Research Techniques Made Simple: Developing and Validating QOL Outcome Measures for Skin Diseases

Hayley Braun[1], Howa Yeung[1,2] and Suephy C. Chen[1,2]

Skin conditions can significantly impact QOL. Dermatology QOL instruments may measure general skin-specific, disease-specific, or condition-specific QOL. Key components in the development of QOL instruments include (i) instrument and conceptual framework development, (ii) items and conceptual framework refinement, (iii) psychometric property testing, and (iv) clinical meaning and interpretation. First, a theoretical framework based on existing literature and subject experts (i.e., patients living with these conditions) is developed. By administering the pilot instrument to patients, the theoretical framework undergoes testing and revision to improve the instrument. At initial testing, construct validity and internal consistency can be assessed. At 72-hour follow-up, test–retest reliability can be determined. Responsiveness is determined by a follow-up test at a time point reasonable to allow for a detectable change in condition. The clinical meaning of the results must then be determined to allow clinical and statistical interpretation.

Journal of Investigative Dermatology (2020) **140,** 1888–1892; doi:10.1016/j.jid.2020.04.024

INTRODUCTION AND OVERVIEW OF QOL ASSESSMENTS IN DERMATOLOGY

Measuring health-related QOL outcomes is especially pertinent to dermatology because many aspects of dermatologic care aim to improve QOL. One challenging aspect of dermatologic care is that provider assessments may not reflect patient assessment of QOL impact from skin diseases. Clinician assessment of skin disease severity may not always correlate with the disease disruption to a patient's life (Chren, 2012). Instead, the burden of skin disease is best assessed through a combination of clinical assessment, QOL measures, and other patient-reported outcome measures such as patient-reported global assessment scores (Perez-Chada et al., 2019). Objective clinical assessment instruments, such as the PASI for psoriasis, and subjective QOL assessments, such as Skindex and ItchyQoL, are best measured using psychometrically validated instruments. This paper aims to identify the structure, development, and qualities of a sound dermatologic QOL assessment tool through the presentation of two established and validated tools, the Skindex and ItchyQoL.

There are many general health QOL instruments used in clinical trials and observational studies. Dermatology is unique in that multiple instruments exist to assess general skin-related QOL, skin disease–specific QOL, and symptom-specific QOL. For instance, Skindex assesses general skin-related QOL, the RosaQoL specifically assesses rosacea-related QOL, whereas ItchyQoL specifically assesses itch-related QOL. Assessing QOL on the organ system, disease, and symptom levels can provide important opportunities for dermatologic outcomes assessment. Researchers can select from an increasing number of QOL instruments in dermatology. If the goal is to compare QOL impact from different skin conditions across populations, then a general skin-related QOL instrument such as the Skindex would be appropriate. However, if the goal is to detect change across time or due to an intervention, disease- or condition-specific instruments may be more responsive than general instruments. No matter the goal, all QOL instruments should demonstrate sound psychometric properties. Using the Skindex and ItchyQoL instruments as examples, we will review the process of instrument development, including (i) instrument and conceptual framework development, (ii) refinement of items and conceptual framework, (iii) psychometric property testing, and (iv) clinical meaning and/or interpretation.

Instrument and conceptual framework development

Development of an instrument requires hypothesized constructs about how a skin disease can impact QOL and then items to assess those constructs. To organize these constructs hierarchically, a conceptual framework can be a useful aid. In constructing the framework for Skindex and ItchyQoL, items were elucidated through patient interviews in addition to a review of previous instruments and frameworks. An iterative instrument development process is presented in Figure 1.

Skindex. Skindex aims to measure the bother and the frequency of skin diseases on QOL and assess change in QOL over time (Chren, 2012). Prior instruments such as the Dermatology Life Quality Index and input from patients and clinicians were incorporated into a comprehensive conceptual framework of the impact of skin diseases on

[1]*Department of Dermatology, Emory University School of Medicine, Atlanta, Georgia, USA; and* [2]*Regional Telehealth Service, Veterans Affairs Southeast Network (VISN 7), Decatur, Georgia, USA*

Correspondence: Suephy C. Chen, Department of Dermatology, Emory University, 1525 Clifton Road, First Floor, Atlanta, Georgia 30322, USA. E-mail: schen2@emory.edu

