Validity and reliability of patient reported outcomes measurement information system computerized adaptive tests in systemic lupus erythematous

Mitra Moazzami[1], Patricia Katz[2] , Dennisse Bonilla[3], Lisa Engel[4], Jiandong Su[5], Pooneh Akhavan[6], Nicole Anderson[3], Oshrat E Tayer-Shifman[7], Dorcas Beaton[8] and Zahi Touma[9]

Abstract

Background: The evaluation of Patient Reported Outcomes Measurement Information System (PROMIS) computerized adaptive test (CAT) in adults with systemic lupus erythematous (SLE) is an emerging field of research. We aimed to examine the test–retest reliability and construct validity of the PROMIS CAT in a Canadian cohort of patients with SLE.
Methods: Two hundred twenty-seven patients completed 14 domains of PROMIS CAT and seven legacy instruments during their clinical visits. Test–retest reliability of PROMIS was evaluated 7–10 days from baseline using intraclass correlation coefficient (ICC (2; 1)). The construct validity of the PROMIS CAT domains was evaluated against the commonly used legacy instruments, and also in comparison to disease activity and disease damage using Spearman correlations. A multitrait-multimethod matrix (MMM) approach was used to further assess construct validity comparing selected 10 domains of PROMIS and SF-36 domains.
Results: Moderate to excellent reliability was found for all domains (ICC [2;1] ranging from lowest, 0.66 for Sleep Disturbance and highest, 0.93 for the Mobility domain). Comparing seven legacy instruments with 14 domains of PROMIS CAT, moderate to strong correlations (0.51–0.91) were identified. The average time to complete all PROMIS CAT domains was 11.7 min. The MMM further established construct validity by showing moderate to strong correlations (0.55–0.87) between select PROMIS and SF-36 domains; the average correlations from similar traits (convergent validity) were significantly greater than the average correlations from different traits.
Conclusions: These results provide evidence on the reliability and validity of PROMIS CAT in SLE in a Canadian cohort.

[1]Department of Medicine, The George Washington University School of Medicine and Health Sciences, Washington, DC, USA

[2]Department of Medicine and Health Policy, University of California San Francisco, San Francisco, CA, USA

[3]Division of Rheumatology, Department of Medicine, Toronto Western Hospital-Lupus Clinic, University Health Network, Toronto, ON, Canada

[4]Department of Occupational Therapy, College of Rehabilitation Sciences, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada

[5]Centre for Prognosis Studies in the Rheumatic Diseases, University Health Network, Toronto, ON, Canada

[6]Division of Rheumatology, Department of Medicine, Mount Sinai Hospital, University of Toronto Faculty of Medicine, Toronto, ON, Canada

[7]Rheumatology Service, Meir Medical Center, Kfar-Saba, Israel

[8]Health Measurement, St. Michael's Hospital, University of Toronto, Institute for Work and Health, Toronto, ON, Canada

[9]Division of Rheumatology, Department of Medicine; Toronto Western Hospital-Lupus Clinic; Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

Corresponding author:
Zahi Touma, Division of Rheumatology, Department of Medicine; Toronto Western Hospital-Lupus Clinic; Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada.
Email: Zahi.Touma@uhn.ca

Key Messages

1. The data from this SLE cohort demonstrate that the PROMIS CAT domains have moderate to excellent test–retest reliability and moderate to strong construct validity compared to all seven legacy instruments.

2. Using the multitrait-multimethod matrix, six a priori hypotheses were confirmed showing moderate to strong correlations between select PROMIS and SF-36 domains (convergent validity).

3. The PROMIS CAT instruments seem feasible for many clinical environments as the median time to completion of all 14 PROMIS CAT domains was 11.7 min without additional time needed to score.

Introduction

Systemic lupus erythematous (SLE) is a multi-organ autoimmune disease with significant impact on overall health-related quality of life (HRQoL).[1–3] Studies have shown that physician assessments of disease burden do not always align with patient reports.[4] As a result, patient-reported outcome (PRO) measurement tools have become invaluable in clinical practice and are central to patient-centered care. PROs have the potential to increase participation of patients in their medical care and to facilitate earlier identification and access to mental health and social support.[2]

Currently, there are multiple generic and SLE-specific questionnaires that have validity evidence in SLE to evaluate HRQoL called legacy instruments.[5,6] Specifically, these instruments include the following: The Medical Outcomes Study Short-Form 36 Health Survey (SF-36),[7] Lupus Quality of Life (LupusQoL),[8] Beck Anxiety Inventory (BAI),[9] The Perceived Deficits Questionnaire (PDQ-20),[10] Beck Depression Scale-second edition (BDI-II),[11,12] The Assessment of Chronic Illness Therapy Fatigue Scale (FACIT-F),[13] and Epworth Sleepiness Scale (ESS).[14] Often these legacy instruments are administered on paper, and scoring each separately can become cumbersome at point of care.[15]

The Patient Reported Outcome Measurement Information System (PROMIS) was created by the National Institute of Health in 2004 to describe and evaluate physical, mental and social health.[16–19] It purports to have widespread utility in both the general population and in individuals living with chronic conditions.[16–18,20] PROMIS instruments have been shown to capture the experiences of patients across the broad continuum of symptoms and function, especially at low disease activity levels in a variety of rheumatological illnesses.[21]

A derivative of PROMIS tools based on item response theory (IRT), computerized adaptive tests (CAT) have been shown to efficiently (i.e., feasibility), accurately (i.e., validity) and precisely (i.e., reliability) incorporate patient self-report of health into research, potentially reducing research costs.[22,23] Administered on an iPad or computer platform,

PROMIS CAT has the ability to incorporate multiple domains of health. Current literature suggests that CAT-based assessments have the ability to modify the number of questions based on previous answers leading to reductions in responder burden.[21,22] PROMIS CAT instruments or domains incorporate patient self-report of health into research, allowing for reduced number of measurements and therefore, reduced research costs.[23,24]

