Psychometric evaluation of the NIH Patient-Reported Outcomes Measurement Information System (PROMIS®) in a multi-racial, multi-ethnic systemic lupus erythematosus (SLE) cohort

Patricia Katz, PhD[1], **Jinoos Yazdany, MD, MPH**[1], **Laura Trupin, MPH**[1], **Stephanie Rush, BA**[1], **Charles G. Helmick, MD**[2], **DR. Louise B. Murphy, PhD**[2], **DR. Cristina Lanata, MD**[1], **Lindsey A. Criswell, MD, MPH, DSc**[1], **Maria Dall'Era, MD**[1]

[1]University of California San Francisco

[2]Centers for Disease Control and Prevention

Abstract

Background: We examined psychometric performance of PROMIS measures in a racially/ethnically, linguistically diverse SLE cohort.

Methods: Data were from the California Lupus Epidemiology Study (CLUES), a multi-racial/ethnic cohort of individuals with physician-confirmed SLE. The majority (n=332) attended in-person research visits that included interviews conducted in English, Spanish, Cantonese, or Mandarin. Up to 12 PROMIS short-forms were administered (depending on language availability). An additional 99 completed the interview by phone only. Internal consistency was examined with Cronbach's alpha and item-total correlations. Correlations with SF-36 subscales and both self-reported and physician-assessed disease activity assessed convergent validity. All analyses were repeated within each racial/ethnic group. Differences in scores by race/ethnicity were examined in bivariate analyses and by multiple regression analyses controlling for age, sex, disease duration, and disease damage and activity.

Results: The total sample was 30.0% white, 22.3% Hispanic, 10.9% Black, 33.7% Asian, and 3.0% other race/ethnicity. 77.0% of interviews were conducted in-person. Among Hispanics and Asians, 26.0% and 18.6%, respectively, were non-English interviews. Each scale demonstrated adequate reliability and validity overall and within racial/ethnic groups. Minimal floor effects were observed, but ceiling effects were noted. Missing item responses were minimal for most scales, except for items related to work. No differences were noted by mode of administration or by language of administration among Hispanics and Asians. After accounting for differences in disease status, age, and sex, few differences in mean scores between whites and other racial/ethnic groups were noted.

Conclusion: PROMIS measures appear reliable and valid in lupus across racial/ethnic groups.

Lupus is a disease with extreme biological and clinical heterogeneity that makes measurement of outcomes challenging in clinical research. The complexity of lupus is

Corresponding author: Patricia Katz, PhD, University of California San Francisco, Arthritis Research Group, 3333 California Street, San Francisco, CA 94143-0936, Phone: 415-476-5971, Fax: 415-476-9030, Patti.katz@ucsf.edu.

evident in the clinical measures of disease damage and activity, which assess diverse manifestations across multiple organ systems. The range and complexity of patient-reported outcomes (PROs) parallel that of the clinical outcomes[1]. Multiple measures of lupus-specific "quality of life" have been published[1], but none are routinely used in clinical trials, observational studies, or clinical practice. The importance of including PROs as endpoints in clinical trials of novel therapies is gaining momentum and is recognized by the lupus community and the Food and Drug Administration[2,3]. However, there is no consensus on which PROs should be used.

The National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS®) initiative was undertaken to improve and standardize measurement of patient-reported outcomes (PROs)[4]. The PROMIS measures reflect the broad view of health proposed by the World Health Organization (http://www.who.int/about/definition/en/print.html), covering physical, mental, and social health. PROMIS measures were developed using state-of-the-art- psychometric techniques and may be administered via computer adaptive testing (CAT) or through static short forms that range from 4 to 20 items. Item banks and short forms exist for 20+ domains representing a comprehensive model of health and health-related quality of life (HRQL) that includes physical, social, and mental health. This theoretical framework is in accord with domains reported to be important and meaningful to individuals with lupus[1,5]. Notably, many of these content domains are among those of greatest concern to lupus patients (e.g., cognition, sleep)[1,5], yet are not measured in current generic HRQoL questionnaires.

To date, only a handful of published reports have examined the psychometric characteristics of PROMIS measures in cohorts of individuals with SLE[6–9]. In all four of these studies, participants were exclusively English-speaking. In two studies, participants were primarily white, and the remaining two, PROMIS responses by race/ethnicity were not examined. In this paper, we address this gap in the literature by examining the reliability, validity, and usefulness of the PROMIS measures in a racially/ethnically and linguistically diverse cohort of individuals with SLE.

Methods

Subjects

Subjects were participants in the California Lupus Epidemiology Study (CLUES), a multi-racial/ethnic cohort of individuals with physician-confirmed SLE. Participants were recruited from the California Lupus Surveillance Project, a population-based cohort of individuals with SLE living in San Francisco County from 2007 to 2009[10]. Additional participants residing in the geographic region were recruited through local academic and community rheumatology clinics and through existing local research cohorts.

Study procedures involved an in-person research clinic visit, which included collection and review of medical records prior to the visit; a history and physical examination conducted by a physician specializing in lupus; collection of biospecimens for clinical and research purposes; and completion of a structured interview administered by an experienced research assistant. All SLE diagnoses were confirmed by study physicians according to any of the

following definitions: (a) meeting 4 of the 11 American College of Rheumatology (ACR) revised criteria for the classification of SLE as defined in 1982 and updated in 1997[11,12], (b) meeting 3 of the 11 ACR criteria plus a documented rheumatologist's diagnosis of SLE, or (3) a confirmed diagnosis of lupus nephritis[10]. A subgroup of participants was unable to attend the in-person visit. For these individuals, medical records were collected and reviewed, and the same structured interview was administered by telephone. Diagnoses were confirmed through medical record review.

