Development, Validation, and Interpretation of the PROMIS Itch Questionnaire: A Patient-Reported Outcome Measure for the Quality of Life Impact of Itch

Jonathan I. Silverberg[1,2,3,4], **Jin-Shei Lai**[3,5], **Robert W. Kantor**[6], **Prarthana Dalal**[1], **Catherine Hickey**[1], **Sara Shaunfield**[3], **Karen Kaiser**[3], **Helena Correia**[3], **David Cella**[2,3,7,8]

[1]Department of Dermatology, Northwestern University Feinberg School of Medicine, Chicago, Illinois

[2]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois

[3]Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, Illinois

[4]Northwestern Medicine Multidisciplinary Eczema Center, Chicago, Illinois

[5]Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, Illinois

[6]State University of New York Downstate Medical Center, Brooklyn, New York

[7]Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, Illinois

[8]Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, Illinois

Abstract

Current patient-reported outcome measures for itch are limited and may not capture its full impact on health-related quality of life. We sought to develop, calibrate, and validate banks of questions assessing the health-related quality of life impact of itch as part of the Patient-Reported Outcomes Measurement Information System. A systematic process of literature review, content-expert review, qualitative research, testing in a sample of 600 adults, classical test theory methods, and item response theory analyses were applied. Exploratory and confirmatory factor analyses were followed by item response theory model and item fit analyses. Four itch-related item banks were developed: (i) general concerns, (ii) mood and sleep, (iii) clothing and physical activity,

Correspondence: Jonathan I. Silverberg, George Washington University School of Medicine, Suite 2B-425, 2150 Pennsylvania Avenue NW, Washington, DC 20037. JonathanISilverberg@gmail.com.

and (iv) scratching behavior. Item response theory and expert content review narrowed the item banks to 25, 18, 15, and 5 items, respectively. Validity of the item banks was supported by good convergent and discriminant validity with itch intensity, internal consistency, and no significant floor or ceiling effects. In conclusion, the Patient-Reported Outcomes Measurement Information System Itch Questionnaire banks have excellent measurement properties and efficiently and comprehensively assess the burden of itch.

INTRODUCTION

Itch, or pruritus, is a burdensome symptom with many etiologies and triggers, including inflammatory skin, systemic and neurological disorders, medication side-effects, and burns. Itch is reported in 0.8% of all outpatient visits in the United States (Shive et al., 2013). Itch is associated with significant impairment of health-related quality of life (HRQoL), including ability to perform every day physical activities, difficulties with social activities and relationships, and higher rates of depressed mood and anhedonia (Silverberg et al., 2016). Developing comprehensive, efficient, and valid tools to assess the burden of itch in clinical care and research is therefore a high priority.

Itch can be evaluated in several ways. Objective assessments of scratching behavior include nocturnal accelerometry or actigraphy (Petersen et al., 2013) and visual examination of scratching severity on the skin (Udkoff and Silverberg, 2018). Patient-reported outcome measures are more practical and widely used to quantify itch severity and are essential for characterizing patient burden (Schoch et al., 2017). The burden of itch is quite complex, with varying effects of different biological variables, baseline characteristics and functional status, and general health perceptions (Silverberg et al., 2018). A wide variety of published itch instruments address different research and clinical goals. Many instruments assess temporal and quantitative aspects of itch intensity and frequency. However, few instruments characterize its multidimensional impact on HRQoL (Schoch et al., 2017).

We developed a conceptual model of itch based on a systematic literature review (Kantor et al., 2016), panel experts, and semistructured interviews with 33 patients to characterize itch-relevant HRQoL domains and develop items to measure these domains (Silverberg et al., 2018). The model included the following five primary components: biological variables, symptom status, functional status, general health perceptions, and HRQoL. The model proposed a causal relationship beginning with the biological factors, with direct and indirect impacts of itch and its sequelae, including sleep disturbance. In turn, these can impair function, social life, relationships, emotional well-being, general health perception, and HRQoL (Silverberg et al., 2018). The conceptual model identified unmet needs in the evaluation and management of itch. Patient-reported outcome measures such as ItchyQOL and Skindex may not capture the full extent of the patient burden from itch (Silverberg et al., 2018) and suffer from limited or undocumented content validity, structural validity, or cross-cultural validity (Schoch et al., 2017).

The Patient-Reported Outcomes Measurement Information System (PROMIS) Itch Questionnaire (PIQ) item banks were developed to improve assessment of HRQoL in itch. PROMIS (www.healthmeasures.net) is a National Institutes of Health–funded

consortium that aims to build item pools and develop core questionnaires that measure key health-outcome domains manifested in a variety of chronic diseases. PROMIS also provides an electronic Web-based resource for administering computerized adaptive tests, collecting self-report data, and providing instant reports of the health assessments (www.assessmentcenter.net). PROMIS item banks are developed through a systematic process of literature review, expert consensus, qualitative research, classical test theory methods, and item response theory (IRT) analyses. These methods are designed to calibrate individual items for high precision and minimal bias across major symptom domains affecting health status and HRQoL.

