability, a sub-group of consecutive study participants were given a second set of questionnaires to complete at home 3-7 days after the first visit and return using a stamped reply envelope. This short time period was to ensure that the skin condition severity had not changed significantly but at the same time to minimise the chances of recall of previous responses. Participants received a telephoned reminder. Only those with unchanged GQ score or a difference of ± 1 were included in the analysis. Test-retest reliability or stability was tested using ICC for the overall T-QoL scale and three sub-scales. Assessment of T-QoL's sensitivity to change was based on testing the a priori hypothesis that "there will be change in patients' T-QoL scores following changes in their disease severity as assessed by the GQ". In order to test this, subjects were contacted 1-3 months following completion of stage I. Patients who participated in stages I and II were posted a pack with a stamped reply envelope. Only subjects with a GQ score that changed by ± 2 points were included in the analysis. Wilcoxon signed rank test was used to assess the difference in T-QoL scores on two occasions in two subject groups: improved and worsened. When testing sensitivity to change, what matters is the validity of the change scores, rather that the magnitude of the change scores, as a high magnitude of change gives little indication of the ability of an instrument to detect change over time. Although ES and SRM values are not appropriate measures for this matter, we report them in order that comparisons may be drawn with similar data reported in the description of other measures.

Responsiveness of the T-QOL was analysed by comparing T-QoL score changes across patient groups to DLQI and CDLQI score changes. Patients were divided into three groups (improved, no change, and worsened) according to score changes from visit 1 to visit 3, on the GQ score, DLQI and CDLQI.

A GQ score change of ≥ 2 or ≤ -2 represented worsened or improved, respectively, whilst scores between > -2 and < 2 represented no clinically relevant change i.e. no change. For the DLQI and CDLQI, a score change of ≥ 3 or ≤ -3 represented worsened or improved, respectively, whilst scores between > -3 and < 3 represented no change (i.e. clinically non-significant change). The cut-off for the DLQI is less than the minimal clinically important difference (MCID) of 4[23] and is similarly likely to be less than the MCID for the CDLQI.

Comparator instruments

Skindex-Teen

Skindex-Teen is a self-administered 21-item dermatology specific QoL instrument for teenagers with different skin diseases.[14] The questions ask the respondent about the impact of their skin condition on various aspects of QoL during the previous 4 weeks period. Each item is scored on a 5-point scale: Never, Rarely, Sometimes, Often and All the time. Scores of the individual items (0-4) are added to yield to a total score (0-84); the higher the score, the greater the QoL impairment.

Dermatology Life Quality Index (DLQI)

The DLQI is one of the most commonly used QoL instruments for adult patients (16 years and above) with skin disease.[2,3] It is a self-administered, user-friendly questionnaire with a mean completion time of around 2 minutes.[24] It consists of 10 questions concerning patient perception of the impact of skin disease on different aspects of patient's health related QoL over the last week. The items of the DLQI encompass aspects such as symptoms and feelings, daily activities, leisure, work or school, personal relationships and the side effects of treatment. Each item is scored on a 4-point scale: Not at all/Not relevant, A little, A lot and Very much. Scores of the individual items (0-3) are added to yield to a total score (0-30); higher scores mean greater impairment of patients' QoL.

Childrens Dermatology Life Quality index (CDLQI)

The CDLQI is an instrument specifically designed for children between 4 and 16 years of age, to measure the impact of skin disease on their QoL.[4,5] There are text and text plus cartoon[25] versions of the CDLQI. The questionnaire was designed to be completed by the child, with the help of an adult, preferably the child's parent. Each item is scored on a four point scale: Not at all, A little, Quite a lot, Very Much. Scores of the individual items (0-3) are added to give a total score of from 0 to 30; the higher the score, the greater the impairment of child QoL.

Statistical analysis

Statistical analysis was performed using the Statistical Package for the Social Sciences (SPSS, versions 15.0, 16.0, 17.0 and 22.0). Probability of type I error was set at 5% (i.e. $p<0.05$).

Results

In total 426 adolescent patients with skin diseases were recruited from the dermatology outpatient clinics, University Hospital of Wales (Table 1). Four expert dermatology staff members also took part.

1. Conceptual framework of a QoL measure for adolescents

Direct input from adolescent patients with skin diseases was gained through face to face qualitative interviews.[1] Semi-structured interviews were conducted with 50 subjects (M=17, F=33, mean age 16 years, range=12-18). Patients were asked: 'Can you tell me all the ways you can think of how your skin condition has affected your life?' Interviews were transcribed verbatim. Saturation point was reached at interview number 32, however, 18 further interviews were conducted. Twelve main themes were generated (Table 2). Various aspects of QoL identified by thematic analysis of interview scripts using the "grounded theory" formed the basis of the new instrument, building on the preliminary hypothesised conceptual framework (Figure 2) as described below.

