HHS Public Access

Author manuscript

J Burn Care Res. Author manuscript; available in PMC 2022 June 03.

Published in final edited form as:

J Burn Care Res. 2021 March 04; 42(2): 212–219. doi:10.1093/jbcr/iraa133.

Development of proxy and self-report Burn Model System Pediatric Itch Interference scales: A National Institute on Disability, Independent Living, and Rehabilitation Research Burn Model System Study

Dagmar Amtmann, PhD,
University of Washington, Seattle, WA, United States

Alyssa M. Bamer, MPH,
University of Washington, Seattle, WA, United States

Kara McMullen, MPH,
University of Washington, Seattle, WA, United States

Karen Kowalske, MD,
University of Texas Southwestern, Dallas, TX, United States

Jeffrey C. Schneider, MD,
Spaulding Rehabilitation Hospital, Boston, MA, United States

Nicole S. Gibran, MD
University of Washington, Seattle, WA, United States

Abstract

Pruritus is a commonly reported symptom after burn injury. Valid and reliable scales to measure itch in pediatric burn survivors is important for treatment and epidemiological studies. This study sought to develop psychometrically sound, publicly available self- and proxy-report measures of itch for use in pediatric burn survivors suitable for use in research and clinical practice. A panel of burn experts developed a definition of itch interference and a set of parallel self- and proxy-report candidate item that covered important activities affected by itch. Candidate items were evaluated in cognitive interviews with pediatric burn survivors (n=4) and proxies (n=2). Items were translated to Spanish and administered in both English and Spanish to a sample (N=264) of pediatric burn survivors and/or their proxy enrolled in the Burn Model System (BMS) longitudinal database. The mean age of the pediatric sample was 13 years and average time since burn 5 years. The final itch interference measures each included 5 parallel items calibrated using a one parameter graded response item response theory model, with a mean of 50 representing the average itch interference of the sample. Reliability of the scores is excellent between the mean and two standard deviations above. Initial analyses provide support for validity of the score. Concordance between the self- and proxy-report scores was moderate (ICC=0.68). The results support the reliability and validity of the itch scale in children and youth with burn

Corresponding Author: Dagmar Amtmann, University of Washington, Rehabilitation Medicine, Box 354237, Seattle, WA 98195; Phone: 206-543-4741, Fax: 206-685-3244; dagmara@u.washington.edu.

injury. The new BMS Pediatric Itch Interference scales are freely and publicly available at https://burndata.washington.edu/itch.

Keywords

pruritus; burn injury; burn model systems; scale development

Introduction:

Pruritus is a commonly reported and unpleasant symptom after burn injury, and has been reported to affect up to 90% of adult and pediatric burn survivors at discharge from acute care.[1,2] Whereas pruritus does resolve over time for many patients, it remains problematic for up to 60% to 70% of pediatric and adult burn survivors two-years post injury.[1,2] Pruritis can be a significant cause of distress, and can impact wound healing, sleep, concentration, and overall quality of life.[3-5]

Studies in adults and children indicate that pruritus can be difficult to treat due to the complexity of treatment options and potential need for referral to specialists.[6] In addition, the evidence for effectiveness of treatments for pruritus in burns is of low quality and a lack of treatment consensus exists across clinical settings.[6] Very few studies have specifically examined the effectiveness of pruritus treatments in children with burn injuries,[6,7] though a recent study indicated that only 30% of those provided with anti-itch treatment achieved complete symptom relief.[7] Given the paucity of research and the high prevalence of pruritus in pediatric burn survivors, there is a need for future research to identify effective treatment protocols and potential new therapies.

In order to evaluate interventions or medications to improve pruritis, measurement tools that are sensitive and specific are necessary. Two measures of itch, the Toronto Pediatric Itch Scale and The Itch Man Scale, have been developed for use specifically in pediatric burn patients.[8,9] While other measures of itch have been utilized in studies with burn survivors, including the Burns Itch Questionnaire (BIQ),[10] the 5-D Itch scale,[11] and the commonly used Visual Analogue Scale (VAS),[12] they have not been validated in pediatric burn survivors. In addition, the 5-D and BIQ were developed for adults and contain items which may be inappropriate for the pediatric patient (i.e. items relating to work or partner relationship), and some evidence suggests the VAS may not be suitable for use in individuals with cognitive limitations, including children.[12] Of the two pediatric itch measures currently available, the Itch Man Scale[8] was the first tool to be developed for pediatric burn patients and consists of a graphic image containing five different stick figure images with corresponding descriptions. It was designed for burn patients age 6 years or older, and it generates an itch severity score between zero and four based on the image they select. Alternatively, the Toronto Pediatric Itch Scale[9] is designed for use in children under 5 years of age and consists of a single proxy report item which asks raters to assign a score for itch severity between zero and three. While some reliability and/or validity evidence is available for both scales, this evidence is generally weak[8,9] and interpretation of the scores is problematic as neither scale has been evaluated for sensitivity to change, suggesting limited

usefulness for use in assessing the impact of pruritis treatments. While a proxy version of a scale can be used to assess pruritus when a child is unable to self-report or used in addition to pediatric report, validity of proxy report on the Itch Man scale has not been evaluated. The inter-observer agreement for the Toronto scale was moderate overall (kappa=0.52) and weak for those with intermediate itch.[9] Lastly, neither scale was designed to assess interference of itch with daily activities or changes in itch after treatment, but rather are single item measures of itch severity.

