

Validity and Reliability of Patient Reported Outcomes Measurement Information System Computerized Adaptive Tests in Systemic Lupus Erythematosus

Shanthini Kasturi, Jackie Szymonifka, Jayme C. Burket, Jessica R. Berman, Kyriakos A. Kirou, Alana B. Levine, Lisa R. Sammaritano, and Lisa A. Mandl

ABSTRACT. Objective. The aims of this study were to assess the construct validity and the test-retest reliability of Patient Reported Outcomes Measurement Information System (PROMIS) computerized adaptive tests (CAT) in patients with systemic lupus erythematosus (SLE).

Methods. Adults with SLE completed the Medical Outcomes Study Short Form-36, LupusQoL-US version (“legacy instruments”), and 14 selected PROMIS CAT. Using Spearman correlations, PROMIS CAT were compared with similar domains measured with legacy instruments. CAT were also correlated with the Safety of Estrogens in Lupus Erythematosus National Assessment–Systemic Lupus Erythematosus Disease Activity Index (SELENA-SLEDAI) disease activity and the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI) scores. Test-retest reliability was evaluated using ICC.

Results. There were 204 outpatients with SLE enrolled in the study and 162 completed a retest. PROMIS CAT showed good performance characteristics and moderate to strong correlations with similar domains in the 2 legacy instruments ($r = -0.49$ to 0.86 , $p < 0.001$). However, correlations between PROMIS CAT and the SELENA-SLEDAI disease activity and SDI were generally weak and statistically insignificant. PROMIS CAT test-retest ICC were good to excellent, ranging from 0.72 to 0.88.

Conclusion. To our knowledge, these data are the first to show that PROMIS CAT are valid and reliable for many SLE-relevant domains. Importantly, PROMIS scores did not correlate well with physician-derived measures. This disconnect between objective signs and symptoms and the subjective patient disease experience underscores the crucial need to integrate patient-reported outcomes into clinical care to ensure optimal disease management. (First Release April 15 2017; *J Rheumatol* 2017;44:1024–31; doi:10.3899/jrheum.161202)

Key Indexing Terms:

SYSTEMIC LUPUS ERYTHEMATOSUS
QUALITY OF LIFE

PATIENT OUTCOME ASSESSMENT
VALIDATION STUDY

From the Division of Rheumatology/Department of Medicine, Hospital for Special Surgery; Weill Cornell Medicine, New York, New York, USA.

Supported by the Rheumatology Research Foundation Scientist Development Award.

S. Kasturi, MD, MS, Division of Rheumatology/Department of Medicine, Hospital for Special Surgery, and Weill Cornell Medicine; J. Szymonifka, MA, Division of Rheumatology/Department of Medicine, Hospital for Special Surgery; J.C. Burket, PhD, Healthcare Research Institute, Hospital for Special Surgery; J.R. Berman, MD, Division of Rheumatology/Department of Medicine, Hospital for Special Surgery, and Weill Cornell Medicine; K.A. Kirou, MD, Division of Rheumatology/Department of Medicine, Hospital for Special Surgery, and Weill Cornell Medicine; A.B. Levine, MD, Division of Rheumatology/Department of Medicine, Hospital for Special Surgery, and Weill Cornell Medicine; L.R. Sammaritano, MD, Division of Rheumatology/Department of Medicine, Hospital for Special Surgery, and Weill Cornell Medicine; L.A. Mandl, MD, MPH, Division of Rheumatology/Department of Medicine, Hospital for Special Surgery and Weill Cornell Medicine.

Address correspondence to Dr. S. Kasturi, Hospital for Special Surgery, 535 East 70th St., New York, New York 10021, USA.

E-mail: kasturis@hss.edu

Accepted for publication March 14, 2017.

The accurate measurement of health-related quality of life (HRQOL), an important patient-reported outcome (PRO), is critical to providing patient-centered care. This is especially important in diseases such as systemic lupus erythematosus (SLE), in which dramatically lower mortality rates have refocused care on minimizing morbidity¹. Physicians and patients have different perceptions of the effect of SLE. For example, patients focus on functional status whereas physicians focus on laboratory values². Further, it is well known that SLE significantly decreases HRQOL³, but exactly how HRQOL should best be defined and measured is unclear.

The US Food and Drug Administration, the European Medical League, and the Outcome Measures in Rheumatology Clinical Trials (OMERACT) group have identified HRQOL as a crucial outcome measure for clinical trials and observational studies in SLE^{4,5,6}. They recommend the use of both generic and disease-specific measures that would

allow comparisons with healthy individuals while also ensuring the inclusion of domains that are meaningful to patients.

A number of generic and disease-specific instruments have been validated for the measurement of PRO in SLE, but all have significant limitations^{7,8}. The Medical Outcomes Study Short Form-36 (SF-36)⁹ is a widely used generic measure in SLE, but has variable longitudinal responsiveness^{10,11,12} and lacks multiple domains of relevance to patients with SLE, such as fatigue, sleep, and cognition^{13,14,15}. The LupusQoL, the most extensively validated SLE-specific instrument, includes several of these SLE-specific domains, such as fatigue, body image, and planning, but has significant floor and ceiling effects¹⁶. In addition, both measures can be challenging to administer and score at the point of care.