Initial studies of the PROMIS CAT instruments provide promising measurement property evidence supportive of use in English speaking North American SLE populations. Kasturi et al. (2017) reported that in 204 patients with SLE, 10 PROMIS CAT domains had moderate to strong correlations ($\rho = -0.49$ to 0.86, $p < 0.001$) with SF-36, and LupusQoL-US version and moderate to good test-retest reliability (ICC were >0.7 across all domains).[25] Despite the supportive initial evidence, gaps in evidence exist and research is still needed to endorse the use of the PROMIS CAT instruments in SLE populations. No study to date has compared all 14 domains of PROMIS CAT with legacy instruments in patients with SLE in Canada. Furthermore, no study has used multitrait-multimethod (MMM) analysis as proposed by Campbell and Fiske to examine the validity of similar and different traits of PROMIS CAT and legacy instruments simultaneously for use in SLE.[26] Creating a multi-dimensional measurement model has the added benefit of further examining whether PROMIS actually measures the given traits it purports to evaluate.[26] Having more comprehensive reliability and validity evidence for a PRO tool to assist rheumatologists, the health care team and patients in the assessment and early identification of physical, mental, and social concerns is crucial for optimal disease management of patients with SLE.

Hence, this study aims to: (1) compare time to completion of the PROMIS CAT domains with legacy instruments; (2) assess intra-rater test–retest reliability of PROMIS CATs; (3) evaluate construct validity of PROMIS CATs compared to legacy PRO measures, in addition to SLE disease activity and organ damage metrics; and (4) evaluate convergent validity of PROMIS CAT in relation to SF-36 using an MMM analysis comparing similar and different traits.

Materials and methods

Study design

This was a prospective observational study whereby participants were asked to complete selected legacy instruments and the PROMIS CAT on a laptop or tablet in clinic or at home. Patients were also asked if they would complete the PROMIS CAT 7–10 days after baseline at home to study intra-rater test–retest reliability. Of 227 patients, 38% ($n = 87$) agreed to complete the PROMIS CAT again 7–10 days after baseline at home. This study was approved by the hospital-based research institutional ethics board.

Recruitment and sample

All consecutive English-speaking adults (≥18 years old) with SLE receiving care at a Toronto Lupus Center between July 2018 and Jan 2020 were screened for potential inclusion.

To be included in the study, recruited patients fulfilled at least four of the ACR revised criteria for the classification of SLE or three ACR criteria and a biopsy (lupus nephritis and cutaneous lupus).[27] Patients consented to participate at their regular appointments. Patients either agreed to complete the questionnaires in clinic on an iPad or computer or at home accessed through an email.

Measures

Patients completed the PROMIS CAT during their clinical visit and 7 days later at home (for test–retest intra-rater reliability) assessing 14 domains of health, specifically: Physical Function (V2.0), Applied Cognitive Abilities (V2.0), Applied Cognitive General (V2.0), Mobility (V2.0), Pain Behavior (V2.0), Pain Interference (V1.1), Ability to Participate in Social roles (V2.0), Satisfaction with Social Roles (V2.0), Sleep Disturbance (V1.0), Sleep Related Impairment (V1.0), Fatigue (V1.0), Anger (V1.1), Anxiety (V1.0), and Depression (V1.0).

The PROMIS CAT is scored using T score metrics where the mean ± standard deviation T score is 50 ± 10 in the US general population. Higher T scores reflect more of a particular domain. For some domains, a higher score is more desirable such as Physical Function, but for others it is less desirable such as Fatigue. The number of questions administered for each PROMIS CAT domain ranges from 4 to 12. Questions within domains are based on a 7-day recall period except for the physical and social health domains which do not specify a recall time frame.

Patients were also asked to complete: The SF-36,[7] LupusQoL,[8] Beck Anxiety Inventory (BAI),[9] The Perceived Deficits Questionnaire (PDQ-20),[10] Beck Depression Scale-second edition (BDI-II),[11,12] The Assessment of Chronic Illness Therapy Fatigue Scale (FACIT-F),[13] and Epworth Sleepiness Scale[14] (see Supplement 1 for details on scores and recall periods for each). High scores on SF-36 in any domain indicate better health.[7,19] Disease activity was quantified using the Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K).[28] SLEDAI-2K ranges from 0 (no disease activity) to 105 (most organ involvement). SLE-related organ damage was quantified using Systemic Lupus International Collaborating Clinics (SLICC)/ACR Damage Index (SDI).[27,29] SDI ranges from 0 (no organ damage) to 46 (most organ damage).

Statistical analysis

Sample characteristics. Study participants socio-demographic and disease related information was collected at baseline

including: age at study enrolment, SLE disease duration at study enrolment, sex, self-identifying ethnicity and highest education at enrolment. The mean and standard deviation (SD) for continuous variables and frequency (percent) for categorical variables) was calculated. Differences between patients who agreed to participate (participants) and those who declined to participate (non-participants) were evaluated by examining clinical and socio-demographic characteristics using un-paired t-test for continuous variables, chi-square test for binary and Cochran–Armitage trend test for categorical variables. Baseline T scores for PROMIS domains are reported as mean ± SD.

Instrument interpretability and feasibility. Floor and ceiling effects for each instrument, an indicator of instrument score interpretability was determined by calculating the percentage of respondents with the minimum and maximum scores (i.e., floor and ceiling effects, respectively). Floor or ceiling effects were considered meaningful when >15% of respondents scored at extremes.[30] Median time to completion for each PROMIS domain and the legacy instruments were displayed in seconds and were used as an indicator of instrument feasibility or usability.