In addition to the limitations already described, one item measures are typically less reliable than multi-items scales, especially multi-item instruments developed using Item Response Theory (IRT). Compared to Classical Test Theory, IRT methodology can significantly improve psychometric properties of multi-item scales and provide scores on the same metric, regardless of the mode of administration or the specific items (from an item bank) administered. Because of these and other benefits, IRT-based instruments have become increasingly popular in health assessment and clinical research.[13-15] Given the limitations of the currently available measures of pruritus, the objective of this study was to develop publicly available self- and proxy-report BMS Pediatric Itch Interference (BMS-PII) measures for use in pediatric burn survivors using modern psychometric theory. We sought to develop psychometrically sound, clinically meaningful instruments suitable for use in research and clinical practice. To further enhance the validity and utility of the measures, we also included burn experts and burn survivors in the measure development process, as is recommended practice in measure development.[16]

Methods:

This study took place as a part of the larger Burn Model System (BMS) national longitudinal database, which has been described in depth elsewhere.[17] Study procedures were approved at all participating BMS sites by their respective Institutional Review Boards. Development of the BMS-PII scales followed many of the procedures used by national measurement initiatives,[13,15] including feedback from the target audience and use of modern measurement theory. The first step in the study was to define the construct of itch interference and develop items that measure it. This was done in consultation with an expert panel of burn researchers, burn clinicians, and psychometricians. Next, items were reviewed by individuals with burn injuries or a proxy using a cognitive interview process. Items were then administered to a sample of children with burn injury and/or their proxy. Using this data, items were analyzed using modern psychometric techniques including item response theory, and preliminary validity analyses were completed.

Participants and Procedures:

Expert Panel: A panel of eight BMS burn researchers from institutions across the United States met multiple times to discuss the definition and purpose of self- and proxy-report BMS-PII measures, appropriate length of the measures, format and wording of items, important life activities that may be affected by itch, and to choose individual item content. Once candidate items were proposed, the panel considered each item’s appropriateness and relevance to itch in pediatric burn injury prior to submission for cognitive interview testing.

Cognitive Interviews: Candidate items chosen by the expert panel were tested in cognitive interviews with pediatric burn injury survivors and parent/proxy responders for children with a burn injury. Cognitive interview participants were recruited from the BMS center at the University of Texas Medical Branch at Galveston and the BMS center at the University of Texas Southwestern. Participants were recruited from a convenience sample of participants already taking part in the larger BMS longitudinal database.[17] Interviews were conducted in either English or Spanish to youth ages 12 to 17 and to parents of burn survivors aged 12 to 15.

Structured cognitive interviews[18] were completed on each of the proposed items. Cognitive interviewers completed written and verbal training sessions with an experienced investigator with training in measurement. Cognitive interviews were used to ensure that the items were understandable, relevant and meaningful to the target audience. The interview process involved asking participants to first respond to all items, followed immediately by probing questions about how they arrived at their answer and questions about clarity and meaningfulness of the question. Participants were also asked to give feedback on the appropriateness and clarity of the response options and instructions. Any problematic items were modified or deleted based on interview feedback. Each proposed type of item (either self-report or the parallel proxy report item) was reviewed by at least 5 individuals (i.e. either a burn injury survivor or a proxy).

Calibration Study: Items deemed clear and appropriate based on cognitive interview feedback were incorporated into the BMS national longitudinal database follow-up surveys for pediatric participants and their proxies (burn survivors being <18 years at follow-up) beginning in June, 2015. A description of the BMS and it's goals and ideology have been previously described [19] and an overview of the type of data available has been published.[19,20] In brief, the BMS currently includes four burn centers that contribute data to an ongoing longitudinal database that was initiated in 1994. Baseline data are collected from eligible consenting participants within 30 days of their acute hospital discharge following their burn injury. Follow-up data are collected during specific follow-up windows around 6 months, 1-year, 2-years, and every 5-years post-injury for the lifetime of the participant. Inclusion criteria for individuals under 18 years of age include burn surgery for wound closure, primary treatment at a BMS center, and one of the following: a) Total Body Surface Area (TBSA) burned of 20% or greater; or b) deep second or third degree burns with serious functional or cosmetic threat that involve face, hands, feet, genitalia, perineum, or major joints; or c) deep electrical burns including lightning injury or inhalation injury with burn injury. Data is collected using a combination of medical record abstraction and self-report surveys administered either via paper/pencil, online, or phone or in-person interview. Participants can complete the survey in either English or Spanish. Self-report and proxy report surveys for pediatric participants are both administered until at least the age of 14, though some BMS sites administer both through the age of 17. While all BMS sites enroll pediatric burn survivors, the majority of the pediatric participants are enrolled at the Shriners Hospitals Pediatric Burn Center in Galveston, Texas location. Only pediatric participants (<18 years) who were over the age of 8 at the time of follow-up assessment, and/or their proxy, were invited to complete the candidate itch items. If participants responded at more

than one follow-up timepoint to the itch items, only their first response was used (i.e. only one record from each individual and/or their proxy was included).

