Validation of PROMIS-29 domain scores among adult burn survivors: A National Institute on Disability, Independent Living, and Rehabilitation Research Burn Model System Study

Kara McMullen, MPH[1], Alyssa Bamer, MPH[1], Colleen M Ryan, MD[2,3], Jeffrey C Schneider, MD[4], Nicole Gibran, MD[5], Barclay T Stewart, MD, PhD, MPH[5,6], Tracy Mroz, PhD[1], Steven Wolf, MD[7], Dagmar Amtmann, PhD[1]

[1]Department of Rehabilitation Medicine, University of Washington, Seattle, WA

[2]Shriners Hospitals for Children – Boston, Boston, MA

[3]Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA

[4]Department of Physical Medicine and Rehabilitation, Spaulding Rehabilitation Hospital, Spaulding Research Institute, Harvard Medical School, Boston, MA

[5]Department of Surgery, University of Washington Harborview, Seattle, WA

[6]Harborview Injury Prevention and Research Center, Seattle, WA

[7]University of Texas Medical Branch, Galveston, TX

Abstract

Background: Patient-reported outcomes are important for understanding recovery after burn injury, benchmarking service delivery and measuring the impact of interventions. PROMIS-29 domains have been validated for use among diverse populations though not among burn survivors. The purpose of this study was to examine validity and reliability of PROMIS-29 scores in this population.

Methods: PROMIS-29 scores of physical function, anxiety, depression, fatigue, sleep disturbance, ability to participate in social roles, and pain interference were evaluated for validity and reliability in adult burn survivors. Unidimensionality, floor and ceiling effects, internal consistency, and reliability were examined. Differential item functioning (DIF) was used to examine bias with respect to demographic and injury characteristics. Correlations with measures of related constructs (Community Integration Questionnaire, Satisfaction with Life Scale, Post-Traumatic Stress Checklist-Civilian, and Veteran's Rand-12) and known-group differences were examined.

Results: 876 burn survivors with moderate to severe injury from 6 months-20 years post burn provided responses on PROMIS-29 domains. Participant ages ranged from 18–93 years

at time of assessment; mean years since injury was 3.4. All PROMIS domain scores showed high internal consistency (Cronbach's a=0.87–0.97). There was a large ceiling effect on ability to participate in social roles (39.7%) and physical function (43.3%). One-factor confirmatory factor analyses supported unidimensionality (all CFI >0.95). We found no statistically significant bias (DIF). Reliability was high (>0.9) across trait levels for all domains except sleep, which reached moderate reliability (>0.85). All known-group differences by demographic and clinical characteristics were in the hypothesized direction and magnitude except burn size categories.

Conclusions: The results provide strong evidence for reliability and validity of PROMIS-29 domain scores among adult burn survivors. Reliability of the extreme scores could be increased and the ceiling effects reduced by administering PROMIS-43, which includes 6 items per domain, or by administering by computerized adaptive testing.

Level of Evidence: This is a Level III psychometric analysis of prospectively collected survey data.

Keywords

PROMIS-29; validation; burn injury

BACKGROUND

More than 450,000 people seek treatment for burn injuries every year, of whom approximately 40,000 are hospitalized.[1] While many people with burn injuries recover fully, growing evidence has highlighted that for some survivors living with the sequelae of burns can be a chronic condition[2]. Long-term sequelae of burns can include pain, itch, anxiety, depression, contracture, amputation, poor body image, and limited physical and/or mental function,[3, 4] as well as difficulty with returning to work and social reintegration.[5] To better understand the short- and long-term impacts of burn injuries on patients' symptoms and overall health related quality of life, it is important to incorporate patient-reported outcome measures (PROMs) into clinical and research settings. PROMs are valuable when evaluating treatment effectiveness[6] and their use can improve patient-physician communication[7], care, and patient satisfaction.[8]

When considering PROMs for clinical practice or research, there is a need for valid, reliable, and efficient measures that assess multiple domains without undue burden on the respondent. While there are injury-specific PROMs, such as the Burn Specific Health Scale-Brief,[9] the Burn Outcomes Questionnaire,[10] and the Life Impact Burn Recovery Evaluation,[11] disease or condition specific measures limit the ability to compare outcomes among burn survivors to other populations. Alternatively, cross condition or general measures of universally applicable health constructs (i.e., constructs that could be experienced by all persons, either healthy or with chronic conditions), such as depression, anxiety, pain, sleep, physical function and social health, can facilitate comparisons across populations, and allow for researchers and clinicians to learn from patients' and survivors' experiences in other fields.

The Patient-Reported Outcomes Measurement Information System (PROMIS®), funded by the National Institutes of Health (NIH), developed measures that assess numerous health related quality of life (HRQOL) domains and are applicable to both healthy people and those living with acute and chronic conditions.[6] PROMIS measures were developed using modern psychometric methodology (e.g., item response theory), resulting in high reliability and validity, and can be administered via short forms or computer adaptive testing (CAT). PROMIS has profiles that include domains researchers determined were most relevant across multiple fields.[12] PROMIS-29 is the shortest of these profiles and contains seven 4-item short forms that assess the domains of physical function, anxiety, depression, fatigue, sleep disturbance, ability to participant in social roles, and pain interference, along with a pain intensity item.

PROMIS measures have been validated in many populations, including multiple cancers,[13, 14] kidney transplant,[15] and knee and shoulder arthroscopy.[16]. PROMIS measures, if found to be valid and reliable among burn survivors, could have broad applications in burn research and clinical practice. Because PROMIS-29 domains are relevant to understanding the full scope of burn recovery, and because they have not yet been validated in the burn population, the purpose of this study was to examine the psychometric properties of PROMIS-29 v2.1 domain scores in a sample of adult burn survivors. We hypothesize that the PROMIS-29 domain scores will function well in people with moderate to severe burn injury and that results will provide strong evidence of validity and reliability.