The Patient Reported Outcomes Measurement Information System (PROMIS) is a novel publicly available psychometrically validated system developed by the US National Institutes of Health to efficiently measure PRO in populations with a wide range of chronic diseases¹⁷. PROMIS instruments increase measurement precision and reduce responder burden relative to traditional instruments because they use item response theory and include computerized adaptive tests (CAT). CAT select the most informative questions from an item bank based on subjects' previous responses, permitting the use of fewer questions per domain with more precision. PROMIS item banks are generic, scored as T scores normalized to the general population in the United States, and include numerous domains of relevance to patients with SLE that are not found in the SF-36, such as fatigue, sleep, and cognition.

The performance characteristics of PROMIS CAT have not yet been evaluated in SLE. Our study describes the validity and reliability of 14 PROMIS CAT compared with both the SF-36 and the LupusQoL in adult outpatients with SLE. Second, we evaluate the correlation of PROMIS CAT with physician assessments of disease activity and damage.

MATERIALS AND METHODS

English-speaking adults ≥ 18 years receiving care at the Hospital for Special Surgery (HSS) Lupus Center of Excellence and meeting ≥ 4 of the American College of Rheumatology (ACR) 1997 SLE criteria were eligible to participate in our prospective cohort study¹⁸. Patients on dialysis and those with active malignancy, other than nonmelanomatous skin cancer, were excluded.

Patients with SLE were identified by their treating rheumatologists and medical records were reviewed to confirm eligibility. Patients consented to participate in our study at the time of an outpatient visit. Patients could complete the Web-based surveys on-site during their visit by computer or tablet with the technical assistance of a study investigator. Alternatively, patients could complete the surveys remotely on a computer, tablet, or smartphone by an e-mailed study-specific URL. Consenting subjects were registered in the Assessment Center (www.assessmentcenter.net), a free, secure online research management tool maintained at the Northwestern University Research Data Center.

Fourteen PROMIS CAT were selected for testing based on prior focus group studies in which patients with SLE identified quality of life domains of critical importance to them^{14,15,19}. Administered CAT included physical

function (version 1.2), mobility (v1.2), pain behavior (v1.0), pain interference (v1.1), ability to participate in social roles (v2.0), satisfaction with social roles and activities (v2.0), fatigue (v1.0), sleep disturbance (v1.0), sleep-related impairment (v1.0), applied cognition-abilities (v1.0), applied cognition-general concerns (v1.0), anger (v1.1), anxiety (v1.0), and depression (v1.0)²⁰. PROMIS items ask about the 7 preceding days, with the exception of items in the physical and social health domains, which do not specify a recall time frame. CAT were programmed to administer enough items to achieve a standard error (precision estimate) of ≤ 0.3 , with a minimum of 4 to a maximum of 12 items per CAT.

Patients completed 2 legacy PRO measures: the SF-36 standard, US version 1.0, a frequently used generic PRO instrument validated for use in SLE clinical trials, and the LupusQoL-US, an extensively validated SLE-specific PRO questionnaire adapted for use in the United States^{9,21}. Both legacy instruments refer to a 4-week recall period.

All self-report questionnaires were administered through the Assessment Center and all participants completed both PROMIS CAT and legacy instruments. Half the participants were randomly assigned to complete PROMIS CAT first, and the other half completed legacy PRO instruments first.

To assess PROMIS CAT test-retest reliability, all participants were contacted by telephone or e-mail within 1 week of enrollment to complete PROMIS CAT a second time. A 7-point Likert scale anchor question was used to identify any changes in patients' disease activity. Only patients reporting that the effect of SLE on their general health was "about the same" were included in the test-retest analysis because their PRO should not have changed.

PROMIS CAT were scored through the Assessment Center using a T score metric, in which the mean T score in the US general population is 50 with an SD of 10. Higher T scores reflect more of the trait being measured, so that higher scores for physical and social function are desirable, whereas higher symptom scores (e.g., fatigue, depression, anxiety) indicate a greater burden of symptoms. The SF-36 is divided into 8 scales, each with a score ranging from 0 to 100, with higher scores reflecting better HRQOL. Scores can also be reported as the physical component summary (PCS) and mental component summary (MCS), in which related scales are grouped and reported as a single score, normalized to the general US population with a score of 50 representing the population mean. The LupusQoL contains 34 questions in 8 domains, with scores ranging from 0 to 100, with higher scores indicating better HRQOL.

Sociodemographic information including age, sex, race, and ethnicity were obtained by patient self-report. Disease activity and damage at the time of the study visit were assessed by the subject's treating rheumatologist using a physician's global assessment (PGA), the Safety of Estrogens in Lupus Erythematosus National Assessment-Systemic Lupus Erythematosus Disease Activity Index (SELENA-SLEDAI), and the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI)^{22,23}. The PGA ranges from 0 to 3, SELENA-SLEDAI scores range from 0 to 105, and SDI scores range from 0 to 46. Higher scores reflect greater disease activity and more end-organ damage.

Statistical analysis. Means and SD were calculated for continuous variables, and frequencies and percentages for categorical variables. Floor and ceiling effects for each instrument were analyzed by calculating the percentage of respondents achieving the minimum and maximum possible scores, respectively. Construct validity of PROMIS CAT was assessed through Spearman correlation coefficients (r) with legacy PRO instruments, with coefficients of at least 0.7 indicating good convergent validity²⁴. Correlations between PROMIS CAT and disease activity and damage measures were also evaluated with Spearman r . Test-retest reliability was evaluated in participants completing the questionnaires twice within the 7-day time frame. Agreement between scores for each questionnaire was assessed with an intraclass correlation coefficient (ICC)²⁵. ICC of at least 0.7 indicate acceptable test-retest reliability²⁶. All statistical analyses were performed with SAS version 9.3.