Abbreviation: MCID, minimal clinically important difference

SUMMARY POINTS

- Dermatology is unique in that QOL instruments exist to assess general skin-related QOL, skin disease–related QOL, and symptom-specific QOL instruments.
- To fully utilize these instruments, they must undergo rigorous psychometric property testing, and this is an iterative process.
- There are key components in the development of QOL instruments that must be considered by the researcher planning to develop an instrument or select an instrument for research purposes: (i) instrument and conceptual framework development, (ii) items and conceptual framework refinement, (iii) psychometric property testing, and (iv) clinical meaning and interpretation.
- With the psychometric tools outlined at our disposal, investigators can now evaluate and choose QOL instruments with the most rigorous psychometric properties and conceptual framework.
- Future directions include to (i) improve on existing tools rather than develop new but redundant ones and (ii) refine to improve psychometric properties and interpretability while decreasing clinical administrative and respondent burden.

patients' QOL. The framework was split into two major subgroups of psychosocial and physical effects. Psychosocial effects included cognitive, social, and emotional aspects, with the emotional aspects including depression, fear, embarrassment, and anger. The physical effects included discomfort and limitations (Chren, 2012). At first, 65 questions were developed to address all aspects of the conceptual framework within the context of the past 4 weeks. These questions were pilot-tested among 46 patients with various skin diseases (Chren, 2012). Repetitive or ambiguous questions were removed, resulting in the 61-item Skindex instrument.

ItchyQoL. ItchyQoL followed a similar development process to Skindex (Desai et al., 2008). In-depth interviews were conducted to elicit a comprehensive understanding of the impact of pruritus on QOL. Pruritus-specific items were constructed related to the frequency of occurrence ranging from never to all the time. The authors hypothesized and confirmed through principal factor analyses that the following items fit into a conceptual framework similar to Skindex: symptoms, functional limitations, and emotional impact. Because pruritus-related symptoms were considered best captured within a shorter time period, items referred to pruritus-related QOL from the past 7 days only.

Items and conceptual framework refinement

Researchers must balance between developing a longer QOL instrument to elicit a comprehensive understanding of QOL impact and a shorter instrument for efficient administration. Longer instruments may lead to survey fatigue, whereas shorter instruments may miss important aspects of QOL. The refinement stage of the instrument development seeks to reach this delicate balance.

Skindex. To condense the 61-item Skindex, item performance criteria were set on the basis of specific domains within the conceptual framework that explained the response variability in items with good psychometric qualities (Chren, 2012). New items were then added to improve evaluative capacity and discrimination of the instrument. A shorter Skindex-29 with 29 items assessing three domains of symptoms, emotions, and functioning was developed (Chren, 2012).

Further refinement was performed in 2001 to reduce the response burden by creating a one-page instrument while also focusing on bother instead of frequency of experience (Chren, 2012). Item analysis was performed on 692 patient responses to Skindex-29 to identify questions with strong ceiling or floor effects, in which >50% of respondents reported never or always. These were compared with interviews in which patients were asked, "What about your skin condition bothers you most?" Items with strong ceiling or floor effects and those that were not brought up by patients during interviews were considered for removal or rephrasing, which resulted in the one-page Skindex-16 (Chren, 2012).

Psychometric property testing

Face validity. Face validity refers to whether the instrument appears to measure information of interest; in this case, the QOL from skin diseases. Querying patients who experience the disease and clinicians who care for patients with specific skin diseases can provide insight on the comprehensiveness of the instrument in gathering relevant information. When developing a QOL instrument, inclusion of patients with different levels of disease severity, gender, age, and ethnicity should be considered. Timing should also be taken into account. For instance, QOL impact from basal cell carcinoma likely differs between patients who are healing from a recent excision and patients who are living with excision scar years later. Patients with a variety of skin conditions were interviewed in the development of Skindex and ItchyQoL (Chren, 2012; Desai et al., 2008).

Internal consistency and test–retest reliability. Internal consistency measures how well items within an instrument domain correlate with one another. When assessing internal consistency, researchers should look for a Cronbach $\alpha > 0.7$. Test–retest reliability measures how well an instrument reproduces similar results through repeated administration over a short-time period in which the skin condition is not expected to have changed. A good measurement of test–retest reliability is a correlation of >0.7 between baseline and repeat testing.