METHODS

Participants and procedures

All participants who responded to PROMIS-29 as a part of the Burn Model System (BMS) National Longitudinal Database (NLDB) research study were included in this study. The BMS was formed in 1994 and currently includes four burn centers that collect outcomes data on people with moderate to severe burn injuries.[17, 18] Consented participants complete surveys at discharge, 6-months, 12-months, 24-months and every 5 years post-injury either in-person at clinic visits, over the phone, via mailed paper and pencil surveys, or online; additionally, some data are abstracted from the medical record (e.g., burn size, etiology of injury, number of surgeries). Current BMS inclusion criteria include surgery for wound closure and at least one of the following: a burn to at least one critical area (e.g., face, hands, feet); an electrical, high voltage, or lightning injury; or percentage of total body surface area (TBSA) burned 10% for adults older than 65 years and 20% for people between the ages of 0 and 64 at the time of injury. Study procedures were approved by institutional review boards at all participating BMS institutions.

The current study utilized data from participants aged 18 years or older at the time of follow-up data collection with complete responses on at least one of the PROMIS domains. PROMIS-29 was added to data collection in 2015 and this study includes data from follow-up timepoints between 6-months and twenty years post-injury between May 2015 and July 2020. For participants with PROMIS data at multiple timepoints only the most complete earliest timepoint available (i.e., timepoint closest to burn injury) was utilized. Study data

were collected and managed using REDCap electronic data capture tools hosted at the University of Washington.[19, 20]

Measures

Demographic information included gender, age, and race and ethnicity. Injury characteristics included size of burn (% TBSA), burn etiology, and length of hospital stay, which were assessed at the time of hospital discharge using medical record abstraction. In addition to the PROMIS measures, symptom and HRQOL measures utilized for validity testing were the Veterans Rand 12,[21] Post-Traumatic Stress Disorder Checklist-Civilian,[22] the Modified 5-D itch,[23] and Community Integration Questionnaire,[24] and a single item asking participants if they had received psychological therapy or counseling.

PROMIS Measures—The published 4-item PROMIS-29 v2.1[12] adult profile was administered for all domains except ability to participate in social roles and activities. For the ability to participate in social roles domain, a custom 4-item short form (SF) was administered that included two items from the standard PROMIS-29 domain SF and two other items thought by BMS researchers to be more relevant to burn survivors. For the domains of anxiety and depression, the two additional items on the v1.0 6a short forms were also administered. Thus, for these two domains, both a four-item SF score and a six-item SF score were analyzed. Pain intensity was assessed with the standard PROMIS-29 one item average numeric pain rating scale (0–10) but was not included in the analyses designed to assess functioning of multi-item measures. All standard PROMIS SFs (i.e., all domains except ability to participate in social roles) were scored using summary score to T-score lookup tables provided in the PROMIS-29 or Anxiety or Depression user guides (see healthmeasures.net for user guides and scoring information). For the ability to participate in social roles, custom SF IRT parameters were used to generate a summary score to T-score scoring table using IRTScore.[25] Scores on all PROMIS domains are centered on the general U.S. population (Mean 50, SD 10), and higher scores indicate more of the trait being measured (i.e., higher scores mean better physical function and ability to participate in social roles but worse anxiety, depression, fatigue, sleep disturbance, and pain interference).[26]

Veteran's Rand 12 (VR-12)—The Veteran's Rand-12 (VR-12)[21] is a standardized, clinically validated global measure of general health that provides Physical Health Composite (PCS) and a Mental Health Composite (MCS) scores. PCS and MCS scores are centered on the U.S. population with a mean of 50 and standard deviation of 10. Lower PCS and MCS scores indicate worse health.[21]

Post-Traumatic Stress Disorder Checklist-Civilian (PCL-C)—The Post-Traumatic Stress Disorder Checklist-Civilian (PCL-C)[22] was developed to screen individuals for PTSD and monitor symptom change due to interventions. The scoring follows the DSM-IV criteria to identify probable PTSD and results in an indicator variable where 1=provisional/probable PTSD diagnosis and 0=no PTSD diagnosis.[22]

Modified 5-D Itch (4-D Itch)—The 5-D itch scale[23] has been used as an outcome measure for itch in clinical trials and includes the domains of itch duration, degree,

direction, disability, and distribution. Three domains (duration, degree, direction) are made up of one item each, and the disability domain includes four items that assess the impact of itching on daily activities. A slightly modified version of this scale was administered in this study and has been found to have acceptable psychometric properties in adult burn survivors.[27]

Community Integration Questionnaire (CIQ) Social Integration Subscale— The 6-item Social Integration Component (SIC) subscale of the Community Integration Questionnaire (CIQ)[24] was included in this study. The SIC focuses on activities outside the house, social settings, and social relationships.[24] This subscale has shown good to adequate internal consistency and is strongly correlated with other measures of social integration following burn injury.[28]

Analyses

Descriptive statistics of demographics and injury characteristics were calculated to describe the study sample using Stata 15.1.[29] Floor and ceiling effects were calculated as a percent of the sample with the lowest and the highest possible scores. These effects were considered present if more than 15% of respondents received the lowest or highest possible score, respectively.[30] Because not everyone experiences symptoms such as depression, anxiety or pain, it is expected that a proportion of respondents will endorse the lowest category on all the questions in the domain reflecting the absence of the symptom. This is not necessarily of concern. On the other hand, most everybody has some level of physical function or ability to participate in social roles. If more than 15% of people endorse the lowest or highest category on each question in the domain it may indicate that the questions do not measure well across the whole continuum of the domain (i.e., presence of floor or ceiling effect). All descriptive, validity, and reliability analyses were completed separately for each PROMIS domain and for both the 4-item and 6-item versions of depression and anxiety.

The unidimensionality of each set of questions in the same domain was examined in order to determine if there is support for a summary score. If a set of items are sufficiently unidimensional, they primarily measure one domain. This provides evidence that the summary score is meaningful and interpretable. Unidimensionality is also, an assumption of item response theory (IRT), and was examined for each domain by fitting a one-factor confirmatory factor analysis (CFA) with weighted least squares mean- and variance-adjusted estimation in Mplus software 8.2.[31] A comparative fit index (CFI) of 0.90 or higher was considered sufficient support for unidimensionality.[32, 33]

Reliability relates to the information in an item or in a measure; the more information and the less error is present in the score, the more reliable and accurate the score. Using item parameters in IRTPRO 4.2,[34] we extracted test information functions for all items and summary IRT-based scores. Test information functions were then converted to Classical Test Theory (CTT) reliability estimates and plotted along the T-score continuum. In CTT, reliability of 0.80 is considered sufficient for group comparisons while 0.90 is required for individual comparisons.[8, 35] We overlaid a histogram of scores from our adult burn survivors to show the frequency of each score in the sample. The percentage of the sample measured with good (0.8) and high (0.9) reliability was calculated for all domains. These

percentages were also calculated for the domains of pain interference, anxiety, depression, fatigue, and sleep disturbance using only the subsample of individuals who reported at least some symptoms (i.e., excluding those at the floor of the measure).