In the setting of the broader PROMIS objectives, we sought to complete the first two stages of the PROMIS Instrument Maturity Model (Conceptualization & Item Pool Development and Calibration Phase; PROMIS, 2013). The specific aims of this project were the following:

1. Develop an archive of self-report measures that assess the HRQoL impact of itch using items from the conceptual model;

2. Develop item banks from these measures for use in adults with chronic itch regardless of the etiology;

3. Test the item banks in a large sample of adults with chronic itch and evaluate the psychometric properties of individual items using IRT models; and

4. Examine the validity of the new item banks.

RESULTS

Subject characteristics

Overall, 600 adults were included in the study; 550 were excluded from the study either for not having chronic itch or based on sampling quotas for severity of impact from itch. The final cohort was 62% female, 79% Caucasian/white, 9% African American/black, 7% Hispanic/Latino, 2% Asian, and 3% multiracial/other, with a distribution of 9%, 19%, 31%, 24%, 15%, and 2% for ages 18–24 years, 25–34 years, 35–44 years, 45–54 years, 55–64 years, and 65+ years, respectively.

Development of item banks

IRT assumptions.—Item descriptive statistics are presented in Supplementary Table S1. Exploratory factor analysis (EFA) results indicate one dominant factor and three potential factors among these 75 items (Figure 1). Unidimensionality of each item bank was evaluated separately. The decision to have four factors was made based on the EFA results as well as clinical judgement of the panel experts. The first factor accounted for 9.53% of the variance; factors 2, 3, and 4 accounted for 5.04%, 3.41%, and 1.62% of the variance, respectively. The first eigenvalue was 51.3 and the second, third, and fourth were 3.6, 2.0, and 1.2, respectively.

Confirmatory factor analysis (CFA) supported four separate unidimensional sets of items with acceptable fit indices (factor 1: 28 items; factor 2: 21 items; factor 3: 20 items; and

factor 4: 5 items). Factor 1 had items related to itch interference with multiple general concerns. Factor 2 had items related to itch interference with mood and sleep. Factor 3 had items related to itch interference with clothing choices and physical activity. Factor 4 had items related to scratching behavior.

The comparative fit index and non-normed fit index/Tucker-Lewis index values were high (>0.94) for the items in all four factors (Table 1). The root mean square error of approximation for factor 4 was 0.014, which is below the published criterion of <0.06. However, the root mean square error of approximation values for factors 1, 2, and 3 were 0.09, 0.144, and 0.137. This finding is likely because of the larger number of items in these factors. Similarly, a four-factor CFA model had high comparative fit index and non-normed fit index/Tucker-Lewis index values (>0.94) and a root mean square error of approximation of 0.06, supporting the existence of four different domains.

Item-to-item residual correlations indicated overall very little local dependence. All items in factor 1 and factor 4 had residual correlations <0.15 and were included in IRT modeling. Initial analysis of factor 2 showed one residual correlation >0.2 (question [Q] 24 vs. Q28). Similar fit indices were found between analyses with versus without Q24 and thus excluded from IRT modeling of factor 2. Initial analysis of factor 3 showed only two possible pairs of items with residuals >0.15 (Q16 vs. Q17, residual = 0.155; Q16 vs. Q18, residual = 0.161). Each of these items asks about itch interference on clothing choices. Based on these results and IRT results described hereafter, these items were excluded from factor 3.

Monotonicity was qualitatively evaluated by reviewing item characteristics curves, in which response category curves within each item were distinct from each other.

IRT models.—Initial analysis of factor 1 showed seven misfit items based on significant $S-X^2$ value ($P < 0.01$). Of these seven items, two were related to cognition, four were related to social functioning, and one was about work. Different combinations were tested, and all items fit after excluding the three items with smallest P-values from the initial analysis (Q58, Q59, and Q61). Initial analysis of factor 2 showed three misfit items (Q29, Q51, and Q52) based on significant $S-X^2$ value ($P < 0.01$), including one emotion- and two sleep-related items. Different combinations were tested, and all items fit after two of the three misfit items from the initial analysis (Q29 and Q52) were excluded. Initial analysis of factor 3 found that there was a potential secondary factor among Q16, Q17, and Q18, all of which were related to itch interference on clothing choices. Based on CFA, the best results occurred when all three items were excluded. However, additional analysis showed that there were still four misfit items (Q4, Q7, Q14, and Q19) based on significant $S-X^2$ value ($P < 0.01$). Of these, two were related to itch interference with physical activity and two were related to itch interference with clothing choices. Different combinations were tested, and all items fit after two of the four misfit items from the initial analysis (Q15 and Q20) were excluded. Analysis of factor 4 found no misfit items. IRT item parameters are presented in Supplementary Table S2.

Unidimensionality of the remaining items was supported. The final item bank 1 (itch interference with general concerns) has 25 items, item bank 2 (itch interference with mood

and sleep) has 18 items, item bank 3 (itch interference with clothing and physical activity) has 15 items, and item bank 4 (scratching behavior) has 5 items (Table 2).

Item characteristic curves for each item bank are presented in Supplementary Figure S1, Supplementary Figure S2, Supplementary Figure S3, Supplementary Figure S4. The item categories worked well for all items in item banks 1, 2, and 4, in that there was some point along the trait continuum when each category had the highest probability of being chosen. One item in factor 3 (Q19) showed response category 2 was problematic (this curve was hidden by other response category curves). Item-bank reliability and information functions are presented in the Supplementary Results section.

T-scores

Scores are reported on a T-score metric and range from 32.7–79.1 for item bank 1, 26.9–77.4 for bank 2, 32.6–77.1 for bank 3, and 32.6–72.7 for bank 4. A T-score of 50 is the average score for people in the United States general population who recently experienced chronic itch for any reason.