The study was reviewed and approved by the HSS Institutional Review Board (IRB# 14125).

RESULTS

The study questionnaires were completed by 204 patients with SLE (Table 1), with 164 (80%) completing them remotely. One hundred sixty-two subjects (79%) completed the retest within 1 week. Subjects were predominantly women (93%) with a mean (SD) age of 40.0 (13.2) years. They were racially and ethnically diverse: 38% identified as white, 30% black, 13% Asian, and 28% Hispanic or Latino. The average (SD) SELENA-SLEDAI score was 4.2 (3.5), indicating mild disease activity, though 19.6% were flaring as per SELENA-SLEDAI at the time of assessment. The mean (SD) SDI was 1.2 (1.7), consistent with minimal end-organ damage.

PROMIS CAT and legacy instrument score distributions are shown in Table 2. The mean CAT scores across all PROMIS domains were worse than the general population by an average of 0.6 SD. Mean SF-36 PCS and MCS scores were 1.3 and 0.7 SD worse than the general population. PROMIS CAT were generally normally distributed, except for pain behavior and fatigue, which had slight positive skews. Similarly, SF-36 scale scores were relatively normally distributed except for the physical function, role physical, and role emotional scales, which were positively skewed. All domains in the LupusQoL were positively skewed. The SF-36 had large floor and ceiling effects in the role physical and role emotional scales (23%–49%), while the LupusQoL had notable ceiling effects across all domains (6%–32%). PROMIS CAT had less significant floor or ceiling effects, with fewer than 5% of patients scoring the lowest or highest possible score in most domains.

The number of items and time per instrument are shown

Table 1. Baseline characteristics of study participants.

Variable	Participants, n = 204
Demographic characteristics	
Age, yrs, mean ± SD (range)	40.0 ± 13.2 (19–73)
Female, n (%)	189 (92.6)
Race, n (%)	
White	77 (37.7)
Black	61 (29.9)
Asian	26 (12.8)
Other	40 (19.6)
Ethnicity, n (%)	
Hispanic/Latino	58 (28.4)
SLE characteristics	
SLE disease duration, yrs, mean ± SD (range)	12.2 ± 8.8 (0–48)
PGA, 0–3, mean ± SD (range)	0.8 ± 0.6 (0–2.8)
SELENA-SLEDAI, 0–105, mean ± SD (range)	4.2 ± 3.5 (0–20)
Flare, n (%)	40 (19.6)
SDI, 0–46, mean ± SD (range)	1.2 ± 1.7 (0–8)

SLE: systemic lupus erythematosus; PGA: physician's global assessment; SELENA-SLEDAI: Safety of Estrogens in Lupus Erythematosus National Assessment–Systemic Lupus Erythematosus Disease Activity Index; SDI: Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index.

in Table 3. On average, PROMIS CAT administered 4 items per domain and the median time per CAT was 32 s.

Correlations of PROMIS CAT with legacy instruments. Correlations between PROMIS CAT and legacy instruments are shown in Table 4. PROMIS physical function and mobility CAT correlated strongly with the physical function domains in the SF-36 and LupusQoL ($r = 0.81$ – 0.86), and with the SF-36 PCS ($r = 0.75$ – 0.81). Correlations between PROMIS pain interference and legacy instrument pain domains were also strong ($r = -0.79$). PROMIS fatigue correlated more strongly with the corresponding domain in the LupusQoL ($r = -0.75$) than with the SF-36 vitality scale ($r = -0.67$). Similarly, in the domain of mental health, PROMIS anger, anxiety, and depression CATs showed strong correlations with the LupusQoL emotional health domain ($r = -0.69$ to -0.75), and more moderate correlations with all of the SF-36 mental health–related scales ($r = -0.49$ to -0.76). PROMIS social function CAT correlated moderately to strongly with the corresponding domains in the SF-36 and LupusQoL ($r = 0.55$ – 0.75). All correlations were statistically significant with $p < 0.001$.

There were no analogous legacy instrument domains with which to compare the 4 PROMIS CAT evaluating cognition and sleep. However, these CAT showed strong correlations with fatigue. Correlations between fatigue and sleep-related impairment and applied cognition-concerns were both 0.68, while correlations between sleep-related impairment and disturbance was 0.62, and applied cognition-abilities and concerns was -0.74 ($p < 0.001$ for all).

Correlations of PROMIS CAT with physician-derived measures. Correlations between PROMIS CAT and physician-derived measures of SLE disease activity and disease-related damage are shown in Table 5. Correlations were generally weak and nonsignificant, with the highest correlations observed between CAT in the domains of physical function and pain and the PGA and SDI ($r = 0.27$ – 0.37 , $p < 0.001$).

Test-retest reliability. Of the 162 participants who completed PROMIS CAT a second time within 7 days (average 6.9 days), 90 reported no change in the effect of SLE on their health. Among these 90 subjects, ICC were > 0.7 across all domains (Table 6).