Measures:

The BMS longitudinal database includes demographic and multiple health and symptom related measures. Burn related variables, including Total Body Surface Area (TBSA) burned, etiology of burn, and location of burn, are collected via medical record abstraction. Demographics and health outcome measures are collected via self-report and/or proxy-report. Specifically, self-reported pain interference, physical function, and peer relationships were measured using the PROMIS Pediatric Profile 25 v2.0, which includes four items that measure each of six domains.[21] Proxy-reported peer relationships, pain interference, and physical function were assessed using the PROMIS Peer Relationships Short Form 7a proxy, PROMIS Pain Interference Short Form 8a proxy, and PROMIS Physical Function Mobility Short Form 8a proxy. The PROMIS profile and short forms were scored following developer guidelines.[22-24] Items are scored on the T-score metric and a mean of 50 represents the mean of the general United States population. Higher scores indicate more of the trait being measured (i.e. more pain interference, better physical function, and better peer relationships).

Itch Interference: The candidate BMS-PII self- and proxy-report items were included in the BMS longitudinal database at all follow-up timepoints. The candidate items are described in detail below, but included five items measuring important aspects of itch interference. The items utilized the time frame of the past 7 days and a five-point frequency response scale of “Never” to “Almost Always”. A Spanish language version of the items was also created for this study as a significant percentage of participants are Spanish speaking. The Spanish translation was reviewed by at least two different research coordinators fluent in Spanish and then consensus meetings were held to discuss the translation and decide upon the final version.

Data Analyses:

Descriptive statistics were calculated for the calibration sample on demographic and burn related variables. All analyses, with the exception of IRT analyses, were completed using Stata 14.2.[25]

IRT Analyses: Data from the calibration study were used to calibrate items to a one parameter graded response IRT model[26] and evaluate score reliability. All analyses were completed on self- and proxy-report responses separately (i.e. they were treated as separate measures). As a first step, the IRT assumptions of unidimensionality and local independence were examined after fitting a one factor confirmatory factor analysis (CFA) using the WLSMV estimator in Mplus 7.2.[27] Unidimensionality is supported by a comparative fit index (CFI) of 0.90 or higher,[28] while local independence violations can be examined using the residual correlations from the CFA, as residual correlations greater than 0.2 suggest local dependance.[29] Item pairs identified as having significant local dependance were examined and problematic items dropped from the banks if necessary. Items that met the assumptions of unidimensionality and local independence were calibrated to a one parameter graded

response IRT model (i.e. discrimination parameters were set to be equal across items) using IRTPRO 4.2.[30] Item fit was evaluated by examining the stability and size of the item calibration parameter standard errors.If the standard errors were judged to be sufficiently small across all discrimination and difficulty parameter estimates no items were dropped from the bank.

Final item parameter estimates were calculated for both proxy and self-report banks and were centered in the calibration study sample (i.e. a score of 50 is the mean of the calibration study sample). In addition, item response curves and individual and combined test information functions were extracted from IRTPRO. The response curves were examined to evaluate the function of the response options. Ideally, category response curves have distinct peaks, indicating the most likely response category to be selected by a respondent at each level of the trait (i.e. itch interference).[31] The test information functions were converted to Classical Test Theory reliability estimates and plotted along the T-score continuum with a histogram of participant scores. Reliability of 0.8 and 0.9 in the Classical Test Theory framework corresponds to scale information in the IRT framework of 5 and 10 respectively.[32] In general, reliability of 0.8 is considered sufficient for group comparisons while 0.9 is required for individual comparisons.[33,34] The range of scores for which the BMS-PII scales measure above 0.9 or 0.8 reliability was also calculated. Floor and ceiling effects were examined and are considered to be present if more than 15% of respondents achieved the lowest or highest possible score, respectively.[35]

Validity Analyses: The construct validity of the BMS-PII scales was evaluated by examining Pearson's correlations between BMS-PII scores and other self- or proxy-report outcome measures. Based on prior literature we hypothesized that BMS-PII scores would be moderately positively correlated with pain interference scores ($r \approx 0.5$),[1] while having a low degree of correlation with less related constructs including physical function and peer relationships ($\sim r \leq -0.3$). Though the current literature is somewhat conflicting on the association between TBSA and itch,[1,7] we examined known groups validity by comparing BMS-PII scores between TBSA burn groups. We hypothesized that BMS-PII scores would be significantly higher for individuals with higher TBSA burn (>40%) than below.

Self and Proxy Report Comparison: To assess agreement between self and proxy BMS-PII responses a number of analyses were completed. Overall scale agreement was examined by calculating average differences between self and proxy total scores and a Pearson correlation coefficient. In addition, a 2-way mixed effects model (absolute agreement, individual measures) intraclass correlation coefficient (ICC)[36] was performed. We considered ICC values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 indicative of poor, moderate, good, and excellent reliability, respectively.[37] Individual item agreement and concordance was examine using weighted versions (quadratic) of both Cohen's kappa and an alternative to kappa, the Gwet's AC_2 [38] Gwet's AC is a more stable inter-rater reliability coefficient, less affected by prevalence and marginal probability, and does not depend on the assumption of independence of raters.[39] We considered a kappa or Gwet's AC value of 0.21-0.4 fair agreement, 0.41 to 0.6 moderate

agreement, 0.61 to 0.8 substantial agreement, and greater than 0.81 excellent agreement.[40] Gwet's AC was calculated in Stata[25] with the user written *kappaetc*[41] command.