Psychometric properties

Internal consistency.—Within each factor, correlations among the included items were positive (range of Pearson r for bank 1: 0.60–0.84; bank 2: 0.50–0.84; bank 3: 0.54–0.83; and bank 4: 0.58–0.77) and statistically significant ($P < 0.0001$ for all) (Supplementary Tables S3–S6). Cronbach's alpha was 0.98, 0.97, 0.97, and 0.91, and item-total correlations ranged from 0.77–0.88, 0.75–0.84, 0.70–0.89, and 0.72–0.85 for factors 1, 2, 3, and 4, respectively.

Convergent validity.—T-scores for item banks 1, 2, 3, and 4 had strong correlations with each other and moderate correlations with the numerical rating scale (NRS) for worst or average itch (Pearson correlation) (Table 3). Each item was correlated with both average and worst intensity of itch (Pearson correlation).

Discriminant validity.—Regarding discriminant validity, there were significant and stepwise (monotonic) increases of T-scores for item banks 1, 2, 3, and 4 with each level of severity for NRS for worst or average itch (analysis of variance, $P < 0.0001$ for all) (Supplementary Figure S5). In addition, all items satisfied the monotonicity assumption in that incremental increase in the severity of response options was associated with a stepwise and statistically significant increase in the mean intensity of itch (analysis of variance, $P < 0.00001$ for all items).

T-scores for item banks 1, 2, 3, and 4 also showed good discriminant validity in predicting verbal rating scale for worst or average itch (Table 4). The areas under the curve for item banks 1, 2, 3, and 4 were good to excellent for distinguishing severe versus mild itch and very severe versus mild, moderate, or severe itch, and poor to fair for distinguishing moderate versus mild or severe versus moderate itch, indicating that they were able to distinguish between itch severity groups.

Floor or ceiling effects.—The proportions of respondents with lowest and highest values for item banks 1 (5.7% or 1.5%), 2 (2.0% or 2.2%), 3 (6.5% or 2.3%), and 4 (5.0% or 4.8%) were below 15%, indicating there were no floor or ceiling effects.

Interpretation

Severity thresholds were selected based on maximizing the concordance probability in receiver operating characteristic analysis (Table 5). For all four item banks, the optimal threshold for burden of moderate itch was a T-score of approximately 50. The optimal threshold for burden of severe itch was a T-score of approximately 57 (banks 1 and 2) or 58 (banks 3 and 4), whereas the optimal threshold for burden of very severe itch was a T-score of approximately 63 (banks 1 and 4) or 65 (banks 2 and 3). Different thresholds were identified for burden of mild itch for each item bank.

There was considerable overlap of severe T-scores across multiple item banks for respondents. However, 28.8% and 18.8% had severe T-scores in only one or two banks, respectively; only 36.9% had severe scores for all four banks.

DISCUSSION

The PIQ item banks were developed as measurements of HRQoL impact from itch through a systematic process of literature reviews, content-expert review, qualitative research, testing, and psychometric testing in 600 individuals. This process narrowed an initial list of 75 items to four banks with 63 items, representing overall itch interference on (i) general concerns, (ii) mood and sleep, (iii) clothing and physical activity, and (iv) scratching behavior. Classical test theory assessments including EFA and CFA provided support for the four item banks. The four final item banks demonstrated unidimensionality and item-fit to the IRT model. The results indicate that respondents can have severe scores in one bank independent of one or more other banks. PIQ item banks can be administered individually or in combination with each other when assessing patients with chronic itch. PIQ item banks showed good internal consistency and good convergent and discriminant validity, with no floor or ceiling effects. These results indicate that the PIQ constitutes psychometrically sound item banks for assessing the negative effects of itch on functioning in the range experienced by the vast majority of people who have itch. The final items adequately represent the burden of itch from a clinical perspective, that is, they have good face validity. Taken together, these findings support the validity of the PIQ item banks. Future studies are needed to further test for the reliability and validity of PIQ item banks in multiple diverse populations and develop more mature instruments.

The methodology used to derive the PIQ item banks differed from previously published itch instruments. First, items were developed using the PROMIS methodology, a rigorous and iterative process of collecting, sorting, and standardizing items based on literature reviews, content-expert review, and qualitative research. A conceptual model was used to develop items in the banks. Second, the samples for testing and initial psychometric analysis of the scales had adequate sample sizes to conduct EFA and CFA. Finally, the use of IRT methods to characterize the test performance of individual items lends additional strength to our methods. Scales developed with IRT have several desirable attributes, including

multiple and customizable approaches for administration. Item selection can be customized to specific applications. Because IRT analyzes the measurement properties of each item, precise estimates of the burden of each individual from itch can be obtained by selecting a smaller number of items from the item banks. Computerized adaptive testing is also possible, which uses individual item measurement properties to develop a progressively more precise estimate of an individual's burden of itch. Computerized adaptive testing allows the administration of fewer overall items, typically five to eight.