Skindex-29 and ItchyQoL demonstrated good internal consistency with Cronbach's α 0.87–0.96 and 0.92, respectively. They also demonstrated high test–retest reliability over

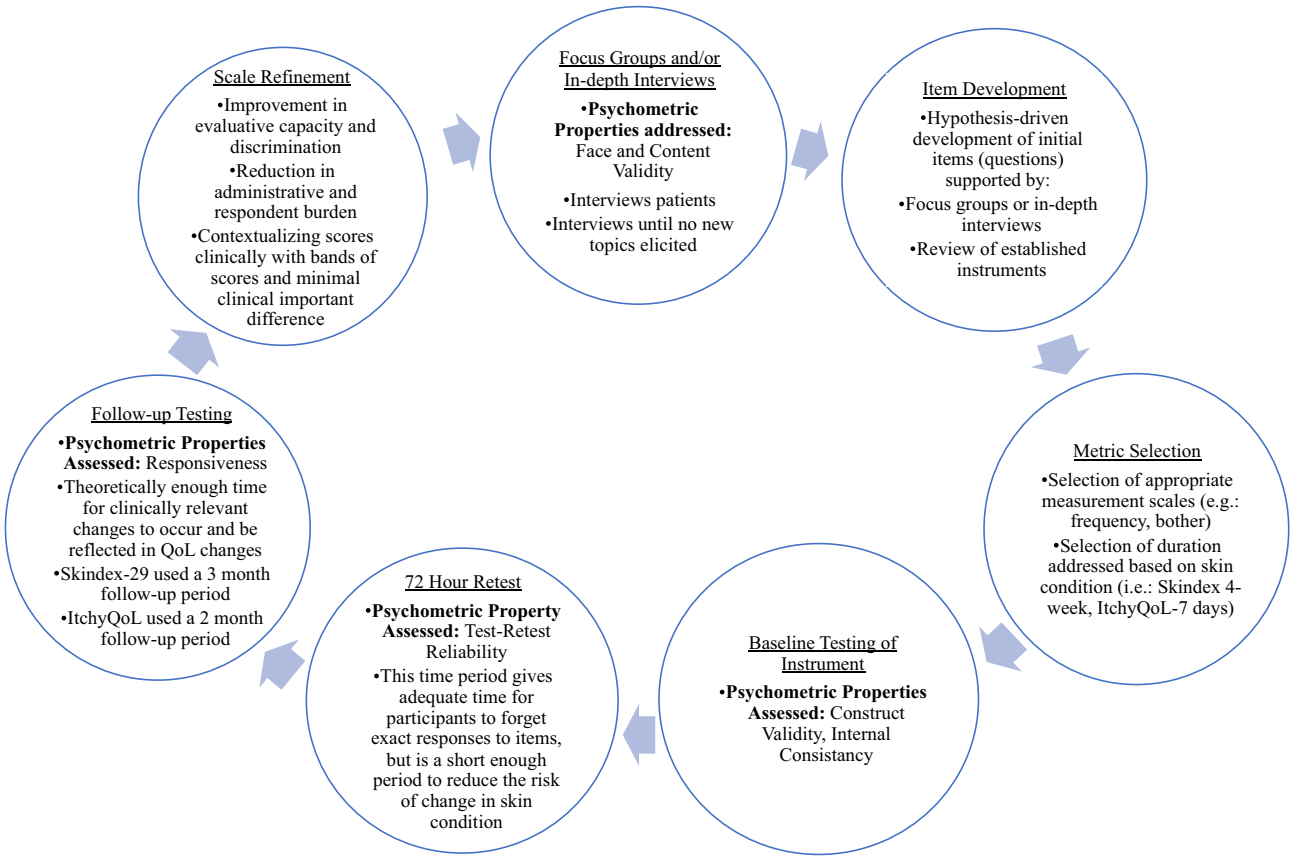

Figure 1. Development of a QOL instrument timeline.

72 hours: Skindex had Pearson's correlation coefficient of 0.88–0.92 and ItchyQoL had an intraclass correlation coefficient of 0.92 (Chren et al., 1997; Desai et al., 2008).

Construct validity. Construct validity assesses whether the instrument measures what it is intended to measure. When the constructs are not already operationally defined, the analyses determine whether the data fit the theorized factors within the proposed conceptual framework on QOL (Cronbach and Meehl, 1955). For both the Skindex and ItchyQoL, factor analysis was utilized to demonstrate how well the items fit the symptom, function, emotional domains that were proposed within the conceptual framework (Chren et al., 1997; Desai et al., 2008).

Responsiveness. Responsiveness measures the sensitivity of an instrument to detect changes over time. For Skindex, the changes in scores were determined by comparing with patient-reported changes in skin conditions. For those who reported a change in skin condition, a significant change in the expected direction on Skindex was seen (Chren et al., 1997). ItchyQoL assessed whether patients reported better, worse, or no change in itch. Patients with improved itch showed a significant change in ItchyQoL score; patients with worse or no change in itching did not demonstrate significant changes in ItchyQoL score potentially owing to small sample sizes (Desai et al., 2008).