Reliability of each PROMIS domain score was also examined using CTT methods. Internal consistency was examined using Cronbach's alpha, with values between 0.7 and 0.9 considered acceptable.[36] Values greater than 0.9 may indicate item redundancy, while values less than 0.7 indicate poor correlations between the items in a domain.[36] Corrected item-total score correlations were calculated for each domain utilizing Spearman's rank order correlations to test for scale homogeneity. Correlation values of >0.40 were considered acceptable[35, 37] evidence of interitem reliability.

Differential item functioning (DIF) analysis is used to assess whether a score is biased due to demographic or clinical characteristics. Presence of meaningful DIF would result in 2 people with the same level of, for instance, depression, receiving different scores due to, for instance, their gender or age. This is referred to as bias. To assess level of DIF, IRT-based parameters generated for the same items in two different groups (e.g., men and women) were compared. If the item parameters differed significantly between the two groups then this suggests bias in the scores. For DIF analyses, a minimum of 200 subjects per group has been recommended.[38] In this study we examined DIF with respect to age (19–34, 35–54, or 55+ years), sex, education (high school graduate or above, no high school diploma), race (White, non-White), ethnicity (Hispanic, non-Hispanic) and total body surface area burned (0–19%, 20%). DIF analyses were conducted using ordinal logistic regression models with the *lordif* (Choi et al 2011) program in R (R Core Team 2016). The PROMIS-recommended criterion of McFadden's pseudo R^2-change of 2% was used to identify items with statistically significant DIF.[39] Items identified as having statistically significant DIF were further analyzed by comparing DIF adjusted and non-adjusted short form scores. While some items may have statistically significant DIF, the impact of that DIF on actual scores may not be clinically meaningful. Only DIF-adjusted scores that differed by >2 points on the T-score metric compared to unadjusted scores were considered to have clinically meaningful DIF.

Convergent validity was examined by calculating Spearman's rank correlations between PROMIS scores and other PRO measures. For each PROMIS domain, correlations with other available measures were assessed and the direction and strength of correlations were hypothesized based on a literature review. Table 1 details the hypothesized magnitude and direction of the correlations and supporting literature. For some domains, correlations were hypothesized based on similarity of measurement constructs and authors' clinical judgment. Correlation coefficients of 0.9–1.0 indicate very strong, 0.7–0.89 strong, 0.5–0.69 moderate, and 0.3–0.49 weak relationships.[40]

Construct validity was examined using known-groups analysis. Known-groups differences in PROMIS scores by pre-defined socio-demographic and clinical groups were examined using Wilcoxon-Mann-Whitney tests for comparisons of two groups and ANOVA for comparisons of more than two groups. These groups were selected based on a literature review and, when not available, authors' clinical judgement. Known-groups hypotheses and literature to support those hypotheses can be found in Table 1.

RESULTS

Participants

A total of 876 BMS participants completed at least one of the five PROMIS domains during the study. The mean age of participants at time of burn injury was 41.2 years (0.7–91.0 years) and at time of data collection was 44.6 years (18.0–93.2 years). The majority of respondents were male (68%) and White (69%). See Table 2 for additional sample characteristics. The average time since injury was 3.4 years and the percentage of data utilized from each follow-up timepoint was: 43% at 6 months, 14% at 12 months, 16% at 24 months, 12% at 5 years, 5% at 10 years, 7% at 15 years, and 3% at 20 years post-burn injury.

Measure performance

Floor and ceiling effects and unidimensionality—Average scores across all PROMIS domains were within 3 points of the general population mean except for ability to participate in social roles, which was slightly higher at 53.5 (Table 3). Substantial portions of the sample reported no symptoms (i.e., they were at the floor of the measure) on anxiety (42.6%), depression (50.9%), fatigue (26.3%), and pain (47.7%). Only 8.9% of the population reported no sleep disturbance. There was a large ceiling effect on social roles (39.7%) and physical function (43.3%). The results of the CFAs supported the IRT assumption of unidimensionality and the use of a summary score across each of the PROMIS domains and SFs (all CFI >0.95).

Reliability—Reliability was high (0.9) across 2–3 SDs of the T-score continuum for all domains but sleep, which was high only between 48 and 65 (Figure 1 and Table 3). Reliability was good (0.8) within +/− 2 SDs of the T-score continuum for all domains. When examining only individuals who reported some level of symptoms, 96% (sleep disturbance) to 100% (6-item anxiety, 4-item depression, fatigue, and pain) of the sample was measured with good reliability and 60% (sleep disturbance) to 97% (4-item depression) of the sample was measured with high reliability. For depression and anxiety, reliability was slightly higher for the 6-item compared to the 4-item short forms, though the difference is small, and the 4-item forms have high reliability at and above the mean (see Figure 1).

Internal consistency—All PROMIS domains showed high internal consistency (Cronbach's a=0.87–0.97). Cronbach's a was above 0.9 for physical function, anxiety 6-item SF, depression 4- and 6-item SFs, fatigue, ability to participate in social roles, and pain interference. Corrected item-total score correlations were above the recommended 0.4 (range: 0.64–0.95) (Table 4).

Differential Item Functioning—For all DIF groups except race (White n=602, non-White n=142), there were at least 200 people in each sub-group. DIF analyses did not identify any items on any domains with statistically significant DIF by age, sex, race, ethnicity, education level, or burn size.