The PIQ has several distinct advantages over other instruments that assess HRQoL impact of itch. First, the PIQ was developed using the rigorous PROMIS methodology as described. This PROMIS methodology incorporates the patient perspective in the instrument development process, unlike in the development of some previous itch scales (Elman et al., 2010, Majeski et al., 2007). Second, the samples for testing and psychometric analysis of the scales were larger than typically reported for itch scale development (Desai et al., 2008, Elman et al., 2010, Majeski et al., 2007). PIQ item banks can be used to develop short forms for particular purposes or samples or administered using computerized adaptive testing. Scores on short forms and on computerized adaptive tests are reported in the same metric and are directly comparable. Third, many important concepts are represented in the PIQ item banks that were not represented in one or more existing instruments to measure the burden of itch. Some examples include "because of itch, I made more mistakes than normal", "…it was hard to do even simple tasks", "…I was absent from work", "…I had to depend on the help of other people", "…I was less active", "…I limited exercising", "…it was hard to do some types of exercise", "…my physical activities were limited", "my itch interfered with my sex life", "I worried that my itch would get worse if people touched me", and "I worried that others would see me scratching". In addition, the PIQ items were refined to exclude double-barrel questions, that is, questions that assess multiple concepts. Fourth, the PIQ scores have inherent meaning, in that T-scores allow for comparison of an individual's scores with those of the general population with chronic itch. This is potentially very useful in clinical practice and research in a setting that may not be generalizable to all patients with chronic itch and/or where healthy controls may not be available. Fifth, the PIQ items are not disease-specific and can be used to assess the burden of itch across various diseases and treatments, similar to other PROMIS item banks, such as pain interference, fatigue, and dyspnea. Itch was similarly determined by the PROMIS leadership to be included in the list of symptoms addressed in the PROMIS framework.

However, it is important to consider these itch item banks within the broader context of measuring itch. A recent review identified 23 different patient-reported outcome measures for itch, with only three patient-reported outcome measures that assess the relationship of itch with HRQoL (Schoch et al., 2017). The PIQ item banks do not objectively measure scratching activity or sleep disturbance secondary to itch, as do accelerometers (Bender et al., 2003, Bringhurst et al., 2004, Petersen et al., 2013). Additional items were developed within the PIQ to assess the intensity, frequency, timing, quality, and triggers of itch (HealthMeasures, 2018) and can be measured in addition to the PIQ item banks. The PIQ item banks quantify unidimensional latent constructs related to qualitative aspects of the HRQoL impact of itch (Silverberg et al., 2018). They include many items related to how itch directly and/or indirectly interferes with activities of daily living, social interaction, sleep

and emotional well-being, clothing choices, and physical activity. These content areas reflect the major dimensions of the burden of itch identified psychometrically in a large sample of respondents, rather than merely assessing the intensity of itch.

Some PIQ items cover concepts that are part of the generic domain framework of PROMIS, such as sleep disturbance, functional limitations, and depression. These concepts were included in the PIQ based on feedback both from patients during qualitative interviews and our expert panel. Patients with chronic itch often have comorbid health disorders that can independently impact these domains. For example, atopic dermatitis is associated with intense pruritus that can lead to sleep disturbance and impairment (Li et al., 2018), symptoms of anxiety and depression (Cheng and Silverberg, 2019, Patel et al., 2019, Silverberg et al., 2019), and limitations of physical activity (Kim and Silverberg, 2016, Silverberg et al., 2016b, Strom and Silverberg, 2016). However, atopic dermatitis is also associated with asthma and other atopic disorders that are themselves associated with sleep and mental health disturbance and limited physical activity (Cordova-Rivera et al., 2018, Zhang et al., 2016). Items aimed at measuring HRQoL impacts attributed to itch may capture the effects of itch per se and have lesser confounding by comorbidities. However, we did not investigate the validity of generic PROMIS item banks to measure the HRQoL concepts from our conceptual model. It is possible that some of the generic PROMIS banks are sufficient to measure different concepts relevant to itch. This limitation should be addressed in future studies. Of note, PIQ item banks can be used together with other PROMIS measures. The choice of which PROMIS measures to use should be tailored to the practice setting, patient population, and relevance of individual items and banks.

Although the PIQ item banks are statistically unidimensional, they measure items from seemingly different constructs. For example, bank 2 includes items for mood and sleep disturbances. Sleep and mood, though distinct symptoms, are inherently intertwined based on the impact of sleep disturbance on mood and vice versa. Previous studies found that sleep deprivation can exacerbate pre-existing mood disturbances, such as anger, depression, and anxiety (Saghir et al., 2018). Sleep and emotional distress consistently belong to the same symptom cluster in patients with cancer as well (Fan et al., 2007). We therefore were not surprised these items were unidimensional enough to be calibrated together. Bank 3 includes items for physical activity and clothing decisions. These items are interrelated by an underlying construct of atopic dermatitis being triggered or worsened by heat, sweat, and friction, which occur with moderate-to-vigorous physical activity and certain clothing types. Although these are considered disparate constructs within the general PROMIS framework, they are related in chronic itch.

Support for the validity of PIQ scores was observed in the correlations with worst and average itch intensity. NRS- and verbal rating scale–itch have both been shown to have high reliability and concurrent validity (Phan et al., 2012). In particular, NRS-itch was found to have fewer missing responses than visual analog or verbal rating scales and was preferred by patients over visual analog scale for itch (Reich et al., 2011). It is not surprising that PIQ scores only correlated moderately with itch intensity. Some individuals with mild intensity may have a high burden of itch and vice versa based on individual characteristics, such as

baseline activities and coping mechanisms. PIQ item banks should be used in conjunction with, not as an alternative to, other measures of itch intensity and characteristics.