Intra-rater test–retest reliability. Test–retest reliability was evaluated with intra-class correlation coefficient (ICC (2; 1)) analysis of PROMIS CAT scores at baseline and 7–10 days after baseline.[31] ICC (2; 1) values were interpreted in the following way: < 0.5 was indicative of poor reliability, between 0.5 and 0.75 indicated moderate reliability, between 0.75 and 0.9 indicated good reliability, and >0.9 indicated excellent reliability.[31]

Construct validity. PROMIS CAT 14 domains were evaluated against the seven commonly used legacy instruments (The SF-36,[7] LupusQoL,[8] Beck Anxiety Inventory [BAI],[9] The Perceived Deficits Questionnaire [PDQ-20],[10] Beck Depression Scale-second edition [BDI-II],[11,12] The Assessment of Chronic Illness Therapy Fatigue Scale [FACIT-F],[13] and Epworth Sleepiness Scale,[14]) in addition to SLEDAI-2K and SDI. We hypothesized that most legacy instruments would have at least a moderate correlation with corresponding PROMIS domain and that SLEDAI-2K and SDI would have a weak correlation with PROMIS domains.

Construct validity was also evaluated using the MMM approach comparing a similar construct using two different tools (in our case PROMIS-CAT and SF-36).[32–34] 10 PROMIS-CAT domain scores (Physical Function, Mobility, Pain Behavior, Pain Interference, Fatigue, Anger, Anxiety, Depression, Ability to Participate in Social Roles, Satisfaction with Social Roles) and SF-36 domains scores (Physical Function, Role Physical, Bodily Pain, Vitality, Role Emotional, General Health, Mental Health, Social Function) were compared. We hypothesized that similar traits (convergent validity) on average would have higher correlations than different traits between SF-36 and PROMIS-CAT domains (similar and different trait correlations are represented in

Table 1. Group Characteristics in comparison to non-participants.

Variable	Value	Participants (*n* = 227)	Non-participants (*n* = 187)	*p*-value
Age (years) at study enrolment	Mean ± SD	48.6 ± 14.1	53.3 ± 14.3	0.001
SLE disease duration at study enrolment (years)	Mean ± SD	18.5 ± 12.4	19.6 ± 12.3	0.37
Female	Yes *n* (%)	205 (90.3%)	172 (92.0%)	0.55
Ethnicity	Black *n* (%)	31 (13.7%)	48 (25.7%)	0.005
	White *n* (%)	144 (63.4%)	101 (54.0%)	
	Asian *n* (%)	25 (11.0%)	11 (5.9%)	
	Others *n* (%)	27 (11.9%)	27 (14.4%)	
Highest level of education	</= Grade 8 *n* (%)	1 (0.4%)	3 (1.6%)	0.07
	>Grade 8 *n* (%)	7 (3.1%)	7 (3.8%)	
	High school graduate *n* (%)	42 (18.8%)	53 (29.0%)	
	College *n* (%)	73 (32.6%)	55 (30.1%)	
	University *n* (%)	101 (45.1%)	65 (35.5%)	
SLEDAI-2K score	Mean ± SD	2.1 ± 2.3	2.8 ± 3.2	0.01
SLEDAI- 2K immunological score (dsDNA and complements)	Mean ± SD	1.4 ± 1.5	1.4 ± 1.6	0.59
Low complements or elevated dsDNA antibodies	Yes (%)	98 (51.9%)	84 (52.2%)	0.95
SDI	Mean ± SD	1.7 ± 1.9	1.8 ± 2.2	0.38
Anti-malarial	Yes (%)	140 (61.7%)	169 (90.4%)	<0.001
Prednisone % (mean dose)	Yes (%)	79 (34.8%)	159 (85.0%) (23.3 ± 18.6)	<0.001
	(mean ± std)	(8.3 ± 7.1)		<0.001
Immunosuppressants (%)	Yes (%)	94 (41.4%)	126 (67.4%)	<0.001

Supplement 2). Using the MMM matrix, construct validity was also evaluated using additional six a priori hypotheses to explore the relationships of PROMIS-CAT domains with corresponding SF-36 domains. We hypothesized that the following associations with at least moderate strength in the same direction would be found (Spearman correlation ρ>0.3): (1) PROMIS-CAT Physical Function and the SF-36 domains of Physical Function and Role Physical; (2) PROMIS-CAT Pain Behavior and Pain Interference and the SF-36 domain Bodily Pain; (3) PROMIS-CAT Anger, Anxiety, and Depression with SF-36 Role Emotional scores; (4) PROMIS-CAT Ability to Participate in Social Roles and Satisfaction with SF-36 Social Roles; (5) PROMIS-CAT Fatigue with SF-36 Vitality; (6) PROMIS-CAT Depression, Anger and Anxiety with SF-36 Mental Health. The values were interpreted in the following way: ρ <0.3 was indicative of weak association, ρ = 0.3–0.7 was indicative of moderate correlation, and ρ > 0.7 was indicative of strong correlation.[35]

Analytical software used in this project was SAS 9.3 (SAS Institute Inc., Cary, NC, USA.), and a *p* value <0.05 was regard as statistically significant.

Results

Sample characteristics

Four hundred and fourteen patients were approached, with 227 patients (55%) agreeing to participate in the study. 211

(93%) completed PROMIS CAT remotely, and 87 (38%) participants completed the retest within 7–10 days. Out of the 14 PROMIS domains, 11 domains had no missing data, the Physical Function domain had 13 (5.7%) missing, while Pain Behavior and Pain Interference had 2 (0.9%) data points missing. For the legacy instruments, less than 5% of the data were missing per domain except for ESS having 12 (5.3%) missing. Subjects were predominantly female (90.3%), with a mean age of 48.6 ± 14 years. Significant differences were found between participants and the 187 patients who declined participation in age, ethnicity, and SLEDAI-2K score (Table 1). Reasons cited by patients that did not want to participate included the following: no time, resistance to use technology, and no email address to administer survey at home among others. The mean disease activity by SLEDAI-2K in participants was 2.1 ± 2.3 and in non-participants was 2.8 ± 3.2 (*p* = 0.01). The mean damage by SDI in participants was 1.7 ± 1.9 and non-participants was 1.8 ± 2.2 (*p* = 0.38). A complete description of characteristic can be found in Table 1.