2. Item generation

Content analysis of interview transcripts identified 32 items that formed the prototype instrument. This represented eight distinct QoL domains: psychological, social, physical, relationships, leisure, studies/jobs/career, support, and impact on daily activities. The items asked adolescents the impact of their skin condition on their current QoL. A 5-point (0-4) adjectival scale was chosen to score each item: Never = 0; Rarely = 1; Sometimes = 2; Often = 3; and Always = 4. In the prototype there was an additional scoring category "Not relevant". Items described by participants as "Important" are listed in Table 2. Most of these items appeared in the first 32-item draft questionnaire. During subsequent stages of item refinement, several items were removed because they did not meet the statistical criteria, even though from the face value perspective they appeared important and relevant to the conceptual model.

Cognitive debriefing, face and content validity of the prototype instrument were carried out with a new cohort of 20 adolescents (Table 1). All 20 participants (100%) responded 'Yes' to

being happy to complete the questionnaire, found it relevant and said they would be willing to complete such a measure every clinic visit. All 20 (100%) stated the questionnaire was not time consuming. Concerning content and face validity, 80% reported that the questionnaire was "easy to understand", 70% thought it the right length and no items were vague or repetitive.

Four dermatology staff members experienced in treating adolescents commented on the prototype instrument, using a study-specific questionnaire to rate it for format, language clarity, completeness, relevance and scaling. No matters of concern were raised. Following the feedback from the two cohorts, two items were removed and wording of a few items simplified to improve readability. This led to a 30-item second version of the T-QoL.

3. Scale refinement

The aim was to test the 30-item T-QoL version for further refinement, including test of unidimensionality, on a new cohort of adolescent patients using Rasch modelling.

In total 155 adolescents attending the dermatology department were invited to participate: one declined for personal reasons and one with other health issues was unsuitable. All patients, with 23 different skin conditions (Table 1), completed the T-QoL. All were evaluable and psychometric analysis was carried out. Figure 3 shows the distribution of T-QoL scores of individual items in 153 participants. The mean item scores were (Figure 3): self-consciousness (2.4); thinking a lot about the condition (2.4); need to cover up affected areas (2.2); feeling annoyed (2.1); relationships with family (0.3); lack of support from healthcare staff (0.3) and family/friends (0.2). Results of the factor analysis of the original 30-item questionnaire, with factor loadings, are given in Supplementary File 1.

Rasch analysis was carried out on the 30-item version T-QoL data from 153 adolescent patients using Rasch Unidimensional Measurement Model (RUMM) software version 2030, with data imported from SPSS version 15. Rasch analysis found significant multidimensionality and did not support the validity of the T-QoL as a unidimensional measure of QoL impairment (Chi Square Probability=0.0001). Factor analysis was carried out to assess the level of multidimensionality. The Scree plot (Figure 4) showed the "elbow" effect, seen when a scale is measuring more than one domain or factor. The initial Eigenvalues, obtained from the Scree plot, were compared with Random Eigenvalues. Three

of the observed Eigenvalues were greater than the generated Eigenvalues, implying the presence of three domains (Table 3). Each of the domains was then subjected to Rasch analysis individually. The finalised domain 1 was approved on the 3rd attempt of Rasch analysis, after the sequential removal of items 19, 7 and 17. These items were identified as mis-fitting the Rasch model, causing skewing of other items and preventing the domain being unidimensional. After the removal of these 3 items the item-reduced domain 1 proved to be unidimensional. When considered separately after the removal of the items 19, 7 and 17, all domains showed adequate fit to the model (Chi Square Probability=0.36, 0.08, 0.11 for domains 1, 2 and 3 respectively), good person separation (-0.846, -1.535, -2.084 for domains 1, 2 and 3 respectively), good internal consistency (α=0.91, 0.782, 0.813 for domains 1, 2 and 3 respectively) and no significant differential item bias for gender or age.

Inspection of the response options for the T-QoL revealed issues regarding response category ordering. Disordered thresholds indicated the failure of respondents to use the response categories in a manner consistent with the level of trait being measured. Disordering of thresholds occurred due to respondents having difficulty discriminating between 5 response options. Too many response options were available and the response option labelling caused confusion. Substantial disordering was identified for some items, e.g. items 5 and 11. The categories “rarely”, “sometimes” and “often” were therefore replaced by a single category “occasionally”, resulting in the 3-response scale “Never”, “Occasionally” and “Always”. On completion of Rasch modeling, seven items were removed from the 30-item T-QoL version based on low item-total correlation and location for person in order to yield a perfect fit. Further adjustments were made to resolve the tension between mathematical modeling and qualitative interview data: four items were removed and two items with similar themes merged. This resulted in the final 18-item T-QoL version (Appendix 1) scored using the scale: Never=0; Occasionally=1; Always=2. Scores of individual items are added to give a maximum score of 36; the higher the score the greater the QOl impact. Scores can also be calculated for the 3 sub-scales or domains: Self-image (items 1-8) with score range=0-16; Physical well-being and future aspirations (items 9-12, score range=0-8); Psychosocial impact and relationships (items 13-18, score range=0-12).