These findings provide support for the validity of the PIQ item banks for use in adults. Future validation studies are underway. Data for this study were obtained from a large sample, representative of individuals with chronic itch. The sample was from an Internet panel and was not well characterized with respect to specific disease groups or differences across racial/ethnic, sex, or age groups. However, this limitation was offset by rapid data acquisition, demonstration of the feasibility of Internet data collection, and the practicalities of studying a large initial calibration sample. Finally, this calibration will be a springboard for further work to determine whether age, sex, race/ethnicity, or disease etiology are associated with differential item functioning. In addition, although we performed a variety of analyses useful for determining the psychometric properties of the PIQ items, additional analyses would be helpful for determining the reliability and minimally clinically important difference. Although our qualitative research supported the assessment of mood, sleep, and physical activity specifically attributed to itch, future studies are needed to determine how the measurement properties of these PIQ item banks compare with other generic PROMIS assessments that cover similar constructs.

In conclusion, the development and calibration of the PIQ item banks using classical test theory and IRT methods supports their validity. The PIQ item banks will permit researchers and clinicians to assess and integrate qualitative aspects of the burden of itch with other PROMIS measures and across a variety of health conditions. PIQ computerized adaptive tests and short forms are publicly available in the PROMIS/Assessment Center.

MATERIALS AND METHODS

After conducting a literature review and qualitative semistructured interviews, a conceptual model of itch was developed to characterize itch-relevant HRQoL domains and develop items for these domains (Silverberg et al., 2018). After initial item writing, cognitive debriefing was performed to determine if respondents understood the items as intended. Factor analytic approaches helped confirm dimensionality assumptions required for IRT analyses, resulting in itch-specific item banks.

Study design

Our conceptual model of itch was used to develop the items in the bank (reviewed in Supplementary Materials and Methods). Once the candidate item bank was developed, 75 items were field tested in 600 US adults with chronic itch (PROMIS, 2013). Participants were 18 years old, able to read and understand English, and members of a nationwide online Internet panel maintained by Op4G including >250,000 members. Op4G participants are volunteers across the US, and although nationally representative demographic quotas were applied, the panel does not necessarily provide a representative sample. Participant selection is described in the Supplementary Materials and Methods.

Statistical analysis

Psychometric analysis.—Development of the new HRQoL item banks and computerized adaptive tests involved identifying unidimensional sets of items and conducting IRT analyses to develop the calibration data needed for functioning item banks.

Data were analyzed using factor analyses implemented in MPLUS (version 6.12, Los Angeles, CA) (Muthén and Muthén, 2011); the sample was randomly divided into two separate datasets, one for EFA and the other for CFA. EFA was conducted to identify the number of potential factors within item sets using the following criteria: eigenvalues >1; scree plot review (i.e., number of factors before the break in the scree plot); and number of factors that explained >5% of the variance. A promax rotation then was used to examine the association among factors by calculating their loadings (criterion >0.4) and interfactor correlations. CFA was then used to confirm the factor structure suggested by EFA. An iterative process including clinical input was taken into account to finalize item exclusion or inclusion.

Once unidimensional item sets were identified, the graded response model (Samejima et al., 1996), as implemented in IRTPRO (version 2.1, Scientific Software International, Inc, Skokie, IL), evaluated item fit and estimated item parameters. T-scores were corrected for the stratification of itch severity using weights based on an expected severity distribution of 57.8% mild, 20.9% moderate, 21.6% severe, and 8.7% very severe among US adults with itch extrapolated from Silverberg et al. (2016).

Validation of PIQ item banks.—Convergent validity of final item bank scores was established using Pearson correlations with NRS worst or average itch. Correlation coefficients scores of 0.70 were considered strong, 0.50–0.69 moderate, and 0.30–0.49 weak. We hypothesized there would be moderate to strong positive correlations between PIQ scores and itch severity.

Discriminant validity was determined using analysis of variance and receiver operating characteristic curves (Supplementary Materials and Methods).

Internal consistency was determined using Cronbach's alpha and Pearson correlation between individual questions (Supplementary Table S3). Significant floor or ceiling effects of total scores were considered present if 15% of responses fell in the lowest or highest scores (Lim et al., 2015, Terwee et al., 2007). Statistical analyses were performed in SAS, version 9.4.3 (SAS Institute, Cary, IN) using complete case analyses. A two-sided P-value of 0.05 was considered statistically significant.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Data availability statement

All PROMIS item banks are freely available from the Assessment Center (https://www.assessmentcenter.net/). Data from the validation data set are available upon request from the authors.

Abbreviations

CFA	confirmatory factor analysis
EFA	exploratory factor analysis
HRQoL	health-related quality of life
IRT	item response theory
NRS	numerical rating scale
PIQ	Patient-Reported Outcomes Measurement Information System Itch Questionnaire
PROMIS	Patient-Reported Outcomes Measurement Information System
Q	question

References

Bender BG, Leung SB, Leung DY. Actigraphy assessment of sleep disturbance in patients with atopic dermatitis: an objective life quality measure. J Allergy Clin Immunol 2003;111:598e602.

Bringhurst C, Waterston K, Schofield O, Benjamin K, Rees JL. Measurement of itch using actigraphy in pediatric and adult populations. J Am Acad Der- matol 2004;51:893e8.

Cheng BT, Silverberg JI. Depression and psychological distress in US adults with atopic dermatitis. Ann Allergy Asthma Immunol 2019;123:179e85.

Cordova-Rivera L, Gibson PG, Gardiner PA, McDonald VM. A systematic review of associations of physical activity and sedentary time with asthma outcomes. J Allergy Clin Immunol Pract 2018;6:1968e81.e2.