CLUES specifically aimed to include a diverse patient sample, with representation from multiple racial/ethnic groups speaking multiple languages. Study interviews were conducted in English, Spanish, Mandarin, or Cantonese. Data for these analyses included a total of 431 individuals, 332 of whom participated in in-person visits.

Variables

PROMIS.—The PROMIS short forms shown in Table 1 were administered as part of the structured interviews. All scales were scored as recommended and converted to T-scores, with a population mean of 50 and standard deviation (SD) of 10, using PROMIS scoring documentation available at http://assessmentcenter.net. For all PROMIS scales, higher scores reflect "more" of the construct being measured. For example, higher Physical Function and Satisfaction with Social Roles scores would reflect better functioning and satisfaction, so would be considered to be "better" scores; higher Fatigue, Pain Interference, Sleep Disturbance, Depression, and Anxiety scores would be considered to be "worse."

As noted above, PROMIS measures can be administered as static, short forms or through computer-adaptive testing (CAT). CAT is intended to administer items that are targeted to the individual respondent, which may lead to greater measurement precision[4]. However, the PROMIS item banks that support CAT are available only in English and Spanish, while the short forms are available in additional languages, including Mandarin and Cantonese. Because we wanted to use the same mode of administration for all CLUES participants, we chose to administer short forms.

Other patient-reported outcomes.—Three other instruments were used to measure PROs. The Medical Outcomes Study SF-36 is a widely used PRO measure and includes 8 subscales: Physical Function, Role Physical, Role Emotional, Vitality, Mental Health, Social Function, and Bodily Pain[13]. Scores for each scale range from 0 – 100, with a population mean of 50 and SD of 10. Higher scores for each scale except Bodily Pain reflect better outcomes. SLE disease activity was measured with the Systemic Lupus Activity Questionnaire (SLAQ)[14,15], a validated, self-report measure of SLE disease activity. SLAQ scores can range from 0 – 44, with higher scores reflecting more disease activity. The SLAQ also includes an item for respondents to rate the activity of their lupus over the past 3 months (0 [no activity – 10 [high activity]). The Brief Index of Lupus Damage (BILD) was used to estimate organ damage[16]. The BILD is based on Systemic Lupus International Cooperating Clinics/American College of Rheumatology Damage Index (SDI)[17], and consists of 28 items capturing information on 26 SDI items including determinations of important comorbid

conditions such as cardiovascular disease and events and diabetes. It has been shown to be predictive of hospitalizations and mortality[18].

Physician-reported measures.—The Systemic Lupus Erythematosus Disease Activity Index-(SELENA-SLEDAI)[19] and SDI[17] were completed by study rheumatologists during the in-person study visit.

Covariates.—Race, ethnicity, age, age at lupus onset, household income, and education level were self-reported. Language was categorized by the language in which interviews were conducted (English, Spanish, Mandarin, or Cantonese). Current medications were recorded during interviews.

Statistical analysis

Descriptive analyses.—Descriptive statistics were calculated for the total sample and for each racial/ethnic group. Differences in characteristics of groups were tested using t-tests and chi-square analyses. The percentage of respondents with missing items and scale scores was calculated. Distributions of PROMIS scores were examined. Because of the difference in the direction of scores (i.e., high scores reflected better health states for some scales and worse health states for other scales), we modified the standard terminology of floor and ceiling as follows: floor refers to the worst score, and ceiling to the best. T-tests and chi-square analyses were used to compare characteristics and PROMIS scores of individuals completing in-person versus telephone interviews

Reliability and validity.—Internal consistency was assessed by examining item-total correlations and Cronbach's alpha. Item-total correlations ≥ 0.4 and alpha values ≥ 0.80 are considered acceptable[20]. For assessment of convergent validity, Pearson and Spearman correlation analyses were used to examine associations of PROMIS scale scores with PROs for similar domains and measures of disease activity and damage. Correlations of 0.3 – 0.5 were considered low, 0.5 – 0.7 moderate, and ≥ 0.7 high[21].

Psychometric analyses by racial/ethnic group.—All descriptive, reliability, and validity analyses were repeated within each racial/ethnic group. Within the relevant racial/ethnic group, t-tests compared PROMIS scores of individuals completing interviews in English or another language.

Differential scores by race/ethnicity.—Differences in PROMIS scores by race/ethnicity were examined to determine if there appeared to be systematic differences in scores that were not attributable to differences in lupus severity, health status, or socioeconomic status. Differences in PROMIS scores between whites and other racial/ethnic groups were examined using multiple linear regression analyses, first with no covariates, and then controlling for age, sex, disease duration, SLEDAI, and SDI to determine if systematic differences among the groups remained. Multiple regression analyses were repeated using self-reported disease activity (SLAQ) and damage (BILD) so that telephone-only participants could be included in the analysis. Individuals categorized as "Other" race/

ethnicity were omitted from the race/ethnic-stratified analyses because of the small number (n=13). All analyses used SAS 9.4 (Cary, NC).