Instrument interpretability and feasibility

PROMIS CAT and legacy instrument score distributions, an indication of instrument interpretability, can be found in Supplement 1. Mean scores on all PROMIS CAT domains were significantly worse than the general population (*p* < 0.05), (participation social roles, satisfaction social roles, and anger were not statistically different from general population

Table 2. PROMIS CAT score distributions.[a]

Domain	Mean ± SD	Floor n (%)	Ceiling n (%)
Physical Function	43.9 ± 9.1	1 (0.5%)	1 (0.5%)
Mobility	44.9 ± 9.4	2 (0.9%)	**39 (17.2%)**
Pain Behaviour	53.3 ± 10.7	**47 (20.9%)**	1 (0.4%)
Pain Interference	54.9 ± 10.6	**47 (20.9%)**	1 (0.4%)
Fatigue	57.3 ± 10.3	2 (0.9%)	1 (0.4%)
Sleep Disturbance	55.4 ± 10.2	4 (1.8%)	1 (0.4%)
Sleep-Related Impairment	56.3 ± 10.3	2 (0.9%)	1 (0.4%)
Anger	51.3 ± 9.8	7 (3.1%)	1 (0.4%)
Anxiety	55.7 ± 8.9	3 (1.3%)	1 (0.4%)
Depression	52.4 ± 9.8	16 (7%)	1 (0.4%)
Ability to Participate in Social Roles	48.3 ± 9.6	1 (0.4%)	19 (8.4%)
Satisfaction with Social Roles	48.4 ± 11.0	2 (0.9%)	22 (9.7%)
Applied Cognitive Abilities	46.9 ± 9.4	2 (0.9%)	14 (6.2%)
Applied Cognitive General	46.5 ± 8.8	1 (0.4%)	9 (4.0%)

[a]Bold values are the only scales for which the criterion for floor or ceiling effects was met.
Mean scores on all PROMIS CAT domains were significantly worse than the general population ($p < 0.05$) (participation social roles, satisfaction social roles, and anger were not statistically different from general population after Bonferroni adjustments for multiple tests).

after Bonferroni adjustments for multiple tests). Ceiling effects were found for the Mobility domain of PROMIS CAT where 17.2% of the participants had the highest score. Floor effects were found with Pain Behavior and Interference whereby 20.9% of the population had the lowest score (Table 2).

Median time taken to complete PROMIS and legacy questionnaires are displayed in Table 3. Each PROMIS domain took less time to complete than the corresponding domain on the legacy instruments. The only exception to this is that the median time to complete the Sleep Related Impairment domain was 87 s (IQR: 59.5–129.5) versus the median time to complete the ESS questionnaire was 72 s (IQR: 56–100). In total, all 14 PROMIS domains took 700.5 s (IQR: 50.75–1016.5) or 11.7 min (IQR: 0.84–16.9) to complete (Table 3). SF-36 (8 domains) took 317 s (IQR: 236–460.5) to complete, LupusQoL (8 domains) took 264 s (IQR: 191–411), while the other five legacy instruments together (BAI, PDQ-20, BDI-II, FACIT-F, ESS) took 567 s (IQR: 421–796) to complete. Total time for all legacy instruments was 22.13 min (IQR: 16.75–31.7).

Test–retest reliability

Eighty-seven (38%) participants completed the retest within 10 days. Moderate to excellent reliability was found for all domains (ICC (2; 1) range 0.66–0.93). The lowest ICC (2; 1) were identified for Sleep Disturbance (ICC (2;1) 0.66; 95% CI: 0.51, 0.80) and Satisfaction with Social Roles (ICC (2;1) 0.70; 95% CI: 0.45, 0.89) (Table 4).

Construct validity

Table 6 explores the relationships between PROMIS CAT domains and the corresponding legacy instruments, in

addition to disease activity (SLEDAI-2K) and damage (SDI). Moderate to strong correlations ($\rho = 0.51$–0.91) were found for each PROMIS CAT domains with most closely corresponding legacy instrument domains (Table 5). The only legacy instrument that did not correlate as highly with PROMIS CAT was ESS ($\rho = 0.19$ with Sleep Disturbance and $\rho = 0.33$ with Sleep Impairment). The highest correlations were found between SF-36 Physician Function and Mobility of PROMIS CAT ($\rho = 0.91$) and Physical Function of PROMIS CAT ($\rho = 0.87$). Correlations between SLEDAI-2K and SDI were generally weak.

Using the MMM, construct validity was tested again. All a priori developed hypotheses in Supplement 2 were satisfied. As we hypothesized, PROMIS CAT demonstrated significant correlations with legacy instruments in comparable domains and weaker correlations with domains of different traits (Table 6). Furthermore, the average correlations of similar traits between the 10 domains of PROMIS CAT and SF-36 were greater than the average correlations from different traits confirming our hypotheses. All six a priori hypotheses were satisfied with moderate to strong correlations (Spearman correlation, $\rho = 0.55$–0.87) between PROMIS-CAT and most SF-36 domains.

1. Patients with lower Physical Function (PF) scores on PROMIS CAT also have lower Physical Function scores across the two related SF-36 domains (Physical Function (PF) and Role Physical (RF)) with at least a moderate correlation ($\rho > 0.3$).
   PROMIS$_{PF}$/SF-36$_{PF}$ ($\rho = 0.87$, $p < 0.0001$)
   PROMIS$_{PF}$/SF-36$_{RF}$ ($\rho = 0.76$, $p < 0.0001$)
2. Patients that scored higher on Pain Behavior (PB) and Pain Interference (PI) on PROMIS CAT scored

Table 3. Demonstration of average number of items administered per questionnaire and median time to completion.