Desai NS, Poindexter GB, Monthrope YM, Bendeck SE, Swerlick RA, Chen SC. A pilot quality-of-life instrument for pruritus. J Am Acad Der- matol 2008;59:234e44.

Elman S, Hynan LS, Gabriel V, Mayo MJ. The 5-D itch scale: a new measure of pruritus. Br J Dermatol 2010;162:587e93.

Fan G, Filipczak L, Chow E. Symptom clusters in cancer patients: a review of the literature. Curr Oncol 2007;14:173e9. [PubMed: 17938700]

HealthMeasures. Available PROMIS® measures for adults. http://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis/list-of-adult-measures. 2018 (accessed 30 December 2018).

Kim A, Silverberg JI. A systematic review of vigorous physical activity in eczema. Br J Dermatol 2016;174:660e2. [PubMed: 26385020]

Kantor R, Dalal P, Cella D, Silverberg JI. Research letter: impact of pruritus on quality of life-A systematic review. J Am Acad Dermatol 2016;75:885e6. [PubMed: 27576705]

Li JC, Fishbein A, Singam V, Patel KR, Zee PC, Attarian H, et al. Sleep disturbance and sleep-related impairment in adults with atopic dermatitis: A cross-sectional study. Dermatitis 2018;29:270e7. [PubMed: 30234614]

Lim CR, Harris K, Dawson J, Beard DJ, Fitzpatrick R, Price AJ. Floor and ceiling effects in the OHS: an analysis of the NHS Proms data set. BMJ Open 2015;5:e007765.

Majeski CJ, Johnson JA, Davison SN, Lauzon CJ. Itch Severity Scale: a self- report instrument for the measurement of pruritus severity. Br J Dermatol 2007;156:667e73.

Muthén LK, Muthén BO. Mplus user's guide. Los Angeles, CA: Muthén & Muthén; 2011.

Patel KR, Immaneni S, Singam V, Rastogi S, Silverberg JI. Association between atopic dermatitis, depression and suicidal ideation: A systematic review and meta-analysis. J Am Acad Dermatol 2019;80:402e10.

Petersen J, Austin D, Sack R, Hayes TL. Actigraphy-based scratch detection using logistic regression. IEEE J Biomed Health Inform 2013;17:277e83.

Phan NQ, Blome C, Fritz F, Gerss J, Reich A, Ebata T, et al. Assessment of pruritus intensity: prospective study on validity and reliability of the visual analogue scale, numerical rating scale and verbal rating scale in 471 patients with chronic pruritus. Acta Derm Venereol 2012;92: 502e7. [PubMed: 22170091]

PROMIS instrument development and validation scientific standards (version 2.0), http://www.healthmeasures.net/images/PROMIS/PROMISStandards_Vers2.0_Final.pdf; 2013 (accessed 30 December 2018).

Reich A, HJ, Ramus M, Ständer S, Szepietowski J. New data on the validation of VAS and NRS in pruritus assessment: minimal clinically important dif- ference and itch frequency measurement. Acta derm-venereol 2011;91: 636.

Saghir Z, Syeda JN, Muhammad AS, Balla Abdalla TH. The Amygdala, Sleep Debt, Sleep Deprivation, and the Emotion of Anger: A possible connection? Cureus 2018;10:e2912. [PubMed: 30186717]

Samejima F, van der Liden WJ, Hambleton R. The graded response model. In: Van der Liden WJ, editor. Handbook of modern item response theory. New York: Springer; 1996. p. 85e100.

Schoch D, Sommer R, Augustin M, Ständer S, Blome C. Patient-reported outcome measures in pruritus: A systematic review of measurement properties. J Invest Dermatol 2017;137:2069e77.

Shive M, Linos E, Berger T, Wehner M, Chren MM. Itch as a patient-reported symptom in ambulatory care visits in the United States. J Am Acad Der-matol 2013;69:550e6.

Silverberg JI, Gelfand JM, Margolis DJ, Boguniewicz M, Fonacier L, Grayson MH, et al. Symptoms and diagnosis of anxiety and depres- sion in atopic dermatitis in U.S. adults. Br J Dermatol 2019;181: 554e65.

Silverberg JI, Hinami K, Trick WE, Cella D. Itch in the general internal medicine setting: A cross-sectional study of prevalence and quality-of-life effects. Am J Clin Dermatol 2016a;17:681e90.

Silverberg JI, Kantor RW, Dalal P, Hickey C, Shaunfield S, Kaiser K, et al. A comprehensive conceptual model of the experience of chronic itch in adults. Am J Clin Dermatol 2018;19:759e69.

Silverberg JI, Song J, Pinto D, Yu SH, Gilbert AL, Dunlop DD, et al. Atopic dermatitis is associated with less physical activity in US adults. J Invest Dermatol 2016b;136:1714e6.

Strom MA, Silverberg JI. Associations of physical activity and sedentary behavior with atopic disease in United States children. J Pediatr 2016;174: 247e53.e3.

Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34e42. [PubMed: 17161752]

Udkoff J, Silverberg JI. Validation of scratching severity as an objective assessment for itch. J Invest Dermatol 2018;138:1062e8.