DISCUSSION

To our knowledge, our study is the first to demonstrate the validity and reliability of PROMIS CAT in outpatients with SLE. PROMIS CAT showed strong correlations with the SF-36 and LupusQoL across analogous domains, supporting the construct validity of the PROMIS measures. PROMIS CAT also showed high test-retest reliability in participants self-reporting no change in the effect of SLE on their health.

Although to our knowledge no prior studies have evaluated PROMIS CAT in adults with SLE, there has been some work evaluating PROMIS short forms (i.e., PROMIS

Table 2. PROMIS CAT and legacy PRO instrument score distributions (n = 204).

Variables	Mean	SD	Median	25th Percentile	75th Percentile	Min	Percent Min Score	Max	Percent Max Score
PROMIS CAT¹									
Physical function	42.6	8.4	41.9	37.7	48.1	24.7	0.5	73.3	1.0
Mobility	42.7	8.4	41.7	36.8	47.9	22.9	0.5	60.0	9.8
Pain interference	57.1	10.2	58.2	52.2	63.6	38.7	15.2	80.1	0.5
Pain behavior	56.0	9.3	58.5	53.6	62.2	35.3	12.3	73.8	0.5
Fatigue	58.3	9.5	59.1	51.8	64.0	24.3	0.5	82.9	0.5
Anger	53.9	11.2	54.6	47.4	61.4	28.9	3.4	85.2	0.5
Anxiety	55.8	9.4	57.7	51.1	62.0	32.9	2.5	84.8	0.5
Depression	53.0	10.5	54.3	45.6	60.9	34.2	12.8	78.1	1.0
Ability to participate in social roles	46.8	9.2	45.6	41.7	51.2	21.5	0.5	67.5	6.9
Satisfaction with social roles	45.9	9.9	45.8	40.0	52.3	22.0	1.5	68.7	3.9
Sleep disturbance	56.6	10.7	56.5	51.8	63.0	26.3	2.0	83.8	1.5
Sleep-related impairment	57.7	10.4	59.2	51.9	64.3	36.2	1.5	83.1	0.5
Applied cognition–abilities	45.8	8.9	43.5	39.5	50.2	26.2	0.5	67.7	5.9
Applied cognition–concerns	39.0	11.2	54.6	35.1	44.9	14.5	8.8	69.2	1.0
SF-36²									
Physical function	57.5	28.7	60.0	35.0	85.0	0.0	2.0	100.0	8.0
Role physical	35.6	41.5	25.0	0.0	75.0	0.0	49.0	100.0	23.0
Bodily pain	55.7	26.5	55.0	35.0	78.0	0.0	2.9	100.0	28.8
Vitality	43.2	21.1	45.0	25.0	55.0	0.0	1.0	100.0	0.5
Mental health	65.7	20.1	68.0	52.0	80.0	4.0	0.5	100.0	3.0
Role emotional	50.1	43.2	66.7	0.0	100.0	0.0	35.3	100.0	35.3
Social function	56.9	26.5	56.3	37.5	75.0	0.0	4.9	100.0	11.8
Global health perceptions	41.4	21.1	40.0	27.5	55.0	0.0	2.0	100.0	1.5
PCS	36.7	11.5	35.4	28.0	47.0	8.0	0.5	61.0	0.5
MCS	43.3	11.0	45.3	34.9	52.0	16.5	0.5	64.8	0.5
LupusQoL³									
Physical health	60.8	26.8	59.4	40.6	82.8	0.0	1.0	100.0	7.8
Pain	57.9	29.5	58.3	33.3	83.3	0.0	4.4	100.0	15.2
Fatigue*	53.2	27.2	56.3	31.3	75.0	0.0	3.0	100.0	5.9
Emotional health	71.5	25.2	79.2	58.3	91.7	0.0	0.5	100.0	12.7
Planning	65.4	29.4	75.0	50.0	91.7	0.0	3.9	100.0	18.6
Burden to others	53.1	31.7	50.0	33.3	75.0	0.0	10.3	100.0	12.3
Body image**	67.0	28.6	75.0	50.0	89.6	0.0	3.3	100.0	20.0
Intimate relationships	67.1	32.9	75.0	50.0	100.0	0.0	8.3	100.0	31.9

¹ PROMIS CAT are scored by T score from 0 to 100 (higher signifies more of the measured trait), with a score of 50 corresponding to the general population mean. ² SF-36 subscales are scored from 0 to 100 (higher signifies better health-related quality of life). For the PCS and MCS, a score of 50 is equivalent to the general population mean. ³ LupusQoL domains are scored from 0 to 100 (higher signifies better health-related quality of life). * n = 203 because of missing data. ** n = 180, remaining participants replied “Not Applicable.” PROMIS: Patient Reported Outcomes Measurement Information System; CAT: computerized adaptive tests; PRO: patient-reported outcome; min: minimum; max: maximum; SF-36: Medical Outcomes Study Short Form-36; PCS: physical component summary; MCS: mental component summary.

questions administered as part of a standard questionnaire without using computerized adaptive testing) in patients with SLE. The PROMIS-29, a 29-question short form composed of items from 7 PROMIS item banks (physical function, fatigue, pain interference, anxiety, depression, sleep disturbance, and satisfaction with social roles), was administered to 333 patients with self-reported SLE recruited from patient advocacy organizations²⁷. PROMIS-29 domain scores were associated with self-reported disease severity, but the study did not validate cases of SLE or compare the PROMIS-29

with established legacy instruments. Katz, *et al* evaluated the PROMIS-29 in 240 patients with rheumatologist-diagnosed SLE, demonstrating convergent validity with domains of the SF-36, but also noted significantly larger ceiling effects in 5 of the 7 PROMIS-29 domains compared with the SF-36²⁸. In contrast, our study found similar convergent validity between PROMIS CAT and the SF-36, but significantly decreased floor and ceiling effects in PROMIS CAT, suggesting increased precision over both SF-36 and PROMIS-29 short forms.