Validity—All correlations were in the direction hypothesized, with a large majority of the magnitude hypothesized as well (see Table 1). Several correlations were found to be weaker

than hypothesized, including sleep disturbance and VR-12 PCS (0.34), ability to participate in social roles and the CIQ Social Integration Subscale (0.39), sleep disturbance and PROMIS depression (0.44), and sleep disturbance and VR-12 PCS (0.47). Results supported hypothesized known-group differences by demographic and clinical characteristics except for the hypothesized differences between burn size categories, which were not significantly associated with physical function, fatigue, or pain interference (see Table 1).

DISCUSSION

The findings provide evidence of validity and reliability of the PROMIS-29 profile scores in adult burn survivors. Results supported sufficient unidimensionality, internal consistency, reliability, and validity in this population. Therefore, PROMIS-29 can potentially be used to understand recovery after injury, benchmark the effects of burn care service delivery and measure the impact of care and recovery interventions.

This study found evidence of floor and ceiling effects in burn survivors, which has also been noted in other PROMIS validation projects.[13–15] The large proportion of people who do not report any depression, anxiety, or pain likely reflects lack of symptoms and is consistent with the distribution of these domains in the general population. However, the ceiling effects on physical function and ability to participate in social roles may be of concern depending on the purpose of a study. If the purpose is to identify those with low physical function and examine the relationship between low physical function and other long-term outcomes (e.g., obesity), the lack of discrimination at higher levels of physical function is not problematic. However, if the purpose is the describe the level of physical function across the whole continuum of physical function of people with burn injuries, the lack of discrimination at the higher levels is a limitation. In order to better describe people with higher physical function and ability to participate in social roles, the BMS NLDB could include items that better discriminate at higher levels of physical function and ability to participate in social roles. We also suggest that PROMIS revise the PROMIS-29 physical function and ability to participate in social roles to include an item that requires higher levels of physical function and higher levels of participation in social roles to provide scores that are more reliable at higher end of these continuums. Sometimes disease specific instruments can be used to supplement generic instruments, but typically disease specific instruments are helpful in addressing floor effects and not very helpful in addressing ceiling effects, so we don't recommend this approach.

The use of a single IRT-based score for each PROMIS domain was clearly supported. All PROMIS domain scores measured with high reliability for a significant section of the T-score range and therefore are good choices when trying to balance the need for reliable scores with low participant burden. However, the Cronbach's a was above 0.9 on six domains, suggesting potential redundancy among items in each domain. The PROMIS system makes the customization of short forms, including selecting items of particular interest and addressing issues such as redundancy, feasible in both research and clinical settings.[41] However, when deciding between shorter measures and better functioning measures (i.e., higher reliability and fewer floor or ceiling effects), length of administration

and therefore participant burden is an important consideration that needs to be balanced against the performance of the domain.

The lack of statistically significant DIF across demographic and clinical characteristics indicates that the PROMIS domain scores are unbiased by age, gender, education levels and burn size. However, because of the small sample size of non-White people in the race group comparisons, lack of DIF due to race may potentially be due to inadequate power to detect DIF. Further analyses with adequate sample size of non-White participants are required to assess DIF by race.

Convergent validity and known-group analyses supported the validity of the PROMIS-29 domain scores in burn survivors. The lack of significant differences in burn size groups may reflect the fact that burn size is not a good representation of injury severity in and of itself or that burn size alone is not predictive of outcomes.[42, 43] Other factors, such as length of hospital stay, inhalation injury,[44] need for excision and grafting, and burn depth and distribution contribute to injury severity in addition to burn size. Burn size itself may not be the main driver of long-term quality of life outcomes after injury.[45]

One limitation of this study is that results may not be generalizable to people with less severe burn injuries; more studies to evaluate validity and reliability of PROMIS-29 domain scores in people with such injuries would be useful. Additionally, because we only used cross-sectional data, we did not analyze responsiveness to change. Qualitative research (e.g., interviews and/or focus groups) with burn survivors would be useful to examine if PROMIS-29 domains and items should be supplemented with any areas of particular importance to burn survivors. We recommend that researchers investigate whether any important burn-relevant issues are missing from the PROMIS-29 item banks and add relevant items to the item banks and the short forms to ensure that they provide scores meaningful to burn survivors, clinicians, and researchers. In addition, though this study was conducted as part of an ongoing longitudinal research study rather than in a clinical setting, the results suggest that the scores are valid in people with moderate to severe burns, which is an important first step in the clinical adoption of these measures.

Previous studies indicate PROMIS measures have clinical utility; the profile can be used to help clinicians identify burn survivors who may need additional interventions, particularly since most of the domains have established clinically meaningful differences.[5, 46] However, additional studies in clinical settings need to be conducted to determine the feasibility and utility of incorporating them into burn clinical care settings. PROMIS data are currently being collected in at least one burn care clinical setting we are aware of; once more data are available we will evaluate the utility of that data with regards to improving clinical care. Furthermore, because the measures enable cross-population comparisons, learning from one discipline can improve burn care; for example, management of chronic pain in other conditions (such as muscular skeletal pain conditions) can inform prevention and treatment of pain in burn survivors. In other fields, such as cancer care, PROMIS domains have been implemented directly into the electronic health record for clinicians to receive alerts when patient responses indicated elevated symptoms.[47, 48] Advances such as these in the burn care and rehabilitation field could increase patient and provider communication

and aid in identification of emergent issues that can be addressed during routine patient clinic visits.

Conclusion

The results of this study strongly support evidence of validity and reliability the PROMIS-29 profile domain scores in an adult burn population. These measures can be used with confidence by researchers evaluating outcomes after burn and clinicians working to screen and identify individuals who might need additional services. The SFs included in this study are efficient to administer and reduce administration burden compared to lengthier legacy measures. One of the important advantages of PROMIS measures is the ability to compare scores across clinical samples and populations. Because the PROMIS measures are used widely (a Google Scholar search of PROMIS-29 had 1,100 results), their adoption in both research and clinical settings could increase knowledge about the long-term impacts of burn injuries.