Zhang L, Zhang X, Zheng J, Wang L, Zhang HP, Wang L, et al. Co-morbid psychological dysfunction is associated with a higher risk of asthma ex- acerbations: a systematic review and meta-analysis. J Thorac Dis 2016;8: 1257e68.
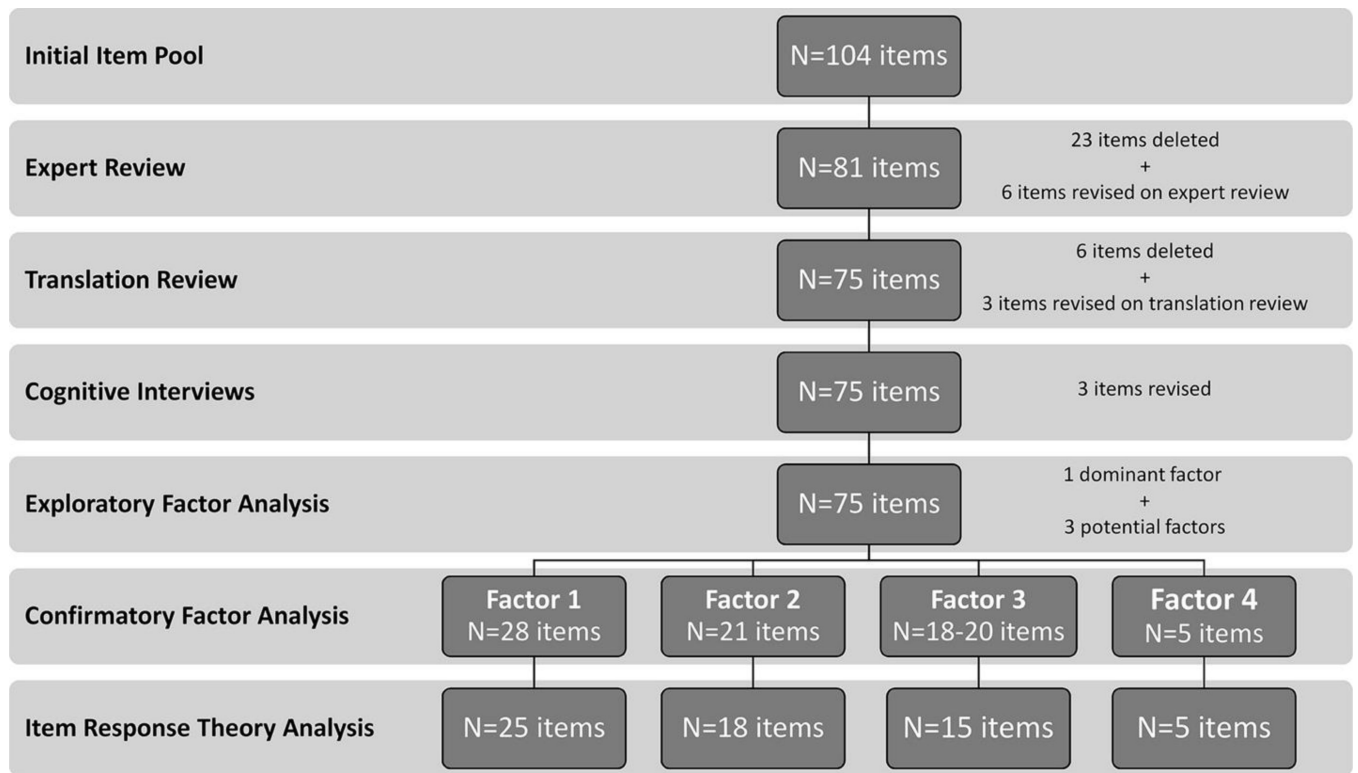
Figure 1.
Analysis overview

Table 1.

Confirmatory Factor Analysis Summary

Characteristic (predefined criteria)

Factor	No. of items	CFI (>0.9)	TLI (>0.9)	RMSEA (<0.1)	R-square (>0.3)	Residual corr (<0.15)
1. General	28	0.976	0.974	0.09	all >0.3	all <0.15
	25	0.981	0.98	0.089	all >0.3	all <0.15
2. Mood and sleep	21	0.949	0.944	0.144	all >0.3	all <0.15
	18	0.953	0.947	0.149	all >0.3	all <0.15
3. Clothing and physical activity	20	0.966	0.962	0.137	all >0.3	Q16 vs. Q17¼0.155, Q16 vs. Q18¼0.161 otherwise, all <0.15
	15	0.989	0.987	0.104	all >0.3	all <0.15
4. Scratching behavior	5	1.000	1.000	0.014	all >0.3	all <0.15

Abbreviations: CFI, comparative fit index; corr, correlation; RMSEA, root mean square error of approximation; TLI, Tucker-Lewis index; Q, question.

Table 2.

Items Identified in Exploratory Factor Analysis and Retained after Item Response Theory Analysis