Results:

Expert Panel:

The expert panel included eight MD or PhD level burn clinicians and researchers, at least one with training in measurement. The panel decided that the new BMS-PII scales should measure how often itch interferes with important aspects of a child's life, including the extent to which itch hinders engagement with social, cognitive, emotional, and recreational activities as well as ability to sleep. Because the panel wanted to include the new scales in the existing BMS survey, brevity of the scales was of extreme importance to the panel. Thus, the panel wanted to include only one question about each of the important aspects of itch interference (social, cognitive, emotional, recreational activities and sleep). The panel decided to use the 7 day time frame (i.e. "In the past 7 days…") to make the scale appropriate for use in clinical trials where the 7 day time frame is the most common and the frequency response options recommended by the PROMIS initiative (i.e. "Never" to "Almost Always").[42] The panel developed 20 candidate items to assess how itch affects engagement with sleep, social, cognitive, emotional, and recreational activities. Of the 20 candidate items, the 5 items (i.e. one per subdomain) that were voted most preferred or most relevant by the expert panel were chosen for the candidate measures (see Table 1). Parallel items were then created for proxy responders with identical response options (i.e. "*I felt angry when I was itching"* became "*My child felt angry when he/she was itching"*).

Cognitive Interviews:

A total of four pediatric participants with a burn injury completed cognitive interviews on the five self-report itch items, and two parent/proxy responders completed the cognitive interview of the five parallel proxy-report items. Pediatric respondents ranged in age from 12 to 15 years (mean age: 13 years) and proxy responders included one mother and one father of different 12-year-old children with burn injury. Mean burn size (% TBSA) of the pediatric burn participants was 41% (range: 5-50%). All six participants found the items clear and appropriate. None indicated difficulties with the response options or instructions. Thus, no changes were made to either of the five item sets prior to administration in the calibration sample.

Calibration Sample:

A total of 264 pediatric participants and/or their proxy completed at least one of the candidate BMS-PII items during at least one follow-up assessment (only their first response was used if they completed it at multiple timepoints). Of these, 247 participants completed the self-report candidate items, 260 completed proxy-reports, and 243 had both self- and proxy-report. Individual item missingness ranged from 0% to 1% across the 10 itch items. The final dataset included 11% (n=28) collected at six months, 27% (n=71) at 12-months, 11% (n=30) at 24-months, 23% (n=61) at 5 years, 22% (n=58) at 10 years, and 6% (n=16) at 15 years post burn injury. The sample was majority male (n=179), mean age of the pediatric burn participant was 13.1 years, and the average time since burn injury at BMS-PII

administration was 5.0 years (see Table 2). The majority of participants completed the items in Spanish (n=207) and were of Hispanic ethnicity (n=210).

IRT Analyses: The results of the CFA for both self- and proxy-report supported unidimensionality (both CFI=0.99) and no items displayed residual correlations greater than 0.2 on either self- or proxy-report. After calibration to the unidimensional graded response IRT model item parameter standard errors were sufficiently small (0.06 to 0.18 for difficulty and 0.35 for discrimination) to retain all 5 items for both self- and proxy-report (see Table 1 for item parameters). Response options were acceptable for all items. There was some support for fewer responses for some items, but the response curves were acceptable and for consistency and simplicity we decided to retain 5 response options for all items. The reliability of both five item self- and proxy-report measures is high (>0.9) between T-scores of 50 and 72 (see Figure 1) and good (>0.8) between T-scores of 48 and 74. When examining only those who report at least some itch interference (i.e. those not answering “Never” to all items), the percentage of participants measured with reliability greater than 0.9 is 94% on self and 95% on proxy and greater than 0.8 is 100% and 95% on self- and proxy-report measures, respectively. As can be seen in Figure 1, approximately half of the sample (n=127 (51%) self-report; n=134 (52%) proxy-report) reported no itch interference (“Never” on all items) and are at the floor of the measure. This is to be expected as individuals with no itching would not experience itch interference. A small number of participants (n=7 (3%) self-report; n=5 (2%) proxy-report) answered “Always” to every item and were at the ceiling of the measure, indicating the ceiling effects are minimal.

Validity Analyses: Correlations between BMS-PII self- and proxy-report and pain interference were 0.50 and 0.55 respectively. Correlations between self-report itch and self-reported physical function and peer relationships were −0.28 and −0.13, respectively. Similarly, correlations between proxy-report BMS-PII and proxy-report physical function and peer relationships were −0.28 and −0.23, respectively. When comparing BMS-PII by percent TBSA burned, self-reported itch interference was higher for those with higher TBSA burned (>40%) (t=−2.3, p=0.02) while proxy-report was not significantly different by TBSA burned (t=−1.6, p=0.10). Thus, these results support construct validity of the BMS-PII scores, as the correlations were of the magnitude and in the direction hypothesized. Known groups validity was also supported by the TBSA burned self-report results, though not by the proxy-report results.