	No. Items patients answered in each domain mean (SD)	Median (IQR) time (in seconds) to complete the question
PROMIS CAT		
Physical Function	4.3 (1.6)	87 (59.5, 129.5)
Pain Interference	5.9 (3.3)	45 (32, 68)
Pain Behavior	5.9 (3.3)	4 (4, 6)
Fatigue	4.2 (1.1)	41.5 (30.5, 60)
Mobility	5.9 (3.3)	49 (33, 78)
Depression	5.4 (2.6)	27 (19.5, 45.5)
Anxiety	4.8 (1.9)	29 (23, 46)
Anger	7.1 (2.3)	49 (29.5, 69)
Ability to Participate in Social Roles	5.0 (2.4)	43 (31, 58)
Satisfaction with Social Roles	5.5 (2.9)	46.5 (32.5, 66)
Applied Cognitive General	5.7 (3.0)	53 (38, 87.5)
Applied Cognitive Abilities	5.3 (2.6)	45 (30, 68.5)
Sleep Disturbance	5.1 (2.2)	39.5 (27.5, 58)
Sleep-Related Impairment	5.1 (2.4)	87 (59.5, 129.5)
Total PROMIS	75.3 (21.1)	700.5 (50.7.5, 1016.5)
The Medical Outcomes Study Short-Form 36 Health Survey (SF-36)	36	317 (236, 460.5)
Lupus Quality of Life (LupusQoL)	34	264 (191, 411)
Beck Anxiety Inventory (BAI)	21	112 (85, 148)
The Perceived Deficits Questionnaire (PDQ-20)	20	136 (103, 182)
Beck Depression Scale-2nd edition (BDI-II)	21	236 (159, 340)
The assessment of chronic illness therapy fatigue scale (FACIT-F)	13	85 (64, 119)
Epworth sleepiness scale (ESS)	8	72 (56, 100)
Total for BAI, PDQ-20, BDI-II, FACIT-F, ESS	83	567 (421, 796)
Total for all legacy instrument	153	1328 (1005, 1902)

Table 4. Test–retest reliability of PROMIS CAT (*n* = 87).

Domain	ICC (2, 1) and 95% CI
Physical Function	0.91 (0.85, 0.95)
Mobility	0.93 (0.89, 0.97)
Pain Behaviour	0.80 (0.59, 0.93)
Pain Interference	0.82 (0.69, 0.93)
Ability to Participate in Social Roles	0.85 (0.78, 0.92)
Satisfaction with Social Roles	0.70 (0.45, 0.89)
Fatigue	0.81 (0.71, 0.90)
Sleep Disturbance	0.66 (0.51, 0.80)
Sleep-Related Impairment	0.77 (0.66, 0.88)
Anger	0.74 (0.60, 0.85)
Anxiety	0.75 (0.62, 0.90)
Depression	0.86 (0.80, 0.91)
Applied Cognitive Abilities	0.71 (0.55, 0.85)
Applied Cognitive General	0.85 (0.77, 0.92)

lower (worse) Bodily Pain (BP) on SF36 with at least a moderate correlation ($\rho > 0.3$).
PROMIS$_{PB}$/SF-36$_{BP}$ ($\rho = -0.80$, $p < 0.0001$)
PROMIS$_{PI}$/SF-36$_{BP}$ ($\rho = -0.82$, $p < 0.0001$)

3. Patients that scored higher on Anger (ANG), Anxiety (ANX), or Depression (DEP) on PROMIS CAT, rated a lower Role Emotional (RE) health on SF-36 with at least a moderate correlation ($\rho > 0.3$).
PROMIS$_{ANG}$/SF-36$_{RE}$ ($\rho = -0.55$, $p < 0.0001$)
PROMIS$_{ANX}$/SF-36$_{RE}$ ($\rho = -0.61$, $p < 0.0001$)
PROMIS$_{DEP}$/SF-36$_{RE}$ ($\rho = -0.65$, $p \leq 0.0001$)

4. Patients with lower scores on Ability to Participate in Social Roles (APSR) and Satisfaction with Social Soles (SSR) on PROMIS CAT also rated lower on Social Function (SF) on SF-36 with at least a moderate correlation ($\rho > 0.3$).
PROMIS$_{APSR}$/SF-36$_{SF}$ ($\rho = 0.78$, $p = <0.0001$)
PROMIS$_{SSR}$/SF-36$_{SF}$ ($\rho = 0.71$, $p = <0.0001$)

5. Patients with higher Fatigue (FA) score on PROMIS CAT would have a lower Vitality (VT) score on SF-36 with at least a moderate correlation ($\rho > 0.3$).
PROMIS$_{FA}$/SF-36$_{VT}$ ($\rho = -0.85$, $p= < 0.0001$)

6. Patients with a higher Depression (DEP), Anxiety (ANX), or Anger (ANG) score on PROMIS CAT would have a lower Mental Health (MH) score on SF-36 with at least a moderate correlation ($\rho > 0.3$).

Table 5. Spearman Correlation, ρ^{a} between 14 domains of PROMIS CAT and seven legacy instruments in addition to SLEDAI-2K and SDI.