External Funding:

References

1. American Burn Association. Burn Incidence Fact Sheet. 2016.

2. Kelter BM, et al. , Recognizing the long-term sequelae of burns as a chronic medical condition. Burns, 2020. 46(2): p. 493–496. [PubMed: 31711801]

3. Carrougher GJ, et al. , Pruritus in adult burn survivors: postburn prevalence and risk factors associated with increased intensity. J Burn Care Res, 2013. 34(1): p. 94–101. [PubMed: 23079565]

4. Esselman PC, et al. , Burn rehabilitation: state of the science. Am J Phys Med Rehabil, 2006. 85(4): p. 383–413. [PubMed: 16554686]

5. Chen JH, et al. , Patient and social characteristics contributing to disparities in outcomes after burn injury: application of database research to minority health in the burn population. Am J Surg, 2018. 216(5): p. 863–868. [PubMed: 29366485]

6. Cella D, 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): p. 1179–94. [PubMed: 20685078]

7. Detmar SB, et al. , Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. JAMA, 2002. 288(23): p. 3027–34. [PubMed: 12479768]

8. Hahn EA, et al. , Precision of health-related quality-of-life data compared with other clinical measures. Mayo Clin Proc, 2007. 82(10): p. 1244–54. [PubMed: 17908530]

9. Kildal M, et al. , Development of a brief version of the Burn Specific Health Scale (BSHS-B). J Trauma, 2001. 51(4): p. 740–6. [PubMed: 11586169]

10. Ryan CM, et al. , Benchmarks for multidimensional recovery after burn injury in young adults: the development, validation, and testing of the American Burn Association/Shriners Hospitals for Children young adult burn outcome questionnaire. J Burn Care Res, 2013. 34(3): p. e121–42. [PubMed: 23511284]

11. Kazis LE, et al. , Development of the life impact burn recovery evaluation (LIBRE) profile: assessing burn survivors' social participation. Qual Life Res, 2017. 26(10): p. 2851–2866. [PubMed: 28493205]

12. Cella D, et al. , PROMIS((R)) Adult Health Profiles: Efficient Short-Form Measures of Seven Health Domains. Value Health, 2019. 22(5): p. 537–544. [PubMed: 31104731]

13. Jensen RE, et al. , Validation of the PROMIS physical function measures in a diverse US population-based cohort of cancer patients. Qual Life Res, 2015. 24(10): p. 2333–44. [PubMed: 25935353]

14. Quach CW, et al. , Reliability and validity of PROMIS measures administered by telephone interview in a longitudinal localized prostate cancer study. Qual Life Res, 2016. 25(11): p. 2811–2823. [PubMed: 27240448]

15. Tang E, et al. , Validation of the Patient-Reported Outcomes Measurement Information System (PROMIS)-57 and −29 item short forms among kidney transplant recipients. Qual Life Res, 2019. 28(3): p. 815–827. [PubMed: 30467780]

16. Giordano NA, et al. , A Longitudinal Comparison of Patient-Reported Outcomes Measurement Information System to Legacy Scales in Knee and Shoulder Arthroscopy Patients. Arthroscopy, 2021. 37(1): p. 185–194 e2. [PubMed: 32721547]

17. Amtmann D, et al. , National Institute on Disability, Independent Living, and Rehabilitation Research Burn Model System: Review of Program and Database. Arch Phys Med Rehabil, 2020. 101(1S): p. S5–S15. [PubMed: 28989076]

18. Goverman J, 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): p. e240–e253. [PubMed: 27294859]

19. Harris PA, et al. , The REDCap consortium: Building an international community of software platform partners. J Biomed Inform, 2019. 95: p. 103208. [PubMed: 31078660]

20. Harris PA, et al. , Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform, 2009. 42(2): p. 377–81. [PubMed: 18929686]

21. Selim AJ, et al. , Updated U.S. population standard for the Veterans RAND 12-item Health Survey (VR-12). Qual Life Res, 2009. 18(1): p. 43–52. [PubMed: 19051059]

22. Blanchard EB, et al. , Psychometric properties of the PTSD Checklist (PCL). Behaviour research and therapy, 1996. 34(8): p. 669–673. [PubMed: 8870294]

23. Elman S, et al. , The 5-D itch scale: a new measure of pruritus. Br J Dermatol, 2010. 162(3): p. 587–93. [PubMed: 19995367]

24. Willer B, Ottenbacher KJ, and Coad ML, The community integration questionnaire. A comparative examination. Am J Phys Med Rehabil, 1994. 73(2): p. 103–11. [PubMed: 8148099]

25. Flora D and Thissen D. User's Guide for IRTSCORE: Item Response Theory Score Approximation Software. Available from: http://www.unc.edu/depts/psychology/dthissen/840F14/IRTScore.pdf.

26. Measures H Interpret PROMIS® Scores. 2020 November 8, 2020]; Available from: https://www.healthmeasures.net/score-and-interpret/interpret-scores/promis..

27. Amtmann D, 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): p. e402–e408. [PubMed: 27388881]

28. Gerrard P, et al. , Validation of the Community Integration Questionnaire in the adult burn injury population. Qual Life Res, 2015. 24(11): p. 2651–5. [PubMed: 25986908]

29. StataCorp, Stata Statistical Software: Release 14. 2015, StataCorp LP: College Station, TX.

30. Terwee CB, et al. , Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol, 2007. 60(1): p. 34–42. [PubMed: 17161752]

31. Muthén LK and Muthén BO, Mplus User's Guide. Seventh Edition. 1998–2012, Muthén & Muthén: Los Angeles, CA.

32. Hu LT and Bentler PM, Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Model, 1999. 6: p. 1–55.

33. Reise SP, A comparison of item-and person-fit methods of assessing model-data fit in IRT. Applied Psychological Measurement, 1990. 14(2): p. 127–137.

34. Cai L, Thissen D, and duToit S, IRTPRO for Windows. Version 4.2. 2015, Scientific Software International: Skokie, IL.

35. Nunnally JC and Bernstein IH, Psychometric theory. 3rd ed. 1994, New York, NY: McGraw-Hill, Inc.

36. Streiner DL and Norman GR, Health Measurement Scales: a practical guide to their development and use. 3rd ed. 2002, Oxford: Oxford Medical Publications.

37. Gliem JA and Gliem RR. Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. 2003. Midwest Research-to-Practice Conference in Adult, Continuing, and Community ….

38. Scott NW, et al. , A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. J Clin Epidemiol, 2009. 62(3): p. 288–95. [PubMed: 18774693]

39. Health Measures. PROMIS® Minimum requirements for the release of PROMIS instruments after translation and recommendations for further psychometric evaluation. 2014.