Self and Proxy-Report Comparison: By design mean BMS-PII scores in the sample were 50.0 (SD:8.8) on self-report and 50.0 (SD:8.8) on proxy-report. The average mean difference was 0.03 points (SD:7.0) for self- minus proxy-report. The Pearson correlation coefficient between the self- and proxy-report T-scores was 0.68 (n=243) and the ICC was 0.68 (95% CI:0.61-0.74), which corresponds to moderate concordance. Individual item weighted Kappa values ranged from 0.63 to 0.66 while Gwet’s AC values ranged from 0.83 to 0.87 (see Table 2). Depending on which statistic is considered, proxy item reliability for all items is either substantial or excellent.

Discussion:

Pruritus is a common symptom after burn injury and can remain problematic for over half of burn survivors at two years post injury.[1,2] The availability of psychometrically sound instruments for measuring itch are essential for studying the efficacy of treatments for improving itch in burn patients, an area in which there are currently no recommended treatment protocols. The availability of such an instrument could also potentially allow clinicians to identify patients in need of further treatment for pruritus when used as a screening tool in clinical settings.

This study reports on the development of new pediatric itch interference scales for use by either self- or proxy- report. These parallel scales are the first itch scales to be developed for use in a pediatric population using modern psychometric methodology, including calibration to an IRT model. In addition, the BMS-PII scales were developed following established guidelines by engaging important stakeholders in the development of the construct definition and item bank content, as well as completion of cognitive interviews. Results of this study indicate that the BMS-PII scales have good psychometric properties while being brief and flexible.

The final BMS-PII scales were calibrated as item banks which allows them to be administered either through computer adaptive testing or as a custom short form, while maintaining comparability of scores across administration modes. The final scores are centered in this study sample such that a score of 50 represents the mean of this sample of pediatric participants as self-reported or reported by caregiver proxy-report of children and youth with severe burn injuries. There was ≤1% missing data on any of the self- or proxy-report items indicating items were applicable to nearly all patients. When administered in full as 5-item scales, the reliability of both BMS-PII scales is very high above the mean T-score (i.e., between 50 and 72), indicating they can be used for individual comparisons within this range. Scores are less reliable for individuals with inch interference below the mean. This is clinically appropriate as it is more important to have higher reliability for assessing and discriminating itch interference in individuals with higher itch interference who may be in need of more or better treatment options. In addition, the pattern of correlations between itch and pain, physical function, and peer relationships all provided support for the construct validity of the measures.

A further strength of this study is the development of parallel self- and proxy-report scales. Whereas it is well documented that proxy-response is not equivalent to patient report and self-report information is preferable when possible,[43] information about children's symptoms is often solicited from caregivers for multiple reasons. For example, parent perceptions of their child's health are the principal determinant of utilization of health care services[44] and may have unique and important value.[45,46] In addition, proxy report may be the only option in many instances when children are unable (e.g. illness, cognitive limitations, reading ability) or unwilling to provide responses. Thus, the availability of a proxy-report option for assessing itch interference for this population is of high value, and our use of parallel item sets makes comparisons between self- and proxy-report easier and more meaningful.[47,48] In this study the individual item agreement ranged between

substantial and excellent, and overall score correspondence between self- and proxy-report scores was moderate. This is similar to or higher than agreement reported between many other health related outcome measures.[43]

Limitations and Future Research:

One limitation of this study is the use of a sample of pediatric burn patients with severe burn injuries, the majority of which were of Hispanic background and from one data collection site within the BMS. Future research should examine the validity and reliability of the BMS-PII scales in children with less severe burn injuries as well as in samples with more diverse racial or ethnic backgrounds. Also, the sample did not include any self- or proxy-report for children less than 8 years of age. While evidence supports the validity of self-report health related quality of life in children as young as 5 years,[49] further validity testing in children 5 to 8 years would need to be done before the scale can be recommended for use below age 8. In addition, because the items are not burn specific the scales could potentially be administered in populations that experience itch for reasons other than burns. However, studies are needed to examine the appropriateness of the BMS-PII scales and their psychometric properties in other populations. Lastly, due to the limited sample size, we were unable to examine if the BMS-PII displayed differential item function (i.e. bias) by language of administration or other participant characteristics such as gender, race, or age group. For example, it is not known if there are cultural differences in development or perceptions of itch, though some research suggests itch may differ across racial or ethnic groups.[50,51] Thus, future studies should examine if cultural or other demographic characteristic influence responses to the items.

Conclusion:

This study found strong evidence of validity and reliability of the BMS-PII scales in pediatric burn survivors. The self- and proxy-report scales are the first pediatric itch measures developed using modern psychometric theory. The reliability of the scores is high for scores at or above the mean, indicating the scale could be used in clinical trials or research evaluating effectiveness of itch interventions. The scales allow for flexible modes of administrations (e.g. computer and paper pencil), including computer adaptive testing. The scales are available in both English and Spanish languages, are publicly available and free for use. User guides including the formatted scales and scoring instructions are available for download at https://burndata.washington.edu/itch.

Acknowledgements:

The authors would like to acknowledge Gretchen J. Carrougher MN, RN, David Herndon, MD, and Walter Meyer, PhD, for their contributions to the conceptualization and development of the scale, including the scale definition and development of candidate items.

Conflicts of Interest and Source of Funding:

The authors have no conflicts of interest to report.

The contents of this manuscript were developed under a grant from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR grant number 90DP0053). NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS). The contents

of this manuscript do not necessarily represent the policy of NIDILRR, ACL, HHS, and you should not assume endorsement by the Federal Government.