PROMIS CAT domains	Legacy instrument domains	Spearman correlation	SLEDAI-2K	SDI
Physical Function	SF-36 Physical function	0.87	0.20	−0.35
	SF-36 Role Physical	0.76		
	SF-36 Physical Component Summary	0.84		
	Lupus QoL Physical	0.82		
Mobility	SF-36 Physical Function	0.91	0.17	−0.39
	SF-36 Role Physical	0.70		
	SF-36 Physical Component Summary	0.83		
	Lupus QoL Physical	0.80		
Pain Behavior	SF-36 Bodily Pain	−0.80	−0.17	0.25
	Lupus QoL Pain	−0.77		
Pain Interference	SF-36 Bodily Pain	−0.82	−0.15	0.24
	Lupus QoL Pain	−0.76		
Fatigue	SF-36 Vitality	−0.85	**0.09**[a]	**0.07**[a]
	Lupus QoL Fatigue	−0.72		
	FACIT	0.82		
Anger	SF-36 Mental Health	−0.64	**−0.13**[a]	**0.02**[a]
	SF-36 Role Emotional	−0.55		
	SF-36 MCS	−0.64		
	Lupus QoL Emotional	−0.61		
Anxiety	SF-36 Mental Health	−0.78	**0.00**[a]	**−0.03**[a]
	SF-36 Role Emotional	−0.61		
	SF-36 MCS	−0.74		
	LupusQoL Emotional	−0.70		
	BAI	0.65		
Depression	SF-36 Mental Health	−0.80	**−0.02**[a]	**0.03**[a]
	SF-36 Emotional Role Functioning	−0.65		
	SF-36 MCS	−0.79		
	LupusQoL Emotional Health	−0.74		
	BECK depression	0.72		
Ability to Participate in Social Roles	SF-36 Social Function	0.78	**0.12**[a]	−0.25
	Lupus QoL Intimate Relationships	0.48		
	Lupus QoL Emotional Health	0.59		
Satisfaction with Social Roles	SF-36 Social Function	0.71	**0.12**[a]	−0.18
	Lupus QoL Intimate Relationships	0.47		
	Lupus QoL emotional health	0.59		
Applied Cognitive General	PDQ20	−0.80	0.16	−0.04[a]
	Lupus-QoL Planning	0.56		
Applied Cognitive Abilities	PDQ20	−0.74	0.21	−0.09[a]
	Lupus-QoL planning	0.52		
Sleep disturbance	ESS	0.19	−0.11[a]	0.12[a]
	Lupus QoL Fatigue	−0.51		
	SF36 Vitality	−0.53		
Sleep-Related Impairment	ESS	0.33	−0.11[a]	0.08[a]
	Lupus QoL Fatigue	−0.65		
	SF36 Vitality	−0.67		

[a]All values were significant at $p < 0.05$ except those in bold.

Table 6. Multitrait-multimethod matrix: SF-36 subscale and PROMIS domain Spearman correlations.

PROMIS	SF-36 domains								Average correlations of the similar traits in comparison to different traits[a]	
	PF	RF	BP	VT	RE	GH	MH	SF	Similar traits	Different traits
PF	**0.87**	**0.76**	0.65	0.61	0.63	0.56	0.41	0.62	0.82	0.58
M	**0.91**	**0.70**	0.62	0.55	0.58	0.50	0.36	0.58	0.81	0.53
PB	−0.75	−0.72	**−0.80**	−0.62	−0.65	−0.55	−0.47	−0.65	−0.80	−0.63
PI	−0.71	−0.76	**−0.82**	−0.65	−0.68	−0.59	−0.48	−0.68	−0.82	−0.65
FA	−0.61	−0.76	−0.65	**−0.85**	−0.67	−0.64	−0.54	−0.73	−0.85	−0.66
ANG	−0.37	−0.50	−0.46	−0.54	**−0.55**	−0.43	**−0.64**	−0.51	−0.55	−0.49
ANX	−0.34	−0.51	−0.40	−0.52	**−0.61**	−0.45	**−0.78**	−0.56	−0.61	−0.51
DE	−0.43	−0.58	−0.49	−0.60	**−0.65**	−0.47	**−0.80**	−0.67	−0.72	−0.54
APSR	0.76	0.79	0.64	0.68	0.71	0.61	0.47	**0.78**	0.78	0.67
SSR	0.68	0.74	0.60	0.69	0.70	0.58	0.54	**0.71**	0.71	0.65

Bold values represent expected convergent correlations. PF, Physical Function; M, Mobility; PB, Pain Behavior; PI, Pain Interference; FA, Fatigue; ANG, Anger; ANX, Anxiety; DE, Depression; APSR, Ability to Participate in Social Roles; SSR, Satisfaction with Social Roles; RF, Role Physical; BP, Bodily Pain; VT, Vitality; RE, Role Emotional; GH, General Health; MH, Mental Health; SF, Social Function.
[a]Average correlations of the similar trait versus different traits were compared using multitrait-multimethod approach.

PROMIS$_{DEP}$/SF-36$_{MH}$ (ρ = −0.80, p < 0.0001)
PROMIS$_{ANX}$/SF-36$_{MH}$ (ρ = −0.78, p < 0.0001)
PROMIS$_{ANG}$/SF-36$_{MH}$ (ρ = −0.64, p < 0.0001)

Discussion

The importance of PRO measurement tools has been previously emphasized in the literature.[2] The measurement evidence, including the interpretability, feasibility, reliability, and validity of the PROMIS Computerized Adaptive Tests (CAT), in a cohort of patient living with SLE in Canada was examined. This is the first study to use an MMM to evaluate construct validity and to test hypothesized relationships developed a priori. This study further provides evidence that PROMIS CAT can perform as well as legacy instruments, encompass many HRQoL domains, and will ultimately save time without the added work of individualized data management platform, grading, and paper forms.