40. Schober P, Boer C, and Schwarte LA, Correlation Coefficients: Appropriate Use and Interpretation. Anesth Analg, 2018. 126(5): p. 1763–1768. [PubMed: 29481436]

41. Broderick JE, et al. , Advances in Patient-Reported Outcomes: The NIH PROMIS((R)) Measures. EGEMS (Wash DC), 2013. 1(1): p. 1015. [PubMed: 25848562]

42. Ryan CM, et al. , The Impact of Burn Size on Community Participation: A Life Impact Burn Recovery Evaluation (LIBRE) Study. Ann Surg, 2020. Publish Ahead of Print.

43. Schneider JC, et al. , The long-term impact of physical and emotional trauma: the station nightclub fire. PLoS One, 2012. 7(10): p. e47339. [PubMed: 23077593]

44. Stockly OR, et al. , Inhalation injury is associated with long-term employment outcomes in the burn population: Findings from a cross-sectional examination of the Burn Model System National Database. PLoS One, 2020. 15(9): p. e0239556. [PubMed: 32966317]

45. Ryan CM, et al. , Recovery trajectories after burn injury in young adults: does burn size matter? J Burn Care Res, 2015. 36(1): p. 118–29. [PubMed: 25501787]

46. Cook KF, et al. , PROMIS measures of pain, fatigue, negative affect, physical function, and social function demonstrated clinical validity across a range of chronic conditions. J Clin Epidemiol, 2016. 73: p. 89–102. [PubMed: 26952842]

47. Papuga MO, et al. , Large-scale clinical implementation of PROMIS computer adaptive testing with direct incorporation into the electronic medical record. Health Syst (Basingstoke), 2018. 7(1): p. 1–12. [PubMed: 31214335]

48. Wagner LI, et al. , Bringing PROMIS to practice: brief and precise symptom screening in ambulatory cancer care. Cancer, 2015. 121(6): p. 927–34. [PubMed: 25376427]

49. Vaishnav AS, et al. , Correlation between NDI, PROMIS and SF-12 in cervical spine surgery. Spine J, 2020. 20(3): p. 409–416. [PubMed: 31678044]

50. Hoch JM, et al. , The Relationship Among 3 Generic Patient-Reported Outcome Instruments in Patients With Lower Extremity Health Conditions. J Athl Train, 2019. 54(5): p. 550–555. [PubMed: 31084504]

51. Druery M, Brown TL, and Muller M, Long term functional outcomes and quality of life following severe burn injury. Burns, 2005. 31(6): p. 692–5. [PubMed: 16129223]

52. Jacobson NC and Newman MG, Anxiety and depression as bidirectional risk factors for one another: A meta-analysis of longitudinal studies. Psychol Bull, 2017. 143(11): p. 1155–1200. [PubMed: 28805400]

53. Pilkonis PA, et al. , Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS(R)): depression, anxiety, and anger. Assessment, 2011. 18(3): p. 263–83. [PubMed: 21697139]

54. Kroenke K, Baye F, and Lourens SG, Comparative Responsiveness and Minimally Important Difference of Common Anxiety Measures. Med Care, 2019. 57(11): p. 890–897. [PubMed: 31415337]

55. Sarwer DB, et al. , Psychiatric diagnoses and psychiatric treatment among bariatric surgery candidates. Obes Surg, 2004. 14(9): p. 1148–56. [PubMed: 15527626]

56. Simko LC, et al. , Fatigue Following Burn Injury: A Burn Model System National Database Study. J Burn Care Res, 2018. 39(3): p. 450–456. [PubMed: 28877130]

57. Mauck MC, et al. , Chronic Pain and Itch are Common, Morbid Sequelae Among Individuals Who Receive Tissue Autograft After Major Thermal Burn Injury. Clin J Pain, 2017. 33(7): p. 627–634. [PubMed: 28145911]

58. Miro J, et al. , Defining mild, moderate, and severe pain in young people with physical disabilities. Disabil Rehabil, 2017. 39(11): p. 1131–1135. [PubMed: 27291566]

59. Prasad A, et al. , The association of patient and burn characteristics with itching and pain severity. Burns, 2019. 45(2): p. 348–353. [PubMed: 30686696]

60. McCallum SM, et al. , Associations of fatigue and sleep disturbance with nine common mental disorders. J Psychosom Res, 2019. 123: p. 109727. [PubMed: 31376877]