References:

1. Schneider JC, Nadler DL, Herndon DN, et al. Pruritus in pediatric burn survivors: Defining the clinical course. J Burn Care Res. 2015;36(1):151–158. [PubMed: 25162949]
2. Carrougher GJ, Martinez EM, McMullen KS, et al. Pruritus in adult burn survivors: Postburn prevalence and risk factors associated with increased intensity. J Burn Care Res. 2013;34(1):94–101. [PubMed: 23079565]
3. Nedelec B, LaSalle L. Postburn Itch: A Review of the Literature. Wounds. 2018;30(1):E118–E124. [PubMed: 29406295]
4. Mendham JE. Gabapentin for the treatment of itching produced by burns and wound healing in children: A pilot study. Burns. 2004;30(8):851–853. [PubMed: 15555801]
5. Parnell LKS, Nedelec B, Rachelska G, et al. Assessment of pruritus characteristics and impact on burn survivors. J Burn Care Res. 2012;33(3):407–418. [PubMed: 22210065]
6. Richardson C, Upton D, Rippon M. Treatment for wound pruritus following burns. J Wound Care. 2014;23(5):227–233. [PubMed: 24810306]
7. Nieuwendijk SMP, de Korte IJ, Pursad MM, et al. Post burn pruritus in pediatric burn patients. Burns. 2018;44(5):1151–1158. [PubMed: 29563015]
8. Morris V, Murphy LM, Rosenberg M, et al. Itch assessment scale for the pediatric burn survivor. J Burn Care Res. 2012;33(3):419–424. [PubMed: 22561307]
9. Everett T, Parker K, Fish J, et al. The construction and implementation of a novel postburn pruritus scale for infants and children aged five years or less: Introducing the toronto pediatric itch scale. J Burn Care Res. 2015;36(1):44–49. [PubMed: 25094009]
10. Van Loey NE, Hofland HW, Hendrickx H, et al. Validation of the burns itch questionnaire. Burns. 2016;42(3):526–534. [PubMed: 26778706]
11. Amtmann D, McMullen K, Kim J, et al. Psychometric Properties of the Modified 5-D Itch Scale in a Burn Model System Sample of People With Burn Injury. J Burn Care Res. 2017;38(1):e402–e408. [PubMed: 27388881]
12. Reich A, Heisig M, Phan NQ, et al. Visual analogue scale: evaluation of the instrument for the assessment of pruritus. Acta Derm Venereol. 2012;92(5):497–501. [PubMed: 22102095]
13. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol. 2010;63(11):1179–1194. [PubMed: 20685078]
14. Tulsky DS, Kisala PA, Victorson D, et al. Overview of the Spinal Cord Injury--Quality of Life (SCI-QOL) measurement system. J Spinal Cord Med. 2015;38(3):257–269. [PubMed: 26010962]
15. Gershon RC, Lai JS, Bode R, et al. Neuro-QOL: Quality of life item banks for adults with neurological disorders: Item development and calibrations based upon clinical and general population testing. Qual Life Res. 2012;21(3):475–486. [PubMed: 21874314]
16. Amtmann D, Cook KF, Johnson KL, et al. The PROMIS initiative: involvement of rehabilitation stakeholders in development and examples of applications in rehabilitation research. Arch Phys Med Rehabil. 2011;92(10 Suppl):S12–9. [PubMed: 21958918]
17. Amtmann D, McMullen K, Bamer A, et al. National Institute on Disability, Independent Living, and Rehabilitation Research Burn Model System: Review of Program and Database. Arch Phys Med Rehabil. 2017.
18. Willis G. Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks, CA: Sage Publications; 2005.
19. Klein MB, Lezotte DL, Fauerbach JA, et al. The National Institute on Disability and Rehabilitation Research burn model system database: A tool for the multicenter study of the outcome of burn injury. J Burn Care Res. 2007;28(1):84–96. [PubMed: 17211206]
20. Goverman J, Mathews K, Holavanahalli RK, et al. The National Institute on Disability, Independent Living, and Rehabilitation Research Burn Model System: Twenty Years of

Contributions to Clinical Service and Research. J Burn Care Res. 2017;38(1):e240–e253. [PubMed: 27294859]