Results

Descriptive, total sample

Descriptive characteristics of the CLUES sample (n = 431) are shown in Table 2. Approximately 90% were female, mean age was 46.6 years, 19.5% had household incomes below the poverty level, and 22.1% had education at the high school level or lower. Thirty percent were White, 22.3% Hispanic, 10.9% Black, 33.7% Asian/Pacific Islander, and 3.0% other race/ethnicity.

Missing item data were greatest in the Satisfaction with Social Roles scale, ranging from 1.1% of items for Hispanic Spanish-speaking respondents to 4.9% for White respondents (Table 3). Specific items with the greatest number of missing values queried satisfaction with the amount of work one could do (9.0% missing), ability to work (6.6% missing), and ability to meet the needs of those who depend on the respondent (3.8% missing). The item in the Ability to Participate in Social Roles scale dealing with work also had a relatively large number of missing responses (5.9%). The two Psychosocial Impact of Illness scales and the Social Isolation scales also had a relatively large number of missing item data, ranging from 0.5% to 2.8% of items, although missingness was not concentrated on specific items.

Mean scores of all PROMIS measures were within one-half standard deviation of the population mean (50.0) (Table 4). Floor effects (worse scores) were minimal, with the largest seen for Fatigue (5.4%). In contrast, over 20% of the cohort scored at the ceiling (best scores) for Physical Function, Positive Psychosocial Impact of Illness, and all four social health scales.

Ninety-nine of the CLUES participants completed the PROMIS measures by telephone. The telephone completion group was older (Phone: 52 years vs. In-person: 45 years, p <.0001), more likely to complete the interview in English (95% vs. 86%, p=.02), and had disease of longer duration (23 years vs. 16 years, p<.0001) and higher BILD scores (2.6 vs. 1.9, p=.03). There were no other significant differences between the two groups in sex, race/ethnicity, education, income, SLAQ, or PROMIS scale scores (data not shown).

Reliability and validity, total sample

All item-total correlations were >0.50, except for three individual items, which were all above 0.40 (Table 4). Cronbach's alpha was acceptable across domains: 0.80 for all scales except Negative Psychosocial Impact (α = 0.79) and Positive Psychosocial Impact (α = 0.79).

PROMIS scores demonstrated moderate to high correlations with SF-36 scores measuring similar constructs (Table 5). The highest correlations were noted for the scales with the most similar content (i.e., physical function, pain, and fatigue). PROMIS scores had moderate correlations with patient-reported disease activity (SLAQ), and low correlations with patient-reported disease damage (BILD) (Table 5). However, there were no associations between

PROMIS measures and physician-assessed disease activity (SLEDAI), and only minimal associations with physician-assessed disease damage (SDI).

Descriptive, by racial/ethnic group

White participants were significantly older, with disease of longer duration, and were significantly less likely to have below-poverty incomes or low education (Table 2). There were no significant differences among groups in the physician-assessed measures of disease activity and damage. In contrast, Asian patients had significantly lower scores on the self-reported disease activity measures. Twenty-six percent of the Hispanic participants and 19% of the Asian participants completed the PROMIS measures in Spanish and Chinese (Cantonese or Mandarin), respectively. There were no significant differences in PROMIS scores by language for these two groups (data not shown. As noted in Table 1, only 10 of 12 PROMIS measures were available in Spanish and only 4 in Chinese languages).

Psychometric analysis, by race/ethnicity

There were no appreciable racial/ethnic differences in the percentage of scores at the floor (Supplemental Table 1). However, racial/ethnic differences were apparent in the percentage of ceiling responses, with the most notable pattern being a lower percentage of Black patients at the ceiling for a number of scales. All Cronbach's alphas were 0.80 when examined by racial/ethnic group, except for Sleep Disturbance (Whites, Blacks), Psychosocial Impact of Illness Negative (Hispanics, Blacks, Asians), and Psychosocial Impact of Illness Positive (Blacks, Asians); each of these alpha coefficients was 0.70. Item-total correlations were 0.50 for all groups, with a few exceptions, most notably the Physical Function item regarding limitations in vigorous activity, the Sleep Impairment item regarding feeling alert upon awakening, and the Psychosocial Impact of Illness item regarding worry about the future.

There were no substantive differences in correlations with SF-36 scales by race/ethnicity (Supplemental Table 2). As with the total cohort, correlations of PROMIS scores with physician-assessed disease activity were minimal for all groups, and correlations with patient reported disease activity were generally moderate. Among White participants only, correlations with SDI and BILD were significant for almost all PROMIS scales, although most correlations were small. No consistent patterns of differences among racial/ethnic groups were noted (Supplemental Table 3).

Differences in scores by race/ethnicity

In bivariate analyses, Asians had significantly better scores than whites for PROMIS Physical Function, Pain Interference, Fatigue, Ability to Participate in Social Roles and Activities, Satisfaction with Discretionary Activities, and Satisfaction with Social Roles (Table 6). Blacks had significantly worse Physical Function scores. In analyses adjusting for age, sex, disease duration, SLEDAI, and SDI, differences between whites and other race/ethnicity groups were seen only for Fatigue and Satisfaction with Discretionary Activities. Again, Asians had significantly better scores than whites for each of these scales (Table 6, Multivariable Model 1). After further adjustment for obesity and smoking, these differences remained (Table 6, Multivariable Model 2). Further adjustment for income did not change

results substantively (data not shown). In sensitivity analyses using patient-reported disease activity and damage instead of SLEDAI and SDI in order to include subjects who participated by phone only, similar results were noted. (Table 6, Model 3).