61. Frech T, et al. , Prevalence and correlates of sleep disturbance in systemic sclerosis--results from the UCLA scleroderma quality of life study. Rheumatology (Oxford), 2011. 50(7): p. 1280–7. [PubMed: 21324979]
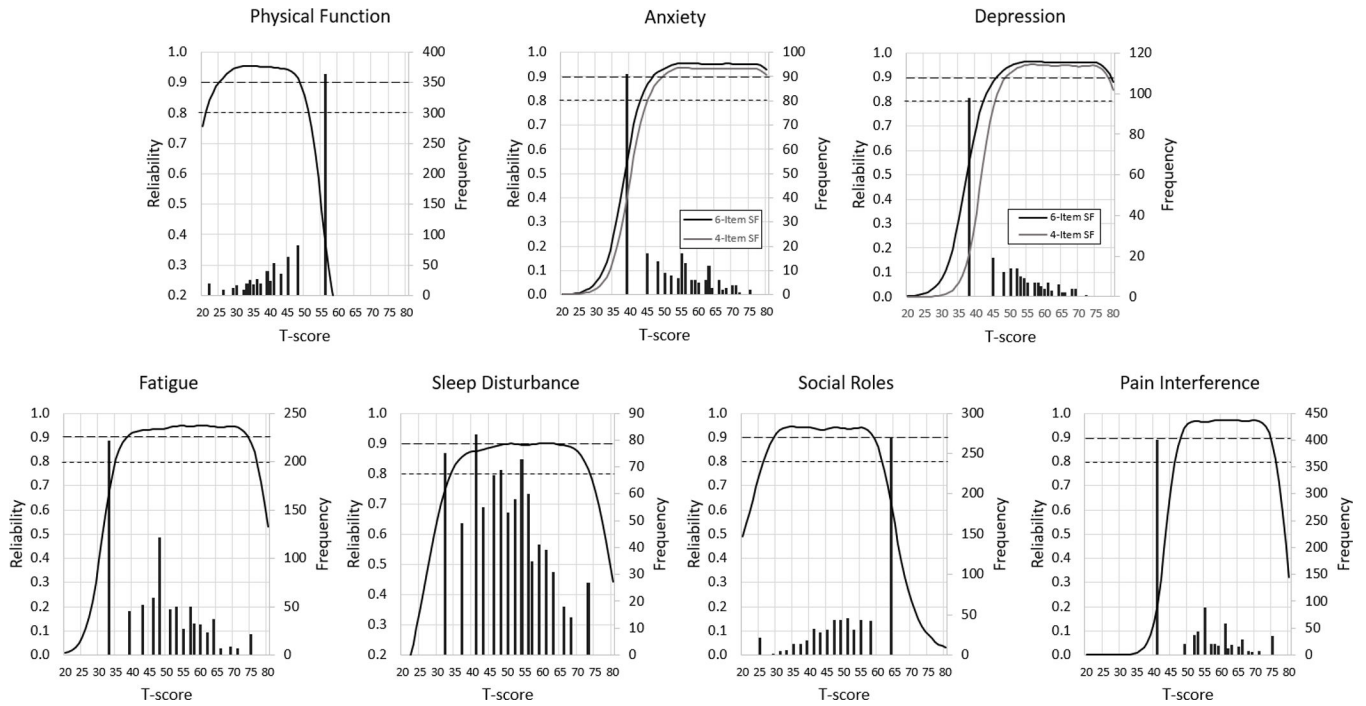
Figure 1.
Reliability (on the left y-axis; referred to in IRT framework as "information") and histogram of the sample PROMIS scores on the T- metric. The histogram represents the frequency of each score in the study sample. Solid lines show reliability of the PROMIS scale 4-item short-form and 6-item short-form where assessed (depression and anxiety) along the T-scale metric (mean =50, SD=10). Scores with reliability above 0.9 are considered highly reliable and adequate for individual comparisons and above 0.8 for group comparisons; these reliability levels are indicated by dotted lines. A T-score of 50 is the mean of the general population sample.

Table 1.

Convergent Validity Correlations and Known-Group Analyses Hypotheses and Results

Domain		Hypothesis	Results	Supporting literature
Physical Function	Correlation	Correlated with VR-12 PCS >0.5	Spearman's rho: 0.76	Vaishnav et al 2019[49] Hoch et al 2019 [50]
	Known-group analysis	People with TBSA >20% will have lower physical function scores than people with TBSA =<10%	People with TBSA =<10% (n=349) had mean PF score of 47.3; TBSA>10% & =<20% (n=119) had mean PF score of 47.3; TBSA>20% (n=372) had mean PF score of 47.0, $F_{(2,839)}$=0.14, p=0.87	Ryan et al 2015[45] Druery et al 2005[51]
		People ages 55+ will have lower physical function scores than people ages <34	People ages 55+ had mean PF score of 44.4 (n=260); ages 35–54 (n=281) had mean PF score of 46.7; ages <35 (n=299) had PF score of 50.0; $F_{(2,837)}$=24.35, p<0.0001	Tang et al 2019[15]
Anxiety	Correlation	Correlated with VR12 MCS >0.5	Spearman's rho: −0.69	Giordano et al 2021[16]
		Correlated with PROMIS Depression >0.7	Spearman's rho: 0.77	Jacobson et al 2017[52] Pilkonis et al 2011[53]
	Known-group analysis	People with PTSD categorized as yes on PCLC will have higher anxiety scores than those with no PTSD on PCLC	People with PTSD categorized as yes on PCLC (n=103) had mean anxiety score of 63.5; those with no PTSD on PCLC (n=702) had mean anxiety score of 48.1, p<0.0001	Kroenke et al 2019[54]
		Females will have higher anxiety scores than males	Females (n=274) had mean anxiety score of 52.7; males (n=569) had mean anxiety score of 49.0, p<0.0001	Tang et al 2019[15]
Depression	Correlation	Correlated with VR12 MCS >0.5	Spearman's rho: −0.73	Giordano et al 2021[16]
		Correlated with PROMIS Anxiety >0.7	Spearman's rho: 0.77	Jacobson et al 2017[52]
	Known-group analysis	People who received psychological therapy or counseling (binary variable) will have higher depression scores than those who did not	People who received psychological therapy or counseling (n=137) had mean depression score of 54.8; people who did not (n=692) had mean depression score of 48.1, p<0.0001	Sarwer et al 2004[55]
		Females will have higher depression scores than males	Females (n=274) had mean depression score of 51.6; males (n=567) had mean depression score of 48.0, p<0.0001	Tang et al 2019[15]
Fatigue	Correlation	Correlated with VR12 Vitality Scale moderately	Spearman's rho: 0.65	Simko et al 2018[56]
	Known-group analysis	People with TBSA >20% will have higher fatigue scores than people with TBSA =<10%	People with TBSA =<10% (n=349) had mean fatigue score of 47.8; TBSA>10% & =<20% (n=118) had mean fatigue score of 48.3;	Simko et al 2018[56]