To date, only one other study has examined the validity and reliability of PROMIS CAT in SLE.[25] Kasturi et al. (2017) examined 204 participants finding a moderate to good agreement for test–retest reliability in all their domains with ICC ranging between 0.72 (for Anger) and 0.88 (for Mobility and Sleep Disturbance).[25] In contrast, this study had excellent test–retest reliability for PROMIS CAT Mobility domain (ICC (2;1) 0.93; 95% CI: 0.89, 0.97) and Physical Function domain (ICC (2;1) 0.91; 95% CI: 0.85, 0.95), and moderate to good agreement for all other domains. It also had the lowest reliability for Sleep Disturbance (ICC (2;1) 0.66; 95% CI: 0.51, 0.80) and Satisfaction with Social Roles (ICC (2;1) 0.70; 95% CI: 0.45, 0.89), while Kasturi et al. (2017) found higher reliability in these domains.[25] This difference may be due to the time between the two measurements. This study allowed 7–10 days for test–retest, and Kasturi et al. (2017) allowed only 7 days.[25] Since these constructs are not very stable overtime, this increases the chances of variability with longer intervals.[36]

Similar to this study, Kasturi et al. (2017) found moderate to strong correlations between two legacy instruments (LupusQoL and SF-36) and 10 corresponding PROMIS CAT domains.[25] However, this study encompassed seven legacy instruments and also examined a greater number of PROMIS CAT domains including Cognition and Sleep.[14] Kasturi et al. (2017) similarly found that the majority of the associations between disease activity and damage and PROMIS CAT were weak.[25] This finding highlights the importance of incorporating PROs in research and clinical settings as disease activity and damage do not cover the entire spectrum of SLE.

Using an MMM, construct validity of PROMIS CAT was further solidified by testing six a priori hypothesis. It was demonstrated that correlations of similar traits on average have higher correlations than correlations of different traits between SF36 and PROMIS CAT. The PROMIS domain Anger had the lowest correlation with SF-36 Role

Emotional but still demonstrated a moderate correlation (ρ = −0.55).

Other studies have examined the time to completion of the PROMIS CAT domains in different rheumatic diseases. However, only one study has measured time to completion in PROMIS CAT in patients with SLE. Kasturi et al. (2017) noted that the average item per PROMIS CAT domain was four and the median time to completion of each PROMIS domain was 32 s.[25] For all 10 domains studied, they found the median time to completion was 7.4 min while for SF-36 it took 5.2 min and LupusQoL 4.6 min.[25] In this study, the average number of items per PROMIS CAT domain (total = 14 domains) was 5.4 (STD 2.5) and the median time to completion per domain was 50 s (IQR 36.3–72.6). It also took the participants about the same time, 5.2 min, to complete SF-36, 4.4 min, to complete Lupus QoL, and 22 min to complete all legacy instruments. In contrast, it took about 11.6 min to complete all 14 PROMIS domains. PROMIS CAT in this study included an additional four domains compared to Kasturi et al. (2017) which, combined with the higher number items per domain, explains the longer time to completion.[25] Overall, the 14 domains of PROMIS CAT take approximately 11 min less to complete than the legacy tests needed to assess the same domains, without the time needed for individualized grading.

PROMIS also provides static short forms ranging in length from four to eight items that can be used to assess different PROMIS domains as opposed to CAT. Multiple studies have evaluated PROMIS short forms (SF) in various rheumatologic conditions.[37–39] These studies all found at least moderate correlations between the PROMIS short form and their corresponding legacy instruments, as well as good test–retest reliability.[37–39]

This study was able to undercover several limitations to PROMIS CAT. PROMIS CAT showed floor or ceiling effects for three domains. Importantly, there were floor effects in one domain: Mobility, which may demonstrate that appropriate functional status may not be captured by this tool.[8] Further, the legacy instrument ESS had weak associations with the PROMIS CAT domain Sleep Disturbance but a moderate association with Sleep Related Impairment. In a previous study comparing PROMIS Sleep Related Impairment and Sleep Disturbance with ESS, only Sleep Related Impairment and Sleep Disturbance correlated with active SLE, whereas ESS did not.[40] The less robust associations with ESS may also be due to differences in instrument content. The ESS assesses daytime sleepiness, which is only one aspect of Sleep Disturbance covered in the PROMIS Sleep Disturbance scale. The Sleep Disturbance scale measures sleep quality, which may or may not have an impact on daytime sleepiness.

Limitations of this study include that consecutive patients from one clinic completed PROMIS CAT only in English, limiting generalizability and internal validity.

Secondly, although all participants were encouraged to participate, non-participants included a higher proportion of Black patients and those with higher levels of disease activity, greater use and higher doses of prednisone, and immunosuppressants. This shows that individuals with more severe burden of SLE disease activity, were less inclined to participate which should be considered when interpreting the measurement property evidence.

This study has four main strengths. First, this study collected data from a large number of patients. Second, this is the first study that examines PROMIS CAT in a Canadian cohort of patients with SLE, providing wider international evidence for instruments developed in the US. Third, the analysis used an MMM approach which can simultaneously measure correlations of similar and different traits. Lastly, it was able to compare seven legacy instruments with 14 domains of PROMIS CAT, allowing the incorporation of a wide range of HRQoL manifestations in this study.

Conclusions

This study provides further evidence that PROMIS CAT can be used as a PRO measurement tool for patients living with SLE to measure a wide range of domains associated with HRQoL. It has the capability of combining many different commonly assessed domains in an easy-to-use platform. In comparison to legacy instruments, it is able to perform well and has moderate to strong construct validity and moderate to excellent reliability for most domains. It takes approximately 11.6 min to complete PROMIS CAT with all its 14 domains without additional time needed for scoring. It can easily be incorporated as part of a regular outpatient clinic visit. The association between PROMIS CAT domains and disease activity and damage was mainly weak and non-significant. Future studies should focus on the responsiveness on PROMIS and score interpretability.

Ethical approval

This study was reviewed and approved by the University Health Network Research Ethics Board (UHN REB#18-5339) This study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. If doubt exists whether the research was conducted in accordance with the 1964 Helsinki Declaration and the Canadian Tri-Council Policy Statement Guidelines on research involving human subject, or comparable ethical standards.

Data availability

The dataset used and analyzed during the current study is available from the senior author on reasonable request.

ORCID iDs

Patricia Katz https://orcid.org/0000-0002-8146-2519
Zahi Touma https://orcid.org/0000-0001-5177-2076

Supplemental material

Supplemental material for this article is available online.