Discussion

Our study presents the first examination of PROMIS short-forms across different racial/ ethnic and language groups in a diverse lupus cohort. Overall, each of the scales demonstrated adequate reliability (internal consistency) and validity (correlations with similar measures). Minimal floor effects were observed, but ceiling effects were noted, particularly in Social Health measures, which could limit responsiveness to change. Missing item responses and resulting missing scale scores were minimal and random for most scales. The notable exceptions were the Satisfaction with Social Roles, Participation in Social Roles and Activities, and the two Psychosocial Impact of Illness scales. For the first two of these, items dealing with work accounted for the majority of missing item responses. It is possible that individuals who were not working felt these items were not applicable to them. For the Psychosocial Impact of Illness scales, there were no clear patterns of missing items. Item-total correlations were lower than optimal for a few items, including ability to engage in vigorous physical activities, feeling alert upon awakening, and worry about the future. The relatively poor performance of these items within their respective scales may reflect lupus-specific biases; i.e., the domains addressed by these items may be affected by lupus in a manner different that in the general population. Further work is needed to determine the underlying reasons for these anomalies and the usefulness of these items among individuals with lupus, particularly those who may be work disabled.

All of the PROMIS short forms demonstrated consistent reliability and validity across racial/ ethnic groups. We found no differences by mode of administration (in-person vs. telephone) or by language of administration among the Hispanic and Asian participants. Bivariate analyses showed significant differences in mean scores by race/ethnicity for more than half of the scales. However, after accounting for differences in disease status, age, and sex, few differences remained between whites and other racial/ethnic groups, suggesting that differences in scale scores may be attributable to differences in disease and demographics rather than race/ethnicity *per se*.

We found that in our population-based cohort of individuals with lupus, scores were generally reflective of better health than have been reported in a sample of lupus patients recruited from a clinical setting[6]. That study also reported fewer ceiling effects, possibly due to the shift toward lower mean scores. The difference between that study and the data reported here may be because individuals are less likely to attend a research visit during episodes of poor health or a flare, while clinical visits are more likely to occur during those times. However, the study of clinic patients also used the CAT versions of the PROMIS scales, so it is possible that the CAT version produces greater precision of item selection and yields fewer ceiling effects, as has been suggested in studies the PROMIS measures of depressive symptoms[22] and physical function[23].

We found no association of PROMIS scores with physician-assessed disease activity and only minimal correlations with physician-assessed disease damage. A similar lack of correlation with the SLEDAI and low correlation with SDI was noted in a previous study by Kasturi et al [6]. However, the lack of correspondence between physician-completed and patient-reported measures in lupus has been well documented[24], so this finding is not surprising.

Strengths of this study include the diverse cohort, with sufficient sample sizes to examine measures by racial/ethnic group. This study included the largest number of PROMIS short forms administered in a lupus cohort. In addition, administration in multiple languages and in both in-person and phone formats permitted comparisons of these subgroups. Only four other studies have been published examining PROMIS scales in lupus. Two examined the PROMIS 29-item profile, which includes 4-item short forms for physical function, fatigue, pain interference, sleep disturbance, satisfaction with social role, anxiety, and depression[7,8], in non-clinical study settings among patients who were primarily white and exclusively English-speaking. The remaining publications were based in a clinical setting, with a more diverse, yet exclusively English-speaking, cohort, but did not examine racial/ethnic differences in scale performance[6,9].

Limitations include the under-representation of clinically active lupus, as noted above. While the cohort was quite diverse, the number of non-English-speakers was relatively small. Yet, this is the first comparison of PROMIS scores of English and non-English speakers. A limited number of legacy measures were available for validity analyses, however the SF-36 is the PRO most commonly used in lupus studies, including clinical trials. A comparison of PROMIS measures with one of the lupus-specific quality of life measures could provide useful information. All questionnaires were interviewer administered, so results may have been different if self-administered. However, interviewer administration provided consistency in mode of administration between the in-person and phone interviews.

PROMIS measures offer several advantages to existing PRO measures, in particular the SF-36, which is the most commonly used PRO in lupus studies. There is evidence that the SF-36 does not adequately cover the broad range of symptoms and outcomes important in lupus[5,25]. With PROMIS, a broader range of domains can be examined, including domains that are relevant to SLE and meaningful to patients, such as sleep quality, cognitive abilities, impact of pain, and satisfaction with social roles, with a relatively small response burden. Like the SF-36, PROMIS short forms have been translated and culturally adapted to multiple languages. PROMIS measures are available in the public domain and are free to use in clinical trials as well as in registries, observational studies, and clinical practice. PROMIS measures have also been adapted to use within many existing electronic health record systems.

It is yet to be seen if the PROMIS measures are responsive to changes in lupus disease activity or severity. The ceiling effects noted in this study may indicate a limitation in responsiveness, but such limitations may not exist in a sample of respondents with more active disease. Additional studies are needed to determine responsiveness to change and minimal clinically important differences in lupus.