Table 3. Items and time per instrument.

Instrument	No. Items Administered				Time, Minutes			
	Mean	Median	Min	Max	Mean	Median	Min	Max
SF-36	36	36	36	36	6.6	5.2	0.7	54.8
LupusQoL	34	34	34	34	5.9	4.6	0.4	48.9
PROMIS physical function	4.3	4	4	12	1.0	0.7	0.1	20.2
PROMIS mobility	5.3	4	4	12	0.8	0.6	0.1	4.0
PROMIS fatigue	4.3	4	4	12	0.8	0.5	0.1	14.1
PROMIS pain interference	5.4	4	4	12	0.8	0.6	0.1	12.0
PROMIS pain behavior	5.0	4	4	12	0.9	0.6	0.1	7.7
PROMIS anger	7.1	6	5	12	0.9	0.5	0.1	19.9
PROMIS anxiety	5.1	4	4	12	0.5	0.3	0.1	6.8
PROMIS depression	5.7	4	4	12	0.4	0.3	0.1	4.9
PROMIS ability to participate in social roles	4.8	4	4	12	0.9	0.5	0.1	16.9
PROMIS satisfaction with social roles	5.1	4	4	12	1.0	0.7	0.1	7.4
PROMIS sleep-related impairment	5.1	4	4	12	0.8	0.5	0.1	14.9
PROMIS sleep disturbance	5.5	4	4	12	0.7	0.5	0.1	5.3
PROMIS applied cognition-abilities	4.7	4	4	12	0.8	0.5	0.1	16.3
PROMIS applied cognition- general concerns	5.5	4	4	12	1.3	0.7	0.1	20.5
All PROMIS CAT	72.8	58	57	170	11.3	7.4	0.6	170.8

Min: minimum; max: maximum; SF-36: Medical Outcomes Study Short Form-36; PROMIS: Patient Reported Outcomes Measurement Information System; CAT: computerized adaptive tests.

Mahieu, *et al* evaluated the internal consistency of 7 PROMIS short forms (physical function, fatigue, pain interference, anxiety, depression, sleep disturbance, and sleep-related impairment) in 123 adults with SLE, finding strong internal consistency among the measures (Cronbach's alpha 0.91–0.98)²⁹. They also showed strong correlations between the PROMIS fatigue short form and the self-report Fatigue Severity Scale scores (Spearman $r = 0.84$). The authors found that physical activity, measured with an accelerometer, was positively associated with PROMIS physical function ($r = 0.33$) and negatively associated with pain interference ($r = -0.29$). While these results legitimize the use of PROMIS short forms in SLE, these pre-set groups of questions are much longer than CAT, which average only 4 items per domain. The ability of PROMIS CAT to decrease responder burden without compromising precision or reliability is a significant advantage over both legacy instruments and PROMIS short forms.

PROMIS instruments have also been evaluated in children with rheumatic disease, with pediatric item banks demonstrating construct validity in 228 children with juvenile idiopathic arthritis and pediatric short forms demonstrating construct validity and responsiveness in 100 children with SLE^{30,31}. Of note, PROMIS instruments were developed with the goal of creating single metrics to measure domains across the lifespan³². This unique advantage of PROMIS over legacy instruments is particularly important in SLE, which can begin in childhood and continue through adulthood with waxing and waning course.

In our study, evaluating PROMIS CAT in SLE, partici-

pants scored one-half SD or worse than the general population across most PROMIS CAT, with the largest differences in the domains of physical function, mobility, pain interference, fatigue, sleep-related impairment, and applied cognition-concerns. These findings are concordant with those of Mahieu, *et al*, who reported that subjects with SLE scored one-half SD worse than the general population in physical function, pain interference, fatigue, sleep disturbance, and sleep-related impairment short forms, which are scored using the same T scale metric as CAT²⁹. These findings also provide face validity because CAT scores should trend lower than average given the known lower HRQOL in patients with SLE³³.

While prior studies have suggested that PROMIS short forms appear to have good reliability and precision in patients with SLE, ours is the first study to compare the performance characteristics of PROMIS CAT with the SF-36 and LupusQoL, 2 legacy PRO instruments commonly used in clinical research. In our study, in contrast to legacy instruments, PROMIS CAT demonstrated a normal distribution across domains and had smaller floor and ceiling effects, with less than 5% of subjects scoring the lowest or highest possible score. Certain domains, notably pain interference, pain behavior, and depression, did exhibit clustering at the lowest observed score, suggesting that perhaps this score represents the “de facto” floor of the instrument. Similar minimum scores for these domains were observed in validations of the PROMIS in other rheumatic disease populations^{34,35}. The significant floor and ceiling effects observed in the SF-36 and LupusQoL are consistent with score distributions reported in

Table 4. Correlations between PROMIS CAT and legacy PRO instruments.