4. Psychometric testing of the final T-QoL

The final stage involved psychometric testing of the 18-item T-QoL to assess validity, reliability and sensitivity to change, primarily employing CTT techniques.

4a) T-QoL scores

A new cohort of 203 patients was recruited from the dermatology department outpatient clinics with the same inclusion criteria (Table 1). Mean item scores are shown in Figure 5. The items related to self-image such as "think a lot about skin", "need to cover up", and "self-consciousness" were the most highly scored items, whereas the lowest scored items were "effect on intimate relationships" and "relationships with friends". This distribution of item scores was also reflected in the mean domain scores with domain 1 **(**Self-image) having the highest mean score (mean score=7; range 0-16) followed by domain 3 (Psychosocial impact and relationships) (3.4, 0-12) and domain 2 (Physical well-being and future aspirations) (1.7, 0-8). Female subjects generally scored higher than male subjects in domains 1 (self image, $p<0.0001$) and 3 (psychosocial and relationships, $p=0.006$) as well as in overall mean score ($p=0.0001$) but not significantly differently in domain 2 (physical well-being and future aspirations). Mean completion time for T-QoL (n=156) was 78 seconds (SD=29.1) and for Skindex-Teen (n=156) was 129 seconds (SD=46.5, $p<0.0001$). Table 4 shows mean scores for the six most common skin conditions reported.

4b) Factor analysis

Once Rasch analysis had confirmed multidimensionality underlying the items of the T-QoL, we carried out exploratory factor analysis. Three factors had been identified from the initial factor analysis of the 30-item draft questionnaire. Further factor analysis was carried out on the 18-item final version: this confirmed the presence of three factors with eigen values >1.00, with the first factor accounting for 38.8% of total variance whereas the three factors together accounted for 55.2% total variance; the eigen values of these were 6.98, 1.67 and 1.29 respectively. Although all items in factor 1 clustered quite logically under the domain of "self-image", there were some complex factor loadings for the other two factors. This is reflected in the way these factors are labelled: factor 2 items cover "physical well-being and future aspiration" and factor 3 items cover "psychological impact and relationships"The 18-item T-QoL had high internal consistency reliability with Cronbach's α of 0.89, though lower for individual domain scores (domains 1, 2 and 3= 0.85, 0.60, 0.74 respectively).

4c) Construct validity

The construct validity of T-QoL was demonstrated by successfully proving the a priori hypotheses: the convergent validity was assessed by correlating the mean total T-QoL score with that of Skindex-Teen, DLQI (subjects aged 16-19 years), CDLQI (subjects <16) and self-assessed disease severity GQ score. Significant correlations were seen with all comparator scales, the highest with Skindex-Teen ($r=0.83$ $p<0.0001$), a scale of similar construct to T-QoL, followed by slightly lower but still high correlations with CDLQI ($r=0.75$, $p<0.0001$) and DLQI ($r=0.74$, $p<0.0001$). There was a moderate correlation with patients' self-assessed disease severity, assessed by GQ ($r=0.50$, $p<0.0001$).

4d) Reliability

For assessment of test-retest reliability, in total 61 (30% of 203) patients agreed to participate and completed the second set of questionnaires after a mean of 7.2 days (SD=6.5). Only T-QoL data is presented here. The value of ICC was highest for the total T-QoL mean score=0.91 (95% CI=0.87-0.94) followed by domain 1=0.9 (95% CI=0.86-0.94) domain 2=0.76 (95% CI=0.65-0.84) and domain 3=0.74 (95% CI=0.6-0.83).

4e) Sensitivity to change

For assessment of sensitivity to change, 41 completed questionnaires (20.2%) were returned (mean time interval=122.5 days, SD=81). Thirty had a GC score decrease, indicating improvement in their self-assessed disease severity, and 11 had an increase indicating deterioration. In the “improved” subgroup, there was significant change in the total T-QoL and each of the three domain scores (Supplementary File 2). In the individuals with GC score increase, there was significant change in the GQ score, and general increase in two out of three domains scores and total T-QoL scores, however, this change did not reach statistical significance (Supplementary File 2). There was an overall change in the T-QoL total scale score with a standardised response mean (SRM) of 0.38 and effect size (ES)=0.34, both indicating low-moderate effect size.