In summary, these results add to the growing evidence supporting PROMIS measures as reliable and valid in lupus. This study adds information regarding the performance of PROMIS measures in lupus across racial/ethnic groups, which is particularly important given the high burden of disease among racial/ethnic minorities. Differences that we observed between racial/ethnic groups appeared to be primarily due to differences in the groups' clinical or sociodemographic characteristics rather than differential scale performance. Overall, the PROMIS measures appear to be well suited to use in lupus and may be particularly useful as outcome measures in clinical trials of targeted therapies and longitudinal studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Yazdany J Health-related quality of life measurement in adult systemic lupus erythematosus. Arthritis Care & Research. 2011;63:S413–S419. [PubMed: 22588761]

2. Bertsias G, Gordon C, Boumpas D. Clinical trials in systemic lupus erythematosus (SLE): lessons from the past as we proceed to the future -- the EULAR recommendations for the management of SLE and the use of end-points in clinical trials. Lupus. 2008;17:437–442. [PubMed: 18490423]

3. U.S. Department of Health and Human Services. Guidance for industry. Systemic lupus erythematosus — developing medical products for treatment. In: Administration FaD, ed2010.

4. Witter J The promise of Patient-Reported Outcomes Measurement Information System -- turning theory into reality: a uniform approach to patient-reported outcomes across rheumatic disease. Rheum Dis Clin North Am. 2016;42:377–394. [PubMed: 27133496]

5. Stamm T, Bauernfeind B, Coenen M, et al. Concepts important to persons with systemic lupus erythematosus and their coverage by standard measures of disease activity and health status. Arthritis Care & Research. 2007;57:1287–1295. [PubMed: 17907225]

6. Kasturi S, Szymonifka J, Burket J, et al. Validity and reliability of patient reported outcomes measurement information system computerized adaptive tests in systemic lupus erythematosus. J Rheumatol. 2017;44:1024–1031. [PubMed: 28412708]

7. Lai J, Beaumont J, Jensen S, et al. An evaluation of health-related quality of life in patients with systemic lupus erythematosus using PROMIS and Neuro-QoL. Clin Rheumatol. 2017;36:555–562. [PubMed: 27848056]

8. Katz P, Andrews J, Yazdany J, Schmajuk G, Trupin L, Yelin E. Is frailty a relevant concept in SLE? Lupus Sci Med. 2017;4:3000186.

9. Kasturi S, Szymonifka J, Burket J, et al. Feasibility, validity, and reliability of the 10-item Patient Reported Outcomes Measurement Information System Global Health short form in outpatients with systemic lupus erythematosus. J Rheumatol. 2018;45:397–404. [PubMed: 29419473]

10. Dall'Era M, Cisternas M, Snipes K, Herrinton L, Gordon C, Helmick C. The incidence and prevalence of systemic lupus erythematosus in San Francisco County, California. Arthrits Rheum. 2017;69:1996–2005.

11. Tan EM, Cohen AS, Fries JF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. 1982;25(11):1271–1277. [PubMed: 7138600]

12. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum. 1997;40(9):1725.

13. Ware JJ, Snow K, Kosinski M, Gandek B. SF-36 Health Survey: manual and interpretation guide. Boston, Massachusetts: The Health Institute, New England Medical Center; 1993.

14. Karlson E, Daltroy L, Rivest C, et al. Validation of a systemic lupus activity questionnaire (SLAQ) for population studies. Lupus. 2003;12:280–286. [PubMed: 12729051]

15. Yazdany J, Yelin E, Panopalis P, Trupin L, Julian L, Katz P. Validation of the systemic lupus erythematosus activity questionnaire in a large obsevational cohort. Arthritis Rheum (Arthritis Care Res). 2008;59:136–143.

16. Yazdany J, Trupin L, Gansky S, et al. The Brief Index of Lupus Damage: a patient-reported measure of damage in systemic lupus erythematosus. Arthritis Care & Research. 2011;63:1170–1177.

17. Gladman D, Ginzler E, Goldsmith C, et al. The development and initial validation of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index for systemic lupus erythematosus. Arthritis Rheum. 1996;39:363–369. [PubMed: 8607884]

18. Katz P, Trupin L, Rush S, Yazdany J. Longitudinal validation of the Brief Index of Lupus Damage (BILD). Arthritis Care & Research. 2014;66:1057–1062. [PubMed: 24376263]

19. Buyon J, Petri M, Kim M, et al. The effect of combined estrogen and progesterone hormone replacement therapy on disease activity in systemic lupus erythematosus: a randomized trial. Ann Intern Med. 2005;142:953–962. [PubMed: 15968009]

20. Nunnally J, Bernstein I. Psychometric theory. 3rd ed. New York: McGraw-Hill, Inc; 1994.

21. Hinkle D, Wiersma W, Jurs S. Applied statistics for the behavioral sciences. 5th ed. Boston: Houghton Mifflin; 2003.

22. Choi S, Reise S, PIlkonis P, Hays R, Cella D. Efficiency of static and computer adaptive short form forms compared to full-length measures of depressive symptoms. Qual Life Res. 2010;19:125–136. [PubMed: 19941077]

23. Rose M, Bjorner J, Gandek B, Bruce B, Fries J, Ware J Jr. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. J Clin Epidemiol. 2014;67:516–526. [PubMed: 24698295]

24. Bertsias G, Ioannidis J, Boletis J, et al. EULAR points to consider for conducting clinical trials in systemic lupus erythematosus: literature based evidence for the selection of endpoints. Ann Rheum Dis. 2009;68:477–483. [PubMed: 18434449]

25. Ow Y, Thumboo J, Cella D, Cheung Y, Fong K, Week H. Domains of health-related quality of life important and relevant to multiethnic English-speaking Asian systemic lupus erythematosus patients: a focus group study. Arthritis Care Res. 2011;63:899–908.