Responsiveness is a critical property for outcome measures used in clinical trials to demonstrate treatment response. For example, in a randomized controlled trial comparing low-dose isotretinoin with placebo for the treatment of difficult-to-treat papulopustular rosacea, more patients in the isotretinoin group reached a 90% reduction in pustules and papules and reported improvement in Skindex scores (Sbidian et al., 2016). This finding suggests that the clinically measured reduction in papules and pustules correlated with QOL improvement and provides insight into the impact these interventions can have on patient's lives. Skindex and Dermatology Life Quality Index were also recently used to assess equivalency between online and in-person psoriasis care (Armstrong et al., 2019). Skindex and Dermatology Life Quality Index scores were calculated before intervention and after intervention and compared between the two groups; the authors found that the changes in QOL scores between the two groups fell within noninferiority margins. This utilization allowed researchers to recommend new paradigms of dermatologic care.

Clinical meaning and/or interpretation

After psychometric verification of QOL instruments, the next important step is contextualizing scores clinically by clinical banding or establishing minimal clinical important difference (MCID). To help understand the meaning of Skindex and ItchyQoL scores, bands of scores (e.g., 1–10, 11–20, etc.) representing clinical interpretations (e.g., mild, moderate, etc.) were developed. ItchyQoL has a statistically established banding system based on comparison of multiple banding options with correlations with the Global Itch Severity

MULTIPLE CHOICE QUESTIONS

1. A scale recommends banding of scores: 0–10 for mild, 11–20 for moderate, and 21–30 for severe QOL impact. In a randomized controlled trial, the scale is used to assess QOL improvement. The treatment group observes a change in QOL from a score of 19 to 12 ($P < 0.05$). How should these data be interpreted?
   A. Not clinically significant and not statistically significant
   B. Clinically significant but not statistically significant
   C. Not clinically significant but statistically significant
   D. Clinically significant and statistically significant
2. An acceptable value for Cronbach's α and test–retest correlation is
   A. 0.5
   B. 0.6
   C. 0.65
   D. 0.7
3. When a participant views an instrument as correctly measuring its stated aims, this is referred to as
   A. Discriminant validity
   B. Construct validity
   C. Face validity
   D. Internal consistency
4. A new QOL scale is developed, and the researchers want to ensure the test has good test–retest reliability. Which testing time period should be used to assess for test–retest reliability?
   A. Baseline + 1 hour after
   B. Baseline + 72 hours
   C. Baseline + 2 months
   D. Baseline + 3 months
5. Why are patient-reported outcomes important to the field of dermatology?
   A. Patient-reported outcomes should drive clinical management because patients' report should drive clinical management
   B. Patient-reported outcomes are not important because clinically detectable skin disease should drive clinical management
   C. Patient-reported outcomes, in conjunction with clinical skin disease severity, can be used to best holistically address the patient's skin condition and its effect on the patient's life
   D. Patient-reported outcomes should drive clinical management because it will result in better physician reviews

Question (0–30: little, 31–50: mild, 51–80: moderate, 81–110: severe) (Love et al., 2015). Several bandings variations have been proposed for Skindex (Rogers et al., 2012). If clinical bandings have been developed, future researchers should critically evaluate different banding options.

The bands allow differentiating clinical significance from statistical significance. For instance, a score difference of two on the ItchyQoL in a large-scale study may demonstrate statistically significant differences with a P-value < 0.05. However, if the mean ItchyQoL score changes from 10 to 8, both of these scores represent a mild impact on QOL and may not represent a clinically meaningful difference for the patient. In the isotretinoin group for the rosacea study described above, the authors looked at the change in Skindex score (Sbidian et al., 2016). Future research can consider evaluating both raw score differences and ordinal band changes to assess whether patients had transitioned from one clinically significant band to another.

The MCID identifies the smallest difference in QOL score that is meaningful to patients. The MCID has been identified as 10 points for the Skindex (Chren et al., 2007). An MCID has not yet been developed for ItchyQoL, but researchers have recommended approximately half an SD as an MCID (Norman et al., 2003).

Future directions

Researchers and clinicians should appraise the quality and appropriateness of various QOL instruments when deploying them in research and practice. QOL tools for many skin conditions and diseases exist for researchers to adopt or refine. Future work should shift from development to refinement through these iterative processes to improve psychometric properties and interpretability while decreasing clinical administrative and respondent burden.

ORCIDs

Hayley Braun: http://orcid.org/0000-0001-5485-0615
Howa Yeung: http://orcid.org/0000-0002-4815-4936
Suephy C. Chen: http://orcid.org/0000-0002-0678-7380

CONFLICT OF INTEREST

HB states no conflict of interest. HY received an honorarium from Syneos Health. SCC receives royalties from for-profit users of the ItchyQoL.