In the analyses anchored on GQ score change, the ‘improved’ and ‘worsened’ groups showed moderate to large change in comparison with the ‘no change’ group (Supplementary File 3). The magnitude of score change in the improved group was larger than in the worsened group

e.g. standard response mean (SRM) for T-QoL total score was -0.68 in improved group, -0.09 in the 'no change' group, and 0.33 in the 'worsened' group. Similar results were obtained from the other responsiveness metrics. Results based on DLQI and CDLQI as anchors are consistent with findings from GQ anchor. As expected, change in T-QoL scores showed strong correlation with score changes in related PROs i.e. DLQI (rho = 0.44), the CDLQI (rho = 0.66) and the SkindexT total (rho = 0.73) (Supplementary File 4). These findings supported the a priori hypothesis that T-QoL is sensitive to change. These findings supported the a priori hypothesis that T-QoL is sensitive to change.

Overall, the current results suggest that the T-QoL is capable of detecting meaningful changes in the patient's condition - whether worsening or improving. Expected moderate to strong correlations between the change scores on the T-QoL and that on other related PRO measures provide further support for the responsiveness of the T-QoL.

5. Dimensionality of T-QoL

5a) Confirmatory Factor Analysis model (Supplementary File 5)

A confirmatory factor analysis (CFA) model where the T-QoL items (18 items) were hypothesised to have a loading on one of the three domains was estimated (Supplementary File 6, standardised factor loadings). Residual variances for latent variable were fixed to 1. The results on the overall model fit were mixed. The Chi-square test, root mean square error of approximation (RMSEA), and Tucker Lewis Index (TLI) indicate a poor fit. On the other hand, standardised root mean square residual (SRMR) and the comparative fit index (CFI) suggest an acceptable fit. All factor loadings were significant: 0.49 to 0.78 for self-image, 0.4 to 0.71 for physical well-being and 0.42 to 0.71 for psychological impact and relationships. Residual variances for all items were positive. Covariance among the three domains was high, ranging from 0.62 to 0.94.

5b) Bi-factor model

A bi-factor model where the T-QoL items were hypothesised to have a loading on one of the three domains and an additional general factor (overall) was estimated (Supplementary File 7:

Standardised factor loadings). Similar assumptions as in the CFA analysis were made, except for an additional assumption of orthogonality i.e. the exogenous latent variables were assumed to be uncorrelated.

The results on the overall model fit were mixed. The Chi-square test, RMSEA, and TLI indicate a poor fit. On the other hand, SRMR and the CFI suggest an acceptable fit. Across all metrics, the bi-factor showed a better fit in comparison to the CFA model without a general factor. All factor loadings to the general factor as well as to the three domains were significant. The loadings ranged from 0.30 to 0.75 for the general factor, 0.1 to 0.51 for self-image, 0.16 to 0.74 for the physical well-being and - 0.52 to 0.37 for psychological impact and relationships. Two items (T-QoL questions 13 and 14) showed negative loadings. Residual variances for all items were positive.

5c) Comparison of model fit – CFA versus bi-factor model

Overall model fit statistics suggest a slightly better fit for the bi-factor model. This is confirmed by a chi-square test (chi-statistic diff. = 123.5, df = 15, $p < 0.001$). In the bi-factor model all item loadings to the general factor are greater than loadings to the individual domains. Overall, the results from the CFA and bi-factor model support that a total score can be calculated despite the multidensionality of T-QoL.

Discussion

The first part of this study was exploratory; qualitative enquiry was made to identify different aspects of adolescents' QoL affected by skin diseases. Many were distinct from adults and children and not appropriately captured by existing QoL tools for adults or children. For example, one of the major themes identified was related to pre-occupation with self-image. The second part of the study was the development of a dermatology-specific QoL instrument for adolescents, "T-QoL"©. The initial qualitative information formed the basis for conceptual framework of this age-specific instrument.

Both the traditional CTT and the modern IRT of Rasch modeling techniques were used in the development, refinement and psychometric evaluation of this instrument.[26] Rasch modeling is considered the most efficient means of establishing unidimensionality of a measure and

also for removing the items with weak psychometric performance. Application of Rasch modeling during the refinement stage of T-QoL resulted in removing some of the problematic items to yield a robust tool for easy administration and minimal respondent burden. Application of Rasch analysis to data in stage 3 (n=153) also demonstrated that T-QoL does not fulfil the criteria of unidimensionality and three domains were identified. The presence of three domains or sub-scales is further supported by results of principal component analysis in stage four. Further analysis using a bi-factor model supported the use of a total score. Furthermore, the variance explained by the first factor was more than 20%,[27,28] suggesting that a total summary score could be used. Ideally the three domains should be scored separately, but if someone wished to use a total summary score based on their specific need for research or clinical use, then we believe this could also be justified based on published criteria, as stated above. A similar provision to use a summary score for two well-identified domains (psychosocial functioning and physical symptoms) has also been given for Skindex-Teen.[14]