Significance and Innovation

- This presents the first examination of PROMIS short-forms across different racial/ethnic and language groups in a diverse lupus cohort, showing adequate reliability and validity, and minimal floor effects for each of four racial ethnic groups.

- No differences were noted by mode of administration (in-person vs. telephone) or by language among Hispanic and Asian participants.

- After controlling for differences in disease status, age, and sex, few differences existed between Caucasians and other racial/ethnic groups, suggesting that differences in scale scores may be primarily attributable to differences in disease and demographics rather than race/ethnicity *per se*.

Table 1.

PROMIS measures administered in CLUES

PROMIS short form	Number of items	English	Spanish	Chinese
Physical Health				
Physical Function	10	✓	✓	✓
Pain Interference	4 (6)[*]	✓	✓	✓
Fatigue	4 (7)[*]	✓	✓	✓
Sleep Disturbance	4	✓	✓	✓
Sleep Impairment	8	✓	✓	
Mental Health				
Applied Cognition, Abilities	4	✓	✓	
Psychosocial illness impact, negative	8[†]	✓		
Psychosocial illness impact, positive	8[†]	✓		
Social health				
Ability to Participate in Social Roles and Activities	4	✓	✓	
Satisfaction with Participation in Discretionary Social Activities	7	✓	✓	
Satisfaction with Participation in Social Roles	7	✓	✓	
Social Isolation	4	✓	✓	

[*] Number of items in Chinese version

[†] Only 4 items are scored.

Table 2.

Baseline characteristics of CLUES cohort (n = 431)

	Total	White	Hispanic	Black	Asian	Other	P
N	431	130 (30.0)	96 (22.3)	47 (10.9)	145 (33.7)	13 (3.0)	
In-person interview	332 (77.0)	96 (73.9)	76 (79.2)	36 (76.6)	118 (81.4)	6 (46.2)	.05
Female	387 (89.8)	116 (89.2)	83 (86.5)	46 (97.9)	129 (89.0)	13 (100)	.19
Age	46.6 ± 14.3	51.4 ± 12.3	42.7 ± 14.1	52.8 ± 14.8	42.7 ± 14.0	48.7 ± 13.3	<.0001
Below poverty	75 (19.4)	7 (5.7)	25 (29.1)	15 (37.5)	24 (19.1)	4 (36.4)	<.0001
Low education	94 (22.1)	11 (8.6)	29 (30.5)	18 (38.3)	34 (23.6)	2 (16.7)	<.0001
Non-English interview	52 (12.1)	0	25 (26.0)	0	27 (18.6)	0	<.0001
Disease duration	17.7 ± 11.1	22.1 ± 10.7	15.0 ± 9.9	18.9 ± 13.1	14.6 ± 9.8	24.3 ± 11.0	<.0001
Current glucocorticoid (GC) use	210 (48.7)	50 (38.5)	50 (52.1)	26 (55.3)	77 (53.1)	7 (53.9)	.09
High dose GC use (7.5 mg for 3 months in past year)	95 (22.4)	27 (20.8)	22 (23.2)	10 (21.7)	32 (22.5	34 (3.3	.90
Current non-GC immunosuppressant use	200 (46.4)	39 (30.0)	48 (50.0)	25 (53.2)	84 (57.9)	4 (30.8)	<.0001
SLEDAI (n = 330)	3.0 ± 3.1	2.4 ± 3.0	3.6 ± 3.6	2.3 ± 2.0	3.1 ± 2.9	4.2 ± 4.5	.07
SDI (n = 331)	1.8 ± 2.0	1.9 ± 2.2	1.9 ± 2.0	2.4 ± 2.2	1.6 ± 1.8	2.8 ± 2.4	.21
BILD	2.1 ± 2.3	2.1 ± 2.3	2.3 ± 2.6	2.4 ± 2.3	1.8 ± 2.0	3.1 ± 2.0	.09
SLAQ	8.8 ± 7.3	9.4 ± 7.5	9.2 ± 7.2	11.3 ± 7.8	7.0 ± 6.6	12.4 ± 8.6	.0007
SLE activity (0 – 10 rating)	3.3 ± 2.7	3.1 ± 2.7	3.7 ± 2.6	4.3 ± 3.0	2.7 ± 2.5	4.5 ± 2.6	.001

Tabled values are n (%) or mean ± standard deviation

Table 3.