Factor/Item No.	Item
Factor 1	
Q12	…because of itch, I had balance problems.
Q25	…because of itch, my thinking was slowed.
Q36	…because of itch, people avoided touching me (ex: holding my hand).
Q37	…because of itch, I avoided certain foods.
Q38	…because of itch, it was hard to shower or take a bath.
Q39	…because of itch, it was hard to watch television.
Q40	…because of itch, it was hard to play games (ex: board games, computer games, phone apps).
Q41	…because of itch, it was hard to read a book.
Q42	…because of itch, it was hard to watch a movie.
Q44	…because of itch, I made more mistakes than normal.
Q46	…because of itch, it was hard to do even simple tasks.
Q49	…my itch interfered with my sex life.
Q54	…because of itch, I was absent from work.
Q55	…because of itch, it was hard to work.
Q56	…I limited activities with others because of how my skin looked.
Q60	…because of itch, people treated me differently.
Q62	…because of itch, I avoided interacting with other people.
Q63	…my Itch bothered people around me.
Q64	…people told me to stop scratching.
Q65	…because of itch, I had to depend on the help of other people.
Q66	…I worried that my itch would get worse if people touched me.
Q67	…because of itch, it was hard to interact with my family.
Q68	…because of itch, I avoided hugging people.
Q69	…I was tired of people asking questions about my itching and scratching.
Factor 2	
Q22	….because of itch, I was nervous.
Q23	…I was afraid my itch would never go away.
Q26	…because of itch, I felt angry.
Q27	…because of itch, I felt sad.
Q28	…because of itch, I felt depressed.
Q30	…because of itch, I felt miserable.
Q31	…because of itch, I felt frustrated.
Q32	…because of itch, I felt embarrassed.
Q33	…because of itch, I was distracted.
Q34	…I worried that others would see me scratching.
Q35	…because of itch, I was self-conscious.
Q43	…because of itch, I had to stop what I was doing to scratch.
Q45	…because of itch, it was hard to sit still.

Factor/Item No.	Item
Q47	…because of itch, it was hard to relax.
Q48	…because of itch, I was restless.
Q50	…because of itch, I had difficulty falling asleep.
Q51	…because of itch, I had trouble staying asleep.
Q53	…because of itch, my sleep was restless.
Factor 3	
Q1	…because of itch, I was less active.
Q2	…because of itch, I limited exercising.
Q3	…because of itch, it was hard to do some types of exercise.
Q4	…because of itch, it was hard to run.
Q5	…because of itch, it was hard to walk.
Q6	…because of itch, it was hard to do vigorous physical activity, such as running, lifting heavy objects, or participating in strenuous sports.
Q7	…because of itch, it was hard to do moderate physical activity, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf.
Q8	…because of itch, it was hard to do light physical activity.
Q9	…because of itch, I sat around more than usual.
Q10	…because of itch, it was hard to do activities that made me sweat.
Q11	…because of itch, I did as little as possible physically.
Q13	…because of itch, I limited the clothing I could wear.
Q14	…because of itch, it was hard to wear short-sleeves.
Q19	…because of itch, it was hard to wear certain shoes.
Q21	…because of itch, my physical activities were limited.
Factor 4	
Q71	…I scratched myself until I bled.
Q72	…I worried about flaking skin from scratching.
Q73	…I worried about getting scars from scratching.
Q74	…it was hard to stop scratching or rubbing.
Q75	…I worried about having open wounds from scratching.

Abbreviation: Q, question.

Table 3.

Pearson Correlation of PIQ Item Bank T-Scores with Each Other and Itch Severity

| | Pearson r | | | | | |
| | Item bank | | | | NRS-itch | |
Item bank	1	2	3	4	Worst itch	Average itch
1	1.00	—	—	—	0.56	0.60
2	0.86	1.00	—	—	0.65	0.66
3	0.87	0.84	1.00	—	0.63	0.65
4	0.78	0.80	0.73	1.00	0.60	0.62

Abbreviations: NRS, numerical rating scale; PIQ, Patient-Reported Outcomes Measurement Information System Itch Questionnaire.

Table 4.

Area Under the ROC Curve for PIQ Item Bank T-Scores as Predictors of Itch Severity

	AUC			
Itch severity	**Item bank 1**	**Item bank 2**	**Item bank 3**	**Item bank 4**
Worst itch				
Moderate vs. mild	0.66	0.76	0.72	0.72
Severe vs. mild	0.82	0.92	0.87	0.85
Severe vs. moderate	0.68	0.75	0.71	0.69
Very severe vs. mild	0.89	0.95	0.92	0.93
Very severe vs. moderate	0.80	0.85	0.83	0.82
Very severe vs. severe	0.66	0.70	0.70	0.69
Average itch				
Moderate vs. mild	0.67	0.74	0.72	0.73
Severe vs. mild	0.86	0.93	0.89	0.88
Severe vs. moderate	0.74	0.77	0.78	0.73
Very severe vs. mild	0.89	0.94	0.91	0.91
Very severe vs. moderate	0.81	0.85	0.85	0.80
Very severe vs. severe	0.66	0.69	0.68	0.64

Abbreviations: AUC, area under the curve; PIQ, Patient-Reported Outcomes Measurement Information System Itch Questionnaire; ROC, receiver operating characteristic.

Table 5.

Optimal Thresholds for T-Scores and their Sensitivity and Specificity for Item Banks as Predictors of Itch Severity

Average itch	Item bank 1			Item bank 2			Item bank 3			Item bank 4		
	Threshold	Sensitivity	Specificity	Threshold	Sensitivity	Specificity	Threshold	Sensitivity	Specificity	Threshold	Sensitivity	Specificity
Mild vs. no itch	46.4	52.2%	100%	44.8	56.1%	75.0%	48.5	58.0%	100%	42.2	36.3%	100%
Moderate vs. mild	50.5	58.5%	67.5%	49.7	70.7%	73.2%	49.8	72.8%	65.6%	49.7	67.1%	71.3%
Severe vs. moderate	56.5	66.2%	69.9%	56.9	69.1%	72.0%	57.9	63.2%	80.5%	58.1	61.8%	77.6%
Very severe vs. severe	62.7	61.4%	70.6%	65.0	59.6%	75.7%	65.3	56.1%	80.9%	63.0	54.4%	71.3%