The T-QoL items enquire about current impact. This reduces respondent burden by avoiding need to recall, encourages greater accuracy and improves usefulness in clinical trials and routine practice. Having only three response options further reduces respondent burden. Overall, the psychometric properties are promising. There was high but not too high[29] internal consistency reliability (Cronbach's α) of the total scale i.e. 0.89. The values ranged from 0.6-0.84 for the three individual domains. The decreasing magnitude of α possibly reflects that Cronbach's α is related directly to the number of scale items.[16,26] Assessment of test-retest reliability demonstrated stability of the measure when the skin condition remained unchanged. Of the four comparator measures used to assess convergent validity, the strongest correlation was between T-QoL and Skindex-Teen (r=0.83): they have similar constructs and are aimed at the same population. The lowest but still significant correlation (r=0.5) was seen with self-assessed disease severity, GQ. In psoriasis there is stronger association between QoL and self-assessed disease severity compared to clinically assessed severity.[30] T-QoL seems to be sensitive to change both at the scale and domain level. The change was significant in the improved patients: it did not reach significance level in those with worsening of their disease severity, possibly related to the small sample size. Further larger sample size studies are required.

The only other age-specific QoL instrument available for adolescents with skin disease is the 21-item questionnaire Skindex-Teen.[14] The T-QoL and Skindex-Teen correlation coefficient value was higher than for the DLQI or the CDLQI, as expected given their similar construct. There were other similarities between the findings of the final validation stage of our study and that of Smidt et al.[14] The five commonest conditions in both were acne, atopic eczema, moles, psoriasis and warts with the highest mean QoL score (i.e. greatest impact) for eczema. Rasch analysis demonstrated both instruments to be multidimensional, with two domains for Skindex-Teen and three for T-QoL. The main difference between T-QoL and Skindex-Teen is the way that the items were generated. Skindex-Teen items were based on pre-existing Skindex questions designed originally for adults, a literature review and expert opinion.[14] T-QoL questions were all based on in-depth face-to-face interviews with the target adolescent population.[1] In-depth interviews with relevant subjects is a critical step in item generation and scale construction since these give valuable first hand insight.[31] This step is considered essential by the Federal Drugs Agency and European Medicines Agency in their PRO guidelines.[32] The other major difference is the much shorter mean completion time for T-QoL (78 seconds) compared to Skindex-Teen (129 seconds). Reasons may include the greater number of questions and response categories in Skindex-Teen (21 and 5) compared to T-QoL (18 and 3). Assessment length and ease of responding are main determinants of respondent burden.[31] The recall period can also affect the time taken to complete questions. For T-QoL, "current" time is used, more straightforward than having to think about the last four weeks as required by Skindex-Teen. Using this "current" time frame may hopefully make T-QoL more responsive to change over time.

Limitations of the study

The study participants included adolescents from a secondary-referral practice and so is not representative of the wider population, introducing selection bias. Patients were not recruited in primary care or from inpatients. Not all adolescents with skin disease are health-care seeking, these results may therefore not be generalizable to all adolescents with skin disease. Most participants were white Caucasians (87.6%), and there was under-representation of adolescents with different religious beliefs, cultural values and ethnic backgrounds. The sample sizes for Rasch measurement (n=153) and responsiveness (n=41) are slightly less than the recommended 10 subjects per item:[33] a larger sample size would have increased the reliability of the results. A small sample size may risk masking misfit of the data to the Rasch model. A larger data set would have ensured a more accurate person-item distribution,

enabling better understanding of how well targeted the T-QoL is for the patient population. In this Rasch analysis, only tests of DIF for gender and age were carried out. The subscales may produce bias between cultural groups and different types of skin disease. These will need to be further assessed for DIF.

We sought to recruit patients with both acute and chronic skin conditions in order to develop a tool applicable to patients with a wide range of conditions. However the mean duration of skin conditions was 5.4 years, indicating that the sample had a predominance of more chronic skin conditions.

A larger, more heterogeneous sample would have increased the representativeness value of our data and not all aspects of validation of T-QoL have yet been established:[34] There are several properties of QoL measurement instruments that should ideally be demonstrated to meet the highest standards of instrument development. Both et al[35] have defined these as: Validity (conceptual, construct and convergent), Interpretability (Norms, categorisation and minimal clinically important difference), Reliability (Internal consistency, retest reliability), Structure, Responsiveness, Item bias, cultural issues (Translation, cultural equivalence), Administrative burden and Alternative forms. At this stage in the development of T-QoL the following have been met: Validity, Reliability, Structure, Respondent burden, Administrative burden.

During the recruitment process no record was kept of the total number of patients attending the clinic or who were invited to take part but did not do so. There was no control group and so it is unclear whether the results are specific to skin disease or the issues can be generalised to adolescents with non-dermatological conditions. Nevertheless, disease specific questionnaires are not applicable to non-disease controls. In addition, response bias is another limitation that is shared by most questionnaire-based studies.