Missing data in PROMIS items and scale scores in CLUES cohort by race/ethnicity [*]

	missing	White (n=130)	Hispanic, Spanish (n=25)	Hispanic, English (n=71)	Black (n=47)	Asian, Chinese (n=27)	Asian, English (n=118)
Physical Health							
Physical Function	# items	1	1	2	0	1	0
(10 items)	% items [†]	0.1	0.4	0.3	0	0.4	0
	# scale scores	0	0	0	0	0	0
Pain Interference	# items	1	1	0	6	5	0
(4 items; 6 items in Chinese)	% items	0.2	1.0	0	3.2	3.1	0
	# scale scores	1	0	0	4	0	0
Fatigue	# items	0	0	0	2	5	4
(4 items; 7 items in Chinese)	% items	0	0	0	1.1	2.7	0.9
	# scale scores	0	0	0	1	0	3
Sleep Disturbance	# items	1	0	0	0	1	1
(4 items)	% items	0.2	0	0	0	0.9	0.2
	# scale scores	1	0	0	0	1	1
Sleep Impairment	# items	2	0	1	3	---	1
(8 items)	% items	0.2	0	0.2	0.8	---	0.1
	# scale scores	0	0	0	0	---	0
Mental Health							
Cognitive	# items	0	1	0	0		1
(4 items)	% items	0	1.0	0	0	---	0.2
	# scale scores	0	1	0	0	---	1
Psychosocial Impact, Negative	# items	12	---	8	1	---	13
(4 items scored)	% items	2.3	---	2.8	0.5	---	2.8
	# scale scores	6	---	5	1	---	8
Psychosocial Impact, Positive	# items	13	---	8	3	---	11
(4 items scored)	% items	2.5	---	2.8	1.6	---	2.3
	# scale scores	7	---	5	3	---	7
Social Health							
Ability to Participate Social	# items	16	0	8	9	---	6
Roles, Activities	% items	3.1	0	2.8	4.8	---	1.2
(4 items)	# scale scores	10	0	5	6	---	6
Satisfaction, Discretionary	# items	16	0	0	9	---	2
Social Activities	% items	1.8	0	0	2.7	---	0.2
(7 items)	# scale scores	2	0	0	1	---	0

	missing	White (n=130)	Hispanic, Spanish (n=25)	Hispanic, English (n=71)	Black (n=47)	Asian, Chinese (n=27)	Asian, English (n=118)
Satisfaction Social Roles	# items	44	2	13	14	---	26
(7 items)	% items	4.9	1.1	2.6	4.3	---	3.2
	# scale scores	3	1	0	1	---	1
Social Isolation	# items	9	1	1	0	---	3
(4 items)	% items	1.7	1.0	0.4	0	---	0.6
	# scale scores	3	1	1	0	---	2

*
"Other" race/ethnicity excluded because of small sample size

†
% items = (number of items in scale with missing responses) / (number of items in scale × number of respondents)

Table 4.

PROMIS scale T-score characteristics in CLUES cohort

	Mean ± SD	% at floor	% at ceiling	Cronbach's alpha
Physical Health				
Physical Function	47.4 ± 9.9	0.2	20.2	.94
Pain Interference	52.4 ± 10.0	0.5	3.3	.95
Fatigue	52.5 ± 11.6	5.4	0.5	.96
Sleep Disturbance	52.7 ± 9.1	2.3	3.7	.80
Sleep Impairment	52.9 ± 10.7	0.3	4.5	.92
Mental Health				
Cognition, Ability	48.7 ± 8.5	2.2	14.7	.90
Psychosocial Illness Impact, Negative	52.1 ± 8.2	0.3	12.3	.78
Psychosocial Illness Impact, Positive	48.2 ± 9.1	0.6	21.7	.79
Social Health				
Ability to Participate Social Roles, Activities	50.5 ± 10.0	3.2	24.8	.96
Satisfaction, Discretionary Social Activities	52.8 ± 10.0	1.8	20.0	.95
Satisfaction, Social Roles	51.1 ± 10.7	2.3	24.1	.96
Social Isolation	46.3 ± 9.4	0.5	27.5	.90

Note: For PROMIS scales, higher T-scores reflect "more" of the domain being measured; i.e., better physical function, more pain interference. Because the directionality of scores is not consistent in terms of "best" or "worst" scores, we defined floor and ceiling to have a consistent meaning. Floor = worst score, Ceiling = best score.

All item-total correlations >0.50 except:

Physical Function: limitations in vigorous activity (r = .48)

Sleep Impairment: "I felt alert when I woke up" (r = .39)

PIN: "I worry about the future" (r = 0.44)

Table 5.

PROMIS scale correlations with SF-36 and measures of disease activity and damage

PROMIS scales	Patient-reported Measures										Physician-reported measures	
	SF-36 PF	SF-36 Pain	SF-36 Vit	SF-36 SF	SF-36 RP	SF-36 RE	SF-36 MH	Activity: SLAQ	Activity: 0-10 rating	Damage: BILD	Activity: SLEDAI	Damage: SDI
Physical Health												
Physical Function	0.94 *							−0.60 *	−0.48 *	−0.38 *	0.09	−0.32 *
Pain Interference		−0.79 *						0.65 *	0.56 *	0.28 *	−0.05	0.19 †
Fatigue			−0.80 *					0.67 *	0.56 *	0.17 †	−0.03	0.13
Sleep Disturbance			−0.55 *					0.44	0.37 *	0.08	−0.01	0.01
Sleep Impairment			−0.78 *					0.60 *	0.51 *	0.16 †	−0.03	0.07
Mental Health												
Cognition, Ability						0.41 *	0.42 *	−0.44 *	−0.35 *	−0.20 *	−0.01	−0.14 †
Psychosocial Illness Impact, Negative						−0.43 *	−0.60 *	0.41 *	0.34 *	0.12 †	−0.03	0.08
Psychosocial Illness Impact, Positive						0.47 *	0.53 *	−0.38 *	−0.29 *	−0.15 †	−0.01	−0.10
Social Health												
Ability to Participate Social Roles, Activities				0.51 *	0.72 *			−0.59 *	−0.48 *	−0.27 *	0.05	−0.22 †
Satisfaction, Discretionary Social Activities				0.67 *	0.68 *			−0.58 *	−0.49 *	−0.22 *	0.01	−0.20 †
Satisfaction, Social Roles				0.66 *	0.74 *			−0.58 *	−0.48 *	−0.23 *	0.01	−0.19 †
Social Isolation				−0.53 *	−0.51 *			0.45 *	0.34 *	0.18 †	−0.04	0.10