AUTHOR CONTRIBUTIONS

Conceptualization: HY, SCC; Writing - Original Draft Preparation: HB, SCC; Writing - Review and Editing: HB, HY, SCC

SUPPLEMENTARY MATERIAL

Supplementary material is linked to this paper. Teaching slides are available as supplementary material.

REFERENCES

Armstrong AW, Ford AR, Chambers CJ, Maverakis E, Dunnick CA, Chren MM, et al. Online Care versus in-person care for improving quality of life in psoriasis: a randomized controlled equivalency trial. J Invest Dermatol 2019;139:1037–44.

Chren MM. The Skindex instruments to measure the effects of skin disease on quality of life. Dermatol Clin 2012;30:231–6, xiii.

Chren MM, Lasek RJ, Flocke SA, Zyzanski SJ. Improved discriminative and evaluative capability of a refined version of Skindex, a quality-of-life instrument for patients with skin diseases. Arch Dermatol 1997;133:1433–40.

Chren MM, Sahay AP, Bertenthal DS, Sen S, Landefeld CS. Quality-of-life outcomes of treatments for cutaneous basal cell carcinoma and squamous cell carcinoma. J Invest Dermatol 2007;127:1351–7.

Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull 1955;52:281–302.

Desai NS, Poindexter GB, Monthrope YM, Bendeck SE, Swerlick RA, Chen SC. A pilot quality-of-life instrument for pruritus. J Am Acad Dermatol 2008;59:234–44.

Love EM, Marrazzo GA, Kini S, Veledar E, Chen SC. ItchyQoL bands: pilot clinical interpretation of scores. Acta Derm Venereol 2015;95:114–5.

Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003;41:582–92.

Perez-Chada L, Taliercio VL, Gottlieb A, Van Beek M, Duffin KC, Reeder M, et al. Achieving consensus on patient-reported outcome measures in clinical practice for inflammatory skin disorders. J Am Acad Dermatol 2019. https://doi.org/10.1016/j.jaad.2019.09.008 (Accessed 18 May 2020).

Rogers A, DeLong LK, Chen SC. Clinical meaning in skin-specific quality of life instruments: a comparison of the Dermatology Life Quality Index and Skindex banding systems. Dermatol Clin 2012;30:333–42, x.

Sbidian E, Vicaut É, Chidiack H, Anselin E, Cribier B, Dréno B, et al. A randomized-controlled trial of oral low-dose isotretinoin for difficult-to-treat papulopustular rosacea. J Invest Dermatol 2016;136:1124–9.

DETAILED ANSWERS

1. **A scale recommends banding of scores: 0−10 for mild, 11−20 for moderate, and 21−30 for severe QOL impact. In a randomized controlled trial, the scale is used to assess QOL improvement. The treatment group observes a change in QOL from a score of 19 to 12 ($P < 0.05$). How should these data be interpreted?**

CORRECT ANSWER: C. Not clinically significant but statistically significant

The results are statistically significant but not clinically significant because the treatment group had moderate symptoms before treatment and after treatment. Clinical bands allow researchers to categorize levels of severity in the clinically relevant categories. It is important to use these to understand the clinical interpretation of data and not just rely on statistical significance to drive conclusions.

2. **An acceptable value for Cronbach's α and test−retest correlation is**

CORRECT ANSWER: D. 0.7

Internal consistency measures the correlation of items within a scale with one another. When assessing internal consistency, researchers should look for a Cronbach's $\alpha > 0.7$.

3. **When a participant views an instrument as correctly measuring its stated aims, this is referred to as**

CORRECT ANSWER: C. Face validity

Face validity refers to the validity perceived by the participants or the participants' perception of how well the test measures its stated aims.

4. **A new QOL scale is developed, and the researchers want to ensure the test has good test−retest reliability. Which testing time period should be used to assess for test−retest reliability?**

CORRECT ANSWER: B. Baseline + 72 hours

72 hours had been determined to be enough time to forget responses to questions but a short enough duration to not have any changes in skin QOL occur. One hour after a test is initially given would be too short because the participants would likely remember all the responses. Months after would be too long because patients are likely to have changes in their skin conditions over this time.

5. **Why are patient-reported outcomes important to the field of dermatology?**

CORRECT ANSWER: C. Patient-reported outcomes, in conjunction with clinical skin disease severity, can be used to best holistically address the patient's skin condition and its effect on the patient's life.

Both clinical assessment and evaluation of QOL are important factors in clinical decision making. This is especially true in the field of dermatology where clinical severity may not directly correlate with impact on QOL.