PROMIS CAT	Legacy Instrument/domain	Spearman r*
Physical function	SF-36/physical function	0.83
	SF-36/role physical	0.64
	SF-36/PCS	0.81
Mobility	LupusQoL/physical health	0.82
	SF-36/physical function	0.86
	SF-36/role physical	0.55
Pain interference	SF-36/PCS	0.75
	LupusQoL/physical health	0.81
	SF-36/bodily pain	-0.79
Pain behavior	LupusQoL/pain	-0.79
	SF-36/bodily pain	-0.69
Fatigue	LupusQoL/pain	-0.71
	SF-36/vitality	-0.67
Anger	LupusQoL/fatigue	-0.75
	SF-36/mental health	-0.62
Anxiety	SF-36/role emotional	-0.50
	SF-36/MCS	-0.62
	LupusQoL/emotional health	-0.69
Depression	SF-36/mental health	-0.69
	SF-36/role emotional	-0.49
	SF-36/MCS	-0.67
Ability to participate in social roles	LupusQoL/emotional health	-0.75
	SF-36/mental health	-0.76
Satisfaction with social roles	SF-36/role emotional	-0.60
	SF-36/MCS	-0.76
	LupusQoL/emotional health	-0.75
	SF-36/social function	0.72
	LupusQoL/planning	0.75
	SF-36/social function	0.60
	LupusQoL/planning	0.55

* All p values < 0.001. PROMIS: Patient Reported Outcomes Measurement Information System; CAT: computerized adaptive tests; PRO: patient-reported outcome; SF-36: Medical Outcomes Study Short Form-36; PCS: physical component summary; MCS: mental component summary.

other studies and may contribute to the variable responsiveness of the measures in longitudinal studies^{10,12,36}. PROMIS CAT are better able to discriminate among individuals at the extremes of the spectrum, and importantly, may be more sensitive to identifying change over time within individuals because their score distribution is less skewed relative to legacy instruments and the PROMIS-29.

Importantly, PROMIS CAT correlated poorly with physician-derived measures of SLE disease activity and damage, supporting the principle that PRO measures identify unique information^{37,38}. In SLE, where defining appropriate outcome measures for clinical trials remains challenging³⁹, the patient perspective is particularly important and needs to be reliably measured. Currently, the SF-36 MCS and PCS are often used to benchmark HRQOL in study populations. Our findings suggest that the MCS and PCS correlate moderately to strongly with domains relevant to patients with SLE; correlations with PROMIS CAT related to physical function and mental health ranged from 0.62 to 0.81. PROMIS CAT offer an improved method of accurately and efficiently measuring

Table 5. Correlations between PROMIS CAT and physician-derived measures.

PROMIS CAT	PGA	SELENA-SLEDAI	SDI
Physical function	-0.29*	-0.15	-0.27*
Mobility	-0.35*	-0.16	-0.31*
Pain behavior	0.31*	0.16	0.22
Pain interference	0.37*	0.22	0.20
Fatigue	0.26	0.19	0.05
Anger	0.26*	0.21	0.07
Anxiety	0.24	0.17	0.12
Depression	0.24	0.17	0.10
Ability to participate in social roles	-0.28*	-0.17	-0.13
Satisfaction with social roles	-0.22	-0.14	-0.08
Applied cognition-abilities	-0.24	-0.16	-0.12
Applied cognition-concerns	0.22	0.18	0.09
Sleep disturbance	0.25	0.10	0.11
Sleep-related impairment	0.16	0.13	0.10

* p value ≤ 0.001. All other p values were nonsignificant. PGA: physician's global assessment; SELENA-SLEDAI: Safety of Estrogens in Lupus Erythematosus National Assessment-Systemic Lupus Erythematosus Disease Activity Index; SDI: Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index.

Table 6. Test-retest reliability of PROMIS CAT (n = 90*).

PROMIS CAT	ICC (95% CI)	SEM
Physical function	0.86 (0.79-0.90)	3.17
Mobility	0.88 (0.82-0.92)	2.91
Pain behavior	0.80 (0.72-0.87)	4.37
Pain interference	0.86 (0.79-0.90)	4.01
Fatigue	0.84 (0.77-0.89)	3.94
Anger	0.72 (0.61-0.81)	5.87
Anxiety	0.78 (0.68-0.85)	4.46
Depression	0.86 (0.79-0.90)	3.92
Ability to participate in social roles	0.87 (0.81-0.91)	3.52
Satisfaction with social roles	0.78 (0.68-0.85)	5.17
Applied cognition-abilities	0.83 (0.75-0.88)	4.21
Applied cognition-concerns	0.83 (0.76-0.89)	5.73
Sleep disturbance	0.88 (0.82-0.92)	3.83
Sleep-related impairment	0.80 (0.71-0.86)	4.85

* No. participants reporting no change in effect of SLE on health at second assessment. PROMIS: Patient Reported Outcomes Measurement Information System; CAT: computerized adaptive tests; SEM: standard error of measurement; SLE: systemic lupus erythematosus.

patient-centered outcomes in SLE while at the same time allowing comparisons with the general population.

In addition to their well-established use in research, PRO measures have great potential to improve the clinical care of SLE. There is increasing recognition that measuring PRO may improve patient engagement and shared decision making^{40,41}. PROMIS CAT are well suited for use at the point of care in SLE because of their favorable performance characteristics and decreased responder burden. Further studies are necessary to evaluate barriers to and facilitators of implementing PROMIS CAT in the clinical care of SLE,

as well as the effect of regular PRO collection on patient engagement and outcomes.