Conclusion

This study confirms that skin conditions have a great impact on adolescent QoL.[1,36] Based on first-hand information directly derived from the target population and within a conceptual framework, T-QoL is a compact, easy to understand and administer measure, with potential for clinical practice, where measurement has many possible benefits.[37] Although some of the items in T-QoL arguably could be applied to adult or children populations, many questions

have a clear adolescent focus and relevance, making it distinct from adult and children QoL measures and reflecting the specific aspects of the transitional life phase of young adults. Preliminary validation results are promising adding to the potential of T-QoL to allow reliable and accurate assessment of QoL of adolescents with a wide range of skin afflictions. Data from T-QoL may be able to influence treatment plans and care strategies for individuals and could be used in outcome research such as in clinical trials. The next challenge is determining the cultural and linguistic equivalence of the measure for use in wider global adolescent populations.

Acknowledgements

We wish to thank Dr P. Kamudoni and Professor Richard Kay for their statistical assistance and advice. We also wish to thank Dr. C. Golics, Z. Tanweer, A Howe, T, Mahbouba, S-J Burton, E. Russell, R. Shaw, and S. Taylor for their assistance with patient recruitment and the patients and dermatology staff members for their contributions to the development of T-QoL.

References

1. Golics CJ, Basra MKA, Finlay AY, Salek MS. Adolescents with Skin Disease Have Specific Quality of Life Issues. *Dermatol* 2009; **218**: 357-66.

2. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI) - a simple practical measure for routine clinical use. *Clin Exp Dermatol* 1994; **19**: 210-6.

3. Basra MKA, Fenech R, Gatt RM *et al*. The Dermatology Life Quality Index 1994–2007: a comprehensive review of validation data and clinical results. *Br J Dermatol* 2008; **159**: 997-1035.

4. Lewis-Jones MS, Finlay AY. The Children's Dermatology Life Quality Index (CDLQI): Initial validation and practical use. *Br J Dermatol* 1995; **132:** 942-9.

5. Salek MS, Jung S, Brincat-Ruffini LA *et al*. Clinical experience and psychometric properties of the Children's Dermatology Life Quality Index (CDLQI), 1995–2012. *Br J Dermatol* 2013; **169**: 734-759.

6. van Geel MJ, Maatkamp M, Oostveen AM *et al*. Comparison of the Dermatology Life Quality Index and the Children's Dermatology Life Quality Index in assessment of quality of life in patients with psoriasis aged 16–17 years. *Br J Dermatol* 2016; **174**: 152-7.

7. Frisen A. Measuring health-related quality of life in adolescence. *Acta Paediatr* 2007; **96**: 963-8.

8. Halvorsen JA, Braae Olesen A, Thoresen M *et al*. Comparison of Self-reported Skin Complaints with Objective Skin Signs Among Adolescents. *Acta Derm Venereol;* **88**: 573-7.

9. Ogedegbe EE, Henshaw EB. Severity and impact of acne vulgaris on the quality of life of adolescents in Nigeria. *Clin Cosmet Investig Dermatol* 2014; **7**: 329–34.

10. Tuong W, Wang AS, Armstrong AW. Effect of Automated Online Counseling on Clinical Outcomes and Quality of Life Among Adolescents With Acne Vulgaris: A Randomized Clinical Trial. *JAMA Dermatol* 2015; **151**: 970–5.

11. Chamlin SL, Chren M-M. Quality-of-life Outcomes and Measurement in Childhood Atopic Dermatitis. *Immunol Allergy Clin North Am* 2010; 30: 281–8.

12. Slattery MJ, Essex MJ, Paletz EM *et al.* Depression, Anxiety, and Dermatologic Quality of Life in Adolescents with Atopic Dermatitis. *J Allergy Clin Immunol* 2011; 128: 668–71.

13. Gonzalez J, Cunningham K, Perlmutter J, Gottlieb A. Systematic Review of Health-Related Quality of Life in Adolescents with Psoriasis. *Dermatology* 2016 Nov 4. [Epub ahead of print]

14. Smidt AC, Lai JS, Cella D *et al*. Development and Validation of Skindex-Teen, a Quality-of-Life Instrument for adolescents with skin diseases. *Arch Dermatol* 2010; **146**: 865-9.

15. Baker SE, Edwards R, Doidge M. How many qualitative interviews is enough? Expert voices and early career reflections on sampling and cases in qualitative research. National Centre for Review Methods, Economic and Social Research Council 2012.