*
p<0.0001

†
p<0.05

SF36 scales: PF = Physical Function; Vit = Vitality; SF = Social Functioning; RP = Role Physical; RE = Role Emotional; MH = Mental Health

SLAQ = Systemic Lupus Activity Questionnaire

0-10 rating = From SLAQ, rating of lupus over the past 3 months. 0 = no activity, 10 = high activity.

BILD = Brief Index of Lupus Damage

SLEDAI = Systemic Lupus Erythematosus Disease Activity Index

SDI = SLICC/ACR Damage Index;

For PROMIS scales, higher T-scores reflect "more" of the domain being measured; i.e., better physical function, more pain interference.

Table 6.

PROMIS scores in CLUES cohort by race/ethnicity

| | Mean scores (± standard deviation) | | | | β from multiple linear regression analysis* | | | | | | | | |
| | Racial/ethnic group | | | | Model 1 | | | Model 2 | | | Model 3 | | |
	White (n = 127)	Hispanic (n = 96)	Black (n=47)	Asian (n = 145)	Hispanic	Black	Asian	Hispanic	Black	Asian	Hispanic	Black	Asian
Physical Health													
Physical Function	47.2 ± 10.2	47.6 ± 10.0	42.1 ± 9.7†	49.7 ± 8.9†	−1.5	−3.3	0.7	−0.8	−1.5	−0.2	−0.5	−3.2	−0.5
Pain Interference	52.8 ± 9.8	53.1 ± 9.6	55.7 ± 10.3	50.0 ± 9.7†	0.9	1.3	−2.6	0.3	−0.3	−1.8	−0.0	1.3	−1.0
Fatigue	54.4 ± 11.8	52.1 ± 11.3	55.4 ± 10.4	49.4 ± 11.2†	−2.5	0.4	−5.0†	−3.2	−1.7	−4.3†	−3.7†	−2.1	−4.0†
Sleep Disturbance	52.2 ± 8.6	54.2 ± 9.3	53.9 ± 8.4	51.6 ± 9.4	1.5	0.9	−0.7	1.0	−0.4	−0.3	1.7	0.1	0.3
Sleep Impairment	53.3 ± 11.1	52.3 ± 10.7	54.4 ± 10.1	51.8 ± 10.4	−1.6	−0.3	−2.1	−2.3	−2.4	−1.5	−2.4	−1.5	−0.7
Mental Health													
Cognition, Ability	48.9 ± 9.3	48.6 ± 8.1	46.8 ± 6.7	49.3 ± 8.6	0.1	1.0	1.1	0.6	2.5	0.6	−0.2	−0.5	−0.6
Psychosocial illness impact, negative	52.7 ± 8.1	51.5 ± 8.0	52.3 ± 9.5	51.6 ± 7.9	−2.2	−0.2	−1.5	−2.2	−1.2	−1.3	−1.9	−1.8	−1.0
Psychosocial illness impact, positive	48.8 ± 8.9	47.2 ± 10.0	46.0 ± 9.0	49.0 ± 9.0	−1.6	−1.3	−0.1	−1.5	−0.2	−0.5	−2.3	−1.8	−1.3
Social Health													
Ability to Participate in Social Roles, Activities	49.6 ± 10.2	51.2 ± 10.4	47.1 ± 8.6	52.9 ± 9.4†	2.2	1.1	2.6	2.8	2.4	2.1	2.2	0.6	1.1
Satisfaction, Discretionary	51.3 ± 10.6	52.5 ± 10.0	52.2 ± 9.1	55.7 ± 9.3†	1.4	3.0	4.4†	2.0	4.7†	3.7†	1.7	3.2†	2.6†
Satisfaction, Social Roles	50.3 ± 10.8	50.8 ± 10.5	49.2 ± 11.0	53.8 ± 10.3†	0.5	0.5	3.0	1.2	2.5	2.2	0.8	1.1	1.4
Social Isolation	46.5 ± 8.9	45.4 ± 10.2	48.3 ± 8.9	45.3 ± 9.3	−1.1	1.6	−1.0	−1.6	0.3	−0.6	−1.7	0.4	−0.2

*
Reference group = White. β indicates difference from score of whites.

†
p<0.05, compared to Whites

Model 1 adjusted for age, sex, disease duration, and physician-reported disease activity and damage (SLEDAI and SDI)

Model 2 adjusted for age, sex, disease duration, obesity, smoking, and physician-reported disease activity and damage (SLEDAI and SDI). Further adjustment for income did not change results substantively

Model 3 adjusted for age, sex, disease duration, and patient-reported disease activity and damage (SLE activity [0 – 10 rating] and BILD)

Note: "Other" racial/ethnic group omitted because of small sample (n = 13)