21. Patient Reported Outcomes Measurement Information System. PROMIS Pediatric and Parent Proxy Profile Instruments Scoring Manual. 2018. http://www.healthmeasures.net. Accessed October 7, 2019.
22. DeWalt DA, Thissen D, Stucky BD, et al. PROMIS pediatric peer relationships scale: Development of a peer relationships item bank as part of social health measurement. Heal Psychol. 2013;32(10):1093–1103.
23. Irwin DE, Stucky BD, Thissen D, et al. Sampling plan and patient characteristics of the PROMIS pediatrics large-scale survey. Qual Life Res. 2010;19(4):585–594. [PubMed: 20204706]
24. Varni JW, Stucky BD, Thissen D, et al. PROMIS Pediatric Pain Interference Scale: an item response theory analysis of the pediatric pain item bank. J Pain. 2010;11(11):1109–1119. [PubMed: 20627819]
25. StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.
26. Samejima F. Graded response models. In: Handbook of Item Response Theory; 2016. doi:10.1201/9781315374512.
27. Muthén LK, Muthén BO. 2007. Mplus User's Guide (Eigth Edition). Los Angeles, CA: Muthén & Muthén.
28. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Model. 1999;6(1):1–55.
29. Kim D, de Ayala RJ, Ferdous AA, et al. The comparative performance of conditional independence indices. Appl Psychol Meas. 2011;35(6):447–471.
30. Cai L, Thissen D, duToit S. 2015. IRTPRO for Windows. Version 4.2. Scientific Software International.
31. Nguyen TH, Han H-R, Kim MT, et al. An introduction to item response theory for patient-reported outcome measurement. Patient. 2014;7(1):23–35. [PubMed: 24403095]
32. Thissen D. Reliability and measurement precision. In: Wainer H, Dorans N, eds. Computerized Adaptive Testing: A Primer. 2nd ed. Mahwah, N.J.: Lawrence Erlbaum Associates Publishers; 2000:159–184.
33. Hahn EA, Cella D, Chassany O, et al. Precision of health-related quality-of-life data compared with other clinical measures. Mayo Clin Proc. 2007;82(10):1244–1254. [PubMed: 17908530]
34. Nunnally J, Bernstein I. Psychometric Theory. 3rd ed. New York, NY: McGraw-Hill, Inc; 1994.
35. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42. [PubMed: 17161752]
36. Koch GG. Intraclass correlation coefficient. In: Kotz S, Johnson NL, eds. Encyclopedia of Statistical Sciences. 4th ed. New York, NY: John Wiley & Sons; 1982:213–217.
37. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155–163. [PubMed: 27330520]
38. Gwet KL. Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement among Raters. 3rd ed. Gaithersburg, MD: Advanced Analytics Press; 2012.
39. Wongpakaran N, Wongpakaran T, Wedding D, et al. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: A study conducted with personality disorder samples. BMC Med Res Methodol. 2013;13(1).
40. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977;33(1):159–174. [PubMed: 843571]
41. Klein D KAPPAETC: Stata module to evaluate interrater agreement," Statistical Software Components S458283, Boston College Department of Economics, revised 06 Jan 2019.
42. DeWalt DA, Rothrock N, Yount S, et al. Evaluation of item candidates: The PROMIS qualitative item review. Med Care. 2007;45(5 SUPPL. 1):S12–21. [PubMed: 17443114]
43. Upton P, Lawford J, Eiser C. Parent-child agreement across child health-related quality of life instruments: A review of the literature. Qual Life Res. 2008;17(6):895–913. [PubMed: 18521721]

44. Varni JW, Seid M, Kurtin PS. PedsQL™ 4.0: Reliability and Validity of the Pediatric Quality of Life Inventory™ Version 4.0 Generic Core Scales in Healthy and Patient Populations. Med Care. 2001;39(8):800–812. [PubMed: 11468499]
45. Varni JW, Limbers CA, Burwinkle TM. Parent proxy-report of their children's health-related quality of life: An analysis of 13,878 parents' reliability and validity across age subgroups using the PedsQL™ 4.0 Generic Core Scales. Health Qual Life Outcomes. 2007;5.
46. Varni JW, Burwinkle TM, Lane MM. Health-related quality of life measurement in pediatric clinical practice: An appraisal and precept for future research and application. Health Qual Life Outcomes. 2005;3.
47. Cremeens J, Eiser C, Blades M. Characteristics of health-related self-report measures for children aged three to eight years: a review of the literature. Qual Life Res. 2006;15(4):739–754. [PubMed: 16688506]
48. Russell KMW, Hudson M, Long A, et al. Assessment of health-related quality of life in children with cancer: Consistency and agreement between parent and child reports. Cancer. 2006;106(10):2267–2274. [PubMed: 16604563]
49. Varni JW, Limbers CA, Burwinkle TM. How Young Can Children Reliably and Validly Self-Report Their Health-Related Quality of Life?: An Analysis of 8,591 Children Across Age Subgroups With the PedsQL 4.0 Generic Core Scales. Health Qual Life Outcomes. 2007;5(1):1–13. [PubMed: 17201920]
50. Mauck MC, Smith J, Shupp JW, et al. Pain and Itch Outcome Trajectories Differ Among European American and African American Survivors of Major Thermal Burn Injury. Pain. 2017;158(11):2268–2276. [PubMed: 28796116]
51. Sood RF, Hocking AH, Muffley LA, et al. Race and Melanocortin 1 Receptor Polymorphism R163Q Are Associated with Post-BurnHypertrophic Scarring: A Prospective Cohort Study. J Invest Dermatol. 2015;135(10):2394–2401 [PubMed: 26030184]