16. Nunnally JC, Bernstein IH. *Psychometric theory.* Third edition. New York: McGraw-Hill Inc 1994.

17. Bland J, Altman D. Statistics notes: Cronbach's alpha. *BMJ* 1997: **314**: 275.

18. DeVellis R. *Scale development: theory and applications.* Thousand Oaks, California: Sage; 2003.

19. Cicchetti DV. "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology." *Psychological Assessment*. 1994; **6**: 284–90.

20. Chinn S. Repeatability and method comparison. *Thorax* 1991; **46**: 454-6.

21. Basra MKA, Sue-Ho R, Finlay AY. The Family Dermatology Life Quality Index: measuring the secondary impact of skin disease. *Br J Dermatol* 2007; **156**: 528-538. Erratum: *Br J Dermatol* 2007; **156**: 791.

22. Ben-Gashir MA, Seed PT, Hay RJ. Quality of life and disease severity are correlated in children with atopic dermatitis. *Br J Dermatol* 2004; **150:** 284-90.

23. Basra MK, Salek MS, Camilleri L *et al*. Determining the minimal clinically important difference and responsiveness of the Dermatology Life Quality Index (DLQI): further data. *Dermatology* 2015; **230**: 27-33.

24. Loo W-J, Diba V, Chawla M, Finlay AY. Dermatology Life Quality Index: influence of an illustrated version. *Br J Dermatol* 2003; **148**: 279-84.

25. Holme SA, Mann I, Sharpe JL *et al*. The Childrens' Dermatology Life Quality Index: validation of the cartoon version. *Br J Dermatol* 2003; **148**: 285-90.

26. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: a practical guide to their development and use*. Fifth Edition. Oxford University Press, Oxford 2014.

27. Reckase M. Unifactor latent trait models applied to multifactor tests: results and implications. *J Educ Stats* 1979; **4**: 207-30.

28. Hattie J. Methodology review: assessing unidimensionality of tests and items. *Appl Psycholog Methodol* 1984; **20**: 1-14.

29. Boyle, Gregory J. Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences* 1991; **12**: 291-294.

30. De Korte J, Bos JD, Sprangers MAG. Mombers FMC. Quality of Life in Patients with Psoriasis: A Systematic Literature Review. *J Invest Dermatol Symp Proceedings* 2004; **9**: 140-7.

31. Turner RR, Quittner AL, Parasuraman BM *et al.* Patient-reported outcomes: Instrument development and selection issues. *Value in Health* 2007; **10**: s86-s93.

32. Bottomley A, Jones D, Claassens L. Patient reported outcomes: Assessment and current perspectives of the guidelines of the Food and Drug Administration and the reflection paper of the European Medicines Agency. *European J Cancer* 2009; **45**: 347-53.

33. Linacre JM. Investigating rating scale category utility. *J Outcome Measurement* 1999; **3**: 103-22.

34. Heinl D, Prinsen CAA, Drucker AM *et al*. Measurement properties of quality of life measurement instruments for infants, children and adolescents with eczema: protocol for a systematic review. *Syst Rev* 2016; **5**: 25.

35. Both H, Essink-Bot, ML, Busschbach J, Nijsten T. Critical review of generic and dermatology-specific health-related quality of life instruments. *J Invest Dermatol* 2007; **127**: 2726-39.

36. Walker N, Lewis-Jones MS. Quality of life and acne in Scottish adolescent schoolchildren: use of the Children's Dermatology Life Quality Index (CDLQI) and the Cardiff Acne Disability Index (CADI). *J Eur Acad Dermatol Venereol* 2006; **20**: 45-50.

37. Finlay AY, Salek MS, Abeni D *et al.* Why quality of life measurement is important in dermatology clinical practice. An expert-based Opinion Statement by the EADV Task Force on Quality of Life. *J Eur Acad Dermatol Venereol* 2017 (in press).

Legends to Tables, Figures, Supplementary Files

Tables

Table 1. Demographic details of study participants.

Table 2. Quality of life aspects identified from interviews with 50 adolescents with skin diseases, showing 12 main themes.

Table 3. Items included in the three domains of the draft 30 item T-QoL.

Table 4. Mean T-QoL total and domain scores in the six most common skin conditions in the study group of 203 subjects.

Figures

Figure 1. Flow chart showing the developmental stages in the creation of T-QoL

Figure 2. Hypothesized conceptual framework for skin disease impacts on adolescents.

Figure 3. Mean item score during scale refinement field testing (range 0-4).

Figure 4. Scree plot showing results of factor analysis based on T-QoL data from 153 patients.

Figure 5. Mean T-QoL item score (range = 0-2) during validation stage, based on 203 patients.

Appendix

Appendix 1. The final 18 item T-QoL questionnaire.

Supplementary Files

Supplementary File 1. The factor loadings of the original 30-item draft T-QoL.

Supplementary File 2. Sensitivity to change of T-QoL from first to final administration (Wilcoxon signed rank test), n=41.

Supplementary File 3. Responsiveness of TQoL, anchored on GQ score change.