Our study has many strengths, including its large and diverse cohort of subjects with SLE, all validated according to the ACR criteria. Patients had a range of disease activity, with almost one-fifth flaring. However, the majority of patients had mild disease, reflecting the reality of most outpatients with SLE. There was also a high rate of participation in the retest questionnaire, supporting the generalizability of our findings. The choice of which PROMIS CAT to administer was informed by literature review, ensuring inclusion of domains that had been identified as important by patients with SLE themselves, and a large number of CAT were administered.

Our study has certain limitations. To decrease responder burden, not all PROMIS domains were validated against gold standard instruments. For example, the sleep- and cognition-related CAT had no corresponding domains in the SF-36 and LupusQoL. Further studies should validate these CAT against relevant legacy instruments, including potentially the LupusPRO¹⁵, which includes both of these domains. Conversely, PROMIS lacks some domains that are present in the LupusQoL, including body image, planning, and intimate relationships, areas valued by patients with SLE. This points to a knowledge gap; these item banks need to be developed. In our study, PROMIS CAT were evaluated in outpatients; the validity among inpatients, who may have worse HRQOL and worse disease activity, may differ and will need to be analyzed. Importantly, our study is cross-sectional and future studies need to evaluate the longitudinal responsiveness of PROMIS CAT. Finally, our study was limited to English speakers. PROMIS item banks have been translated into many other languages (Spanish, German, Dutch, Chinese, etc.)⁴²; additional studies are needed to validate PROMIS CAT in non-English-speaking patients with SLE.

Our study is the first, to our knowledge, to demonstrate the validity and reliability of PROMIS CAT in outpatients with SLE. PROMIS CAT are an efficient method of evaluating HRQOL in patients with SLE. They provide an accurate metric for measuring relevant patient domains, and future work should evaluate their performance in both clinical research and routine clinical care.

ACKNOWLEDGMENT

The authors thank Rima Abhyankar and Kelly McHugh for their assistance with patient recruitment.

REFERENCES

1. Urowitz MB, Gladman DD, Tom BD, Ibañez D, Farewell VT. Changing patterns in mortality and disease outcomes for patients with systemic lupus erythematosus. *J Rheumatol* 2008;35:2152–8.
2. Alarcón GS, McGwin G, Brooks K, Roseman JM, Fessler BJ, Sanchez ML, et al; LUMINA Study Group. Lupus in minority populations: nature versus nurture. Systemic lupus erythematosus in three ethnic groups. XI. Sources of discrepancy in perception of disease activity: a comparison of physician and patient visual analog scale scores. *Arthritis Rheum* 2002;47:408–13.
3. McElhone K, Abbott J, Teh LS. A review of health related quality of life in systemic lupus erythematosus. *Lupus* 2006;15:633–43.
4. U.S. Food and Drug Administration. Guidance for industry systemic lupus erythematosus — developing medical products for treatment. [Internet. Accessed March 20, 2017.] Available from: www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm072063.pdf
5. European Medical League. Guideline on clinical investigation of medicinal products for the treatment of systemic lupus erythematosus and lupus nephritis. [Internet. Accessed March 20, 2017.] Available from: www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/03/WC500184889.pdf
6. Strand V, Gladman D, Isenberg D, Petri M, Smolen J, Tugwell P. Endpoints: consensus recommendations from OMERACT IV. *Outcome Measures in Rheumatology*. *Lupus* 2000;9:322–7.
7. Castelino M, Abbott J, McElhone K, Teh LS. Comparison of the psychometric properties of health-related quality of life measures used in adults with systemic lupus erythematosus: a review of the literature. *Rheumatology* 2013;52:684–96.
8. Holloway L, Humphrey L, Heron L, Pilling C, Kitchen H, Højbjerg L, et al. Patient-reported outcome measures for systemic lupus erythematosus clinical trials: a review of content validity, face validity and psychometric performance. *Health Qual Life Outcomes* 2014;12:116.
9. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
10. Kuriya B, Gladman DD, Ibañez D, Urowitz MB. Quality of life over time in patients with systemic lupus erythematosus. *Arthritis Rheum* 2008;59:181–5.
11. Touma Z, Gladman DD, Ibañez D, Urowitz MB. Is there an advantage over SF-36 with a quality of life measure that is specific to systemic lupus erythematosus? *J Rheumatol* 2011;38:1898–905.
12. Devilliers H, Amoura Z, Besancenot JF, Bonnotte B, Pasquali JL, Wahl D, et al. Responsiveness of the 36-item Short Form Health Survey and the Lupus Quality of Life questionnaire in SLE. *Rheumatology* 2015;54:940–9.
13. Danoff-Burg S, Friedberg F. Unmet needs of patients with systemic lupus erythematosus. *Behav Med* 2009;35:5–13.
14. McElhone K, Abbott J, Gray J, Williams A, Teh LS. Patient perspective of systemic lupus erythematosus in relation to health-related quality of life concepts: a qualitative study. *Lupus* 2010;19:1640–7.
15. Jolly M, Pickard AS, Block JA, Kumar RB, Mikolaitis RA, Wilke CT, et al. Disease-specific patient reported outcome tools for systemic lupus erythematosus. *Semin Arthritis Rheum* 2012; 42:56–65.
16. McElhone K, Abbott J, Shelmerdine J, Bruce IN, Ahmad Y, Gordon C, et al. Development and validation of a disease-specific health-related quality of life measure, the LupusQoL, for adults with systemic lupus erythematosus. *Arthritis Rheum* 2007;57:972–9.
17. HealthMeasures. Intro to PROMIS. [Internet. Accessed March 20, 2017.] Available from: www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis
18. Hochberg MC. Updating the American College of Rheumatology Revised criteria for the classification of Systemic Lupus Erythematosus. *Arthritis Rheum* 1997;40:1725.
19. Ow YL, Thumboo J, Cella D, Cheung YB, Yong Fong K, Wee HL. 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.
20. HealthMeasures. List of adult measures. [Internet. Accessed March 20, 2017.] Available from: www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis/list-of-adult-measures