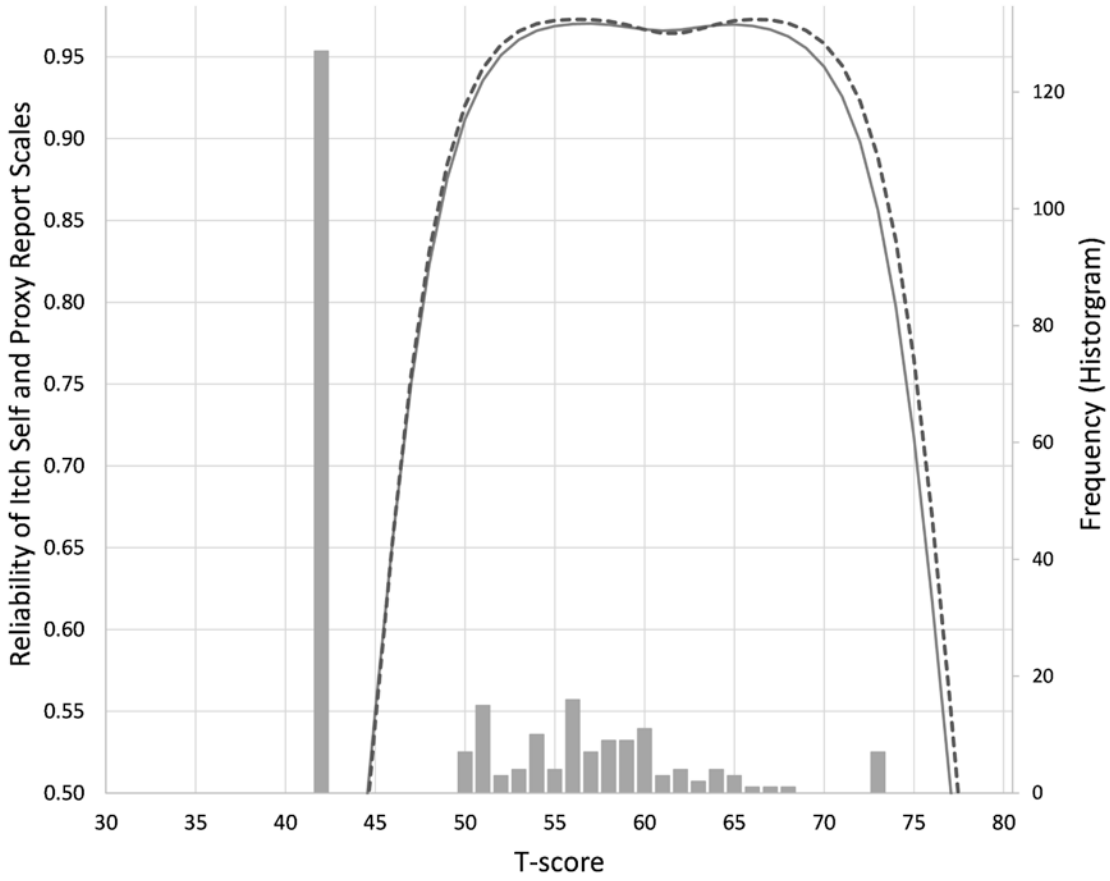

Figure 1.

Reliability of the BMS Itch Interference self- and proxy-report scales compared to the distribution of itch interference self-report T-scores in the scale calibration sample (n=247).

Table 1.

Itch interference scale self- and proxy report items, calibration statistics, and agreement statistics.

Self or Proxy Items	Item Name	Item Content	Discrimination (a)	Difficulty Range (b1 - b4)	Kappa[w] (SE)	Gwet's AC[w] (SE)
Self-Report Items	itch_1	I had trouble sleeping when I was itching	4.71	0.42 - 1.48	0.66 (0.05)	0.83 (0.03)
	itch_2	I felt angry when I was itching	4.71	0.52 - 1.78	0.66 (0.05)	0.88 (0.02)
	itch_3	I had trouble doing schoolwork when I was itching	4.71	0.58 - 1.55	0.63 (0.05)	0.87 (0.02)
	itch_4	It was hard for me to pay attention when I was itching	4.71	0.37 - 1.73	0.66 (0.05)	0.87 (0.02)
	itch_5	It was hard for me to have fun when I was itching	4.71	0.43 - 1.82	0.66 (0.05)	0.87 (0.02)
Proxy-Report Items	itchprxy 1	My child had trouble sleeping when he/she was itching	4.97	0.40 - 1.70	See above	
	itchprxy_2	My child felt angry when he/she was itching	4.97	0.46 - 1.86		
	itchprxy_3	My child had trouble doing schoolwork when he/she was itching	4.97	0.51 - 1.77		
	itchprxy_4	It was hard for my child to pay attention when he/she was itching	4.97	0.44 - 1.83		
	itchprxy_5	It was hard for my child to have fun when he/she was itching	4.97	0.38 - 1.69		

[w]Comparing parallel self- and proxy-report items (e.g. itch_1 and itchpxy_1) and weighted using quadratic weights.

Table 2.

Demographic and burn characteristics of the calibration sample

	Calibration Sample (N=264)
	mean ± SD n (%)
Age (years)	13.1 ± 2.9
Age Group of Pediatric Burn Participant	
8-11 years	99 (38%)
12-17 years	165 (62%)
Sex	
Female	85 (32.%)
Male	179 (68%)
Survey Language	
English	57 (22%)
Spanish	207 (78%)
Race/Ethnicity	
Non-Hispanic White	42 (16%)
Non-Hispanic Black	8 (3%)
Non-Hispanic Other Race	2 (1%)
Hispanic White or Other Race	210 (80%)
Burn Etiology	
Fire/Flame	195 (74%)
Scald	44 (17%)
Electricity	19 (7%)
Other	6 (2%)
Time Since Injury (years)	5.0 ± 4.6
% Total Body Surface Area (TBSA) burned	44.4 ± 18.2