Domain	Hypothesis		Results	Supporting literature
Pain interference	Correlation		TBSA>20% (n=377) had mean fatigue score of 47.0, $F_{(2,841)}=0.08$, p=0.92	Mauck et al 2017[57]
	Correlated to 5-D itch moderately		Spearman's rho: 0.43	
	Known-group analysis	People with lower pain intensity will have lower pain interference	People with pain 0–4 on 0–10 scale (n=605) had mean PI score of 47.1; people with pain 5–6 (n=114) had mean PI score of 58.4; people with pain 7–10 on 0–10 scale had mean PI score of 64.3; $F_{(2,837)}=286.11$, p<0.0001	Miro et al 2017[58]
		People with TBSA >20% will have higher pain scores than people with TBSA =<10	People with TBSA =<10% (n=344) had mean PI score of 50.6; TBSA>10% & =<20% (n=118) had mean PI score of 50.4; TBSA>20% (n=378) had mean PI score of 51.8, $F_{(2,837)}=1.52$, p=0.22	Prasad et al 2019[59]
Ability to Participate in Social Roles	Correlation	Correlated with CIQ moderately	Spearman's rho: 0.35	
		Correlated with VR12 social function subscale moderately	Spearman's rho: 0.65	
	Known-group analysis	No known-group hypothesis made	N/A	N/A
Sleep Disturbance	Correlation	Correlated with VR12 MCS moderately	Spearman's rho: −0.49	McCallum et al 2019[60]
		Correlated with VR12 PCS moderately	Spearman's rho: −0.32	
		Correlated with PROMIS depression moderately	Spearman's rho: 0.45	McCallum et al 2019[60] Frech et al 2011[61]
	Known-group analysis	Females will have higher sleep disturbance scores than males	Females had mean sleep score of 51.5 (n=273); males (n=573) had mean sleep score of 49.6, p=0.016	Tang et al 2019[15]

Table 2.

Description of Sample from the Burn Model System National Longitudinal Database

	Mean (n)	SD	Range
Age at time of burn injury	41.2 (876)	19.3	0.7–91.0
Age at time of PROMIS data collection	44.6 (876)	17.5	18.0–93.2
Length of hospital stay	30.3 (876)	35.1	0–389
TBSA burn	23.9 (870)	23.3	0.1–95.0
	%		n
Sex			
Male	67.8		594
Female	32.2		282
Race			
Black or African-American	9.7		85
Asian	1.4		12
White	68.7		602
American Indian/Alaskan Native	1.4		12
Native Hawaiian or other Pacific Islander	0.6		5
More than one race	0.5		4
Other	2.7		24
Missing/Unknown	15.1		132
Ethnicity			
Hispanic/Latinx	25.9		227
Not Hispanic or Latinx	71.4		625
Missing/Unknown	2.7		24
Etiology of Injury			
Fire/flame	57.0		496
Scald	13.3		116
Grease	10.6		92
Electricity	7.4		64
Other	11.7		105
Timepoint PROMIS measured (post-injury)			
6 months	43.4		380
12 months	14.td4		126
24 months	15.8		138
5 years	12.0		105
10 years	4.9		43
15 years	6.5		57

	Mean (n)	SD	Range
20 years	3.1		27

Table 3.

Summary scores for each PROMIS domain and range of scores with good or high reliability.

	Mean (n)	Median	SD	Score Range	Score Range with Reliability ≥0.8	Score Range with Reliability ≥0.9	% of Entire Sample Measured with Reliability ≥0.8	% of Entire Sample Measured with Reliability ≥0.9	% of Symptomatic Individuals Measured with Reliability ≥0.8	% of Symptomatic Individuals Measured with Reliability ≥0.9
Physical Function	47.2 (840)	48.0	9.9	22.9–56.9	21–51	25–48	57%	54%	N/A	N/A
Anxiety 4-Item	50.2 (843)	48.0	10.3	40.3–81.6	45–80	50–80	55%	48%	99%	83%
Anxiety 6-Item	50.1 (236)	48.8	10.5	39.1–75.6	43–80	47–80	61%	54%	100%	88%
Depression 4-Item	49.2 (841)	41.0	9.7	41.0–79.4	46–80	49–78	49%	48%	100%	97%
Depression 6-Item	48.2 (236)	48.3	9.9	38.4–80.3	42–80	46–79	58%	50%	99%	86%
Fatigue	48.0 (844)	48.6	11.4	33.7–75.8	35–76	38–74	74%	71%	100%	89%
Sleep Disturbance	50.2 (846)	50.5	10.2	32.0–73.3	34–73	48–65	88%	54%	96%	60%
Ability to Participate in Social Roles	53.5 (688)	55.5	10.7	25.7–64.0	26–61	29–58	61%	52%	N/A	N/A
Pain Interference	51.1 (840)	49.6	10.4	41.6–75.6	47–76	48–74	52%	48%	100%	92%
Pain Intensity	2.9 (873)	2.0	2.8	0–10	N/A	N/A	N/A	N/A	N/A	N/A

Table 4.

Cronbach's Alpha and corrected Item-to-total score correlations for each domain

	Cronbach's alpha	Correlation coefficient
Physical Function	**0.9113**	
Are you able to do chores such as vacuuming or yard work?		0.7319
Are you able to go up and down stairs at a normal pace?		0.8357
Are you able to go for a walk of at least 15 minutes?		0.8074
Are you able to run errands and shop?		0.8248
Anxiety 4-item	0.8990	
I felt fearful		0.7304
I found it hard to focus on anything other than my anxiety		0.8003
My worries overwhelmed me		0.8241
I felt uneasy		0.7673
Anxiety 6-item[*]	0.9304	
I felt nervous		0.8424
I felt like I needed help for my anxiety		0.7873
Depression 4-item	0.9356	
I felt worthless		0.8648
I felt helpless		0.8545
I felt depressed		0.8174
I felt hopeless		0.8785
Depression 6-item[*]	0.9510	
I felt like a failure		0.8509
I felt unhappy		0.8393
Fatigue	0.9472	
I feel fatigued		0.8765
I have trouble starting things because I am tired		0.8390
How run-down did you feel on average?		0.8816
How fatigued were you on average?		0.9022
Sleep Disturbance	0.8714	
My sleep quality was		0.7455
My sleep was refreshing		0.6371
I had a problem with my sleep		0.8089
I had difficulty falling asleep		0.7202
Ability to Participate in Social Roles and Activities	0.9000	
I have trouble doing all of my regular leisure activities with others		0.8005
I have trouble keeping up with my family responsibilities		0.8200
I have trouble doing all of my usual work (include work at home)		0.7788

	Cronbach's alpha	Correlation coefficient
Physical Function	**0.9113**	
I have trouble keeping in touch with others		0.6859
Pain Interference	0.9728	
How much did pain interfere with your day to day activities?		0.9197
How much did pain interfere with work around the home?		0.9552
How much did pain interfere with your ability to participate in social activities?		0.9159
How much did pain interfere with your household chores?		0.9434

*
The 6-item scales include the 4 items in the 4-item short form plus the additional items listed