21. Jolly M, Pickard AS, Wilke C, Mikolaitis RA, Teh LS, McElhone K, et al. Lupus-specific health outcome measure for US patients: the LupusQoL-US version. *Ann Rheum Dis* 2010;69:29–33.
22. Petri M, Kim MY, Kalunian KC, Grossman J, Hahn BH, Sammaritano LR, et al; OC-SELENA Trial. Combined oral contraceptives in women with systemic lupus erythematosus. *N Engl J Med* 2005;353:2550–8.
23. Gladman D, Ginzler E, Goldsmith C, Fortin P, Liang M, Urowitz M, 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;39:363–9.
24. Campbell D, Fiske D. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 1959;56:81–105.
25. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–8.
26. Hays R, Revicki DA. Reliability and validity (including responsiveness). In: Fayers P, Hays R, eds. *Assessing quality of life in clinical trials*. New York: Oxford University Press; 2005:25–39.
27. Lai JS, Beaumont JL, Jensen SE, Kaiser K, Van Brunt DL, Kao AH, et al. An evaluation of health-related quality of life in patients with systemic lupus erythematosus using PROMIS and Neuro-QoL. *Clin Rheumatol* 2017;36:555–62.
28. Katz P, Pedro S, Michaud K. Performance of the PROMIS 29-Item Profile in rheumatoid arthritis, osteoarthritis, fibromyalgia, and systemic lupus erythematosus. *Arthritis Care Res* 2016 Dec 28 (E-pub ahead of print).
29. Mahieu MA, Ahn GE, Chmiel JS, Dunlop DD, Helenowski IB, Semanik P, et al. Fatigue, patient reported outcomes, and objective measurement of physical activity in systemic lupus erythematosus. *Lupus* 2016;25:1190–9.
30. Brandon TG, Becker BD, Bevans KB, Weiss PF. Patient-Reported Outcomes Measurement Information System tools for collecting patient-reported outcomes in children with juvenile arthritis. *Arthritis Care Res* 2017;69:393–402.
31. Jones JT, Carle AC, Wootton J, Liberio B, Lee J, Schanberg LE, et al. Validation of Patient-Reported Outcomes Measurement Information System short forms for use in childhood-onset systemic lupus erythematosus. *Arthritis Care Res* 2017;69:133–42.
32. Khanna D, Krishnan E, Dewitt EM, Khanna PP, Spiegel B, Hays RD. The future of measuring patient-reported outcomes in rheumatology: Patient-Reported Outcomes Measurement Information System (PROMIS). *Arthritis Care Res* 2011;63 Suppl 11:S486–90.
33. Jolly M. How does quality of life of patients with systemic lupus erythematosus compare with that of other common chronic illnesses? *J Rheumatol* 2005;32:1706–8.
34. Khanna D, Maranian P, Rothrock N, Cella D, Gershon R, Khanna PP, et al. Feasibility and construct validity of PROMIS and “legacy” instruments in an academic scleroderma clinic. *Value Health* 2012;15:128–34.
35. Bartlett SJ, Orbai AM, Duncan T, DeLeon E, Ruffing V, Clegg-Smith K, et al. Reliability and validity of selected PROMIS measures in people with rheumatoid arthritis. *PLoS One* 2015;10:e0138543.
36. McElhone K, Abbott J, Sutton C, Mullen M, Lanyon P, Rahman A, et al. Sensitivity to change and minimal important differences of the LupusQoL in patients with systemic lupus erythematosus. *Arthritis Care Res* 2016;68:1505–13.
37. Gladman DD, Urowitz MB, Ong A, Gough J, MacKinnon A. Lack of correlation among the 3 outcomes describing SLE: disease activity, damage and quality of life. *Clin Exp Rheumatol* 1996;14:305–8.
38. McElhone K, Castelino M, Abbott J, Bruce IN, Ahmad Y, Shelmerdine J, et al. The LupusQoL and associations with demographics and clinical measurements in patients with systemic lupus erythematosus. *J Rheumatol* 2010;37:2273–9.
39. Thanou A, Merrill JT. Top 10 things to know about lupus activity measures. *Curr Rheumatol Rep* 2013;15:334.
40. Lavalley DC, Chenok KE, Love RM, Petersen C, Holve E, Segal CD, et al. Incorporating patient-reported outcomes into health care to engage patients and enhance care. *Health Aff* 2016;35:575–82.
41. Basch E. Patient-reported outcomes - harnessing patients' voices to improve clinical care. *N Engl J Med* 2017;376:105–8.
42. HealthMeasures. Available translations. [Internet. Accessed March 20, 2017.] Available from: www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis/available-translations