Supplementary File 4. Relationship of T-QoL overall and domain scores to DLQI, CDLQI and SkindexTeen, using Spearman correlation coefficient.

Supplementary File 5. Confirmatory factor analysis of T-QoL (final 18 item version): Model goodness of fit statistics.

Supplementary File 6. A confirmatory factor analysis (CFA) model where the T-QOL items (18 items) were hypothesised to have a loading on one of the 3 domains was estimated (standardised factor loadings). Residual variances for latent variable were fixed to 1.

Supplementary File 7. A bi-factor model where the T-QOL items were hypothesised to have a loading on one of the three domains and an additional general factor was estimated (standardised factor loadings).

Table 1: Demographic details of study participants

	S1-Conceptualisation	S2-Item generation	S3-Scale refinement	S4-Psychometric testing
n	50	20	153	203
Males (%)	34	40	45.8	56.7
Females (%)	66	60	54.2	43.3
Mean age (yrs)	16	16.5	16.5	16.2
Caucasians (%)	92	95	87.6	86.2
Students (%)	86	75	95.5	88.7
Living with family (%)	98	85	98	94
No. of skin diseases (n)	10	5	23	25
Mean duration of skin disease (yrs)	7.2	7.7	5.5	4.7

Table 2: Quality of life aspects identified from interviews with 50 adolescents with skin diseases, showing 12 main themes

Domains and frequency (n, %)		Dimensions and frequency (n, %)	
1. Social impact	37, 74%	• Avoid going out in public	16, 32%
		• Strangers	21, 42%
		• More withdrawn/isolated	3, 6%
		• Effect on friendships/relationships	19, 38%
2. Leisure activities/ Hobbies	37, 74%	• Avoid swimming	24, 48%
		• Use of communal facilities	2, 4%
		• Effect on other sporting activities	12, 24%
3. Holidays	7, 14%	• Affects holiday	7, 14%
4. Support from family and friends	21, 42%	• Support from family and friends	21, 42%
5. Impact on education, interviews & Employment	23, 46%	• Education, interviews and Employment	23, 46%
6. Impact on daily activities	26, 52%	• Washing and shaving	7, 14%
		• Disrupt sleep	13, 26%
7. Need to cover up the affected skin, e.g. make-up	12, 24%	• Applying make-up - taking extra time	12, 24%
8. Psychological impact	44, 88%	• Anger	13, 26%
		• Comments and staring from peers and public	
		• Self-conscious	
		• Low confidence	
		• Annoyed	14, 28%
		• Embarrassment	2, 4%
		• Spots, unclean	1, 2%

		• Feeling different	2, 4%
		• Condition always on mind	4, 8%
		• Mood changes	
		• Feeling down	12, 24%
		• Being judged	12, 24%
9. Physical impact- disease related	32, 64%	• Symptoms e.g. pain, discomfort, itching	32, 64%
10. Side-effects from treatment	17, 34%	• Dry lips, dry skin, skin irritation, painful skin	17, 34%
11. Compliance issues and burden on managing the disease	11, 22%	• Finding medicines hard to take • Continuous trips to clinic- time consuming • Applying/taking medication time consuming	11, 22%
12. Clothing	27, 54%	• Material	3, 6%
		• Style	21, 42%
		• Colour	1, 2%

Table 3. Items included in the three domains of the draft 30-item T-QoL .

Qn No.	Domain 1	Qn No.	Domain 2	Qn No.	Domain 3
1	Self-conscious	21	Affect studies/job	5	Annoyed
2	Upset	22	Concentration	8	Think about it a lot
3	Look different	23	Miss studies/job for treatment	10	Avoid new people
4	Lonely	24	Future career	12	Unfriendly comments
6	People stare at you	27	Pain/discomfort	13	Family relationships
7	Low confidence	28	Sleep	14	Friend relationships
9	Embarrassed	29	Treatment difficult	15	Intimate relationships
11	Uncomfortable with others			25	Family/friends lack of support
16	Stop going places			26	Healthcare lack of support
17	Clothes			30	Treatment side effects
18	Swimming				
19	Communal facilities				
20	Cover-up				

Table 4. Mean T-QoL total and domain scores in the six most common skin conditions in the study group of 203 subjects.

Diagnosis	N	Mean T-QoL scale score Range=0-36	Mean Domain 1 score Range=0-	Mean Domain 2 score Range=0-	Mean Domain 3 score Range=0-
Acne	105	11.6	6.8	1.4	3.5
Eczema	31	15.8	8.4	3.5	3.9
Psoriasis	8	13.25	8.4	1.5	3.4
Moles	13	7.3	4.4	0.8	2.2
Non-specific dermatitis	5	10.4	5.8	1.6	3.0
Warts	6	12.7	6.5	2.3	3.8