References

1. Kwan A, Marzouk S, Ghanean H, et al. (eds) *Seminars in arthritis and rheumatism. Assessment of the psychometric properties of patient-reported outcomes of depression and anxiety in systemic lupus erythematosus.* Elsevier, 2019.
2. Kwan A, Strand V and Touma Z. The role of patient-reported outcomes in systemic lupus erythematosus. *Curr Treat Options Rheumatol* 2017; 3(4): 308–321.
3. Nantes S, Su J and Touma Z. Responsiveness and its magnitude in the 36-item short form health survey and the lupus quality of life questionnaire in patients with active disease. *Arthritis Rheumatol*; 2016; 68.
4. Holloway L, Humphrey L, Heron 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(1): 116.
5. Strand V, Simon LS, Meara AS, et al. Measurement properties of selected patient-reported outcome measures for use in randomised controlled trials in patients with systemic lupus erythematosus: a systematic review. *Lupus Sci Med* 2020 Jun; 7(1): e000373.
6. Izadi Z, Gandrup J, Katz PP, et al. Patient-reported outcome measures for use in clinical trials of SLE: a review. *Lupus Sci Med* 2018; 5(1): e000279.

7. McHorney CA, Ware JE Jr, Rachel Lu JF, et al. The MOS 36-item short-form health survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994; 32: 40–66.

8. Nantes SG, Strand V, Su J, et al. Comparison of the sensitivity to change of the 36-item short form health survey and the lupus quality of life measure using various definitions of minimum clinically important differences in patients with active systemic lupus erythematosus. *Arthritis Care Res* 2018; 70(1): 125–133.

9. Eldeiry D, Zandy M, Tayer-Shifman O, et al. Association between depression and anxiety with skin and musculo-skeletal clinical phenotypes in systemic lupus erythematosus. *Rheumatology (Oxford)* 2000; 59(11): 3211–3220.

10. Vogel A, Bhattacharya S, Larsen J, et al. Do subjective cognitive complaints correlate with cognitive impairment in systemic lupus erythematosus? A Danish outpatient study. *Lupus* 2011; 20(1): 35–43.

11. Roebuck-Spencer TM, Yarboro C, Nowak M, et al. Use of computerized assessment to predict neuropsychological functioning and emotional distress in patients with systemic lupus erythematosus. *Arthritis Rheum* 2006; 55(3): 434–441.

12. McCormick N, Trupin L, Yelin EH, et al. Socioeconomic predictors of incident depression in systemic lupus erythematosus. *Arthritis Care Res* 2018; 70(1): 104–113.

13. Lai J-S, Beaumont JL, Ogale S, et al. Validation of the functional assessment of chronic illness therapy-fatigue scale in patients with moderately to severely active systemic lupus erythematosus, participating in a clinical trial. *J Rheumatology* 2011; 38(4): 672–679.

14. Iaboni A, Ibanez D, Gladman DD, et al. Fatigue in systemic lupus erythematosus: contributions of disordered sleep, sleepiness, and depression. *J Rheumatology* 2006; 33(12): 2453–2457.

15. Hung M, Stuart AR, Higgins TF, et al. Computerized adaptive testing using the PROMIS physical function item bank reduces test burden with less ceiling effects compared with the short musculoskeletal function assessment in orthopaedic trauma patients. *J Orthopaedic Trauma* 2014; 28(8): 439–443.

16. Burgos PI, Alarcón GS, McGwin G Jr, et al. Disease activity and damage are not associated with increased levels of fatigue in systemic lupus erythematosus patients from a multiethnic cohort: LXVII. *Arthritis Rheum* 2009 Sep 15; 61(9): 1179–1186,.

17. Riley WT, Pilkonis P and Cella D. Application of the National Institutes of Health Patient-reported Outcome Measurement Information System (PROMIS) to mental health research. *J Mental Health Policy Econ* 2011; 14(4): 201–208.

18. Cella D., Riley W., Stone A., et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol* 2010; 63(11): 1179–1194.

19. Measures H. *Intro to PROMIS*. Available from: www.healthmeasures.net/.

20. Bingham CO, Bartlett SJ, Merkel PA, et al. Using patient-reported outcomes and PROMIS in research and clinical applications: experiences from the PCORI pilot projects. *Qual Life Res* 2016; 25(8): 2109–2116.

21. Bartlett SJ, Orbai AM, Duncan T, et al. Reliability and validity of selected PROMIS measures in people with rheumatoid arthritis. *PloS One* 2015; 10(9): e0138543.

22. Fries JF, Witter J, Rose M, et al. Item response theory, computerized adaptive testing, and PROMIS: assessment of physical function. *J Rheumatology* 2014; 41(1): 153–158.

23. Oude Voshaar MAH, ten Klooster PM, Glas CAW, et al. Relative performance of commonly used physical function questionnaires in rheumatoid arthritis and a patient-reported outcomes measurement information system computerized adaptive test. *Arthritis Rheumatol* 2014; 66(10): 2900–2908.

24. Hays RD, Spritzer KL, Fries JF, et al. Responsiveness and minimally important difference for the Patient-Reported Outcomes Measurement Information System (PROMIS) 20-item physical functioning short form in a prospective observational study of rheumatoid arthritis. *Ann Rheumatic Diseases* 2015; 74(1): 104–107.

25. Kasturi S., Szymonifka J, Burket JC, et al. Validity and reliability of patient reported outcomes measurement information system computerized adaptive tests in systemic lupus erythematosus. *J Rheumatology* 2017; 44(7): 1024–1031.

26. Campbell DT and Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 1959; 56(2): 81–105.

27. 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.

28. Gladman DD, Goldsmith CH, Urowitz MB, et al. The Systemic Lupus International Collaborating Clinics/American College of Rheumatology (SLICC/ACR) damage index for systemic lupus erythematosus international comparison. *J Rheumatology* 2000; 27(2): 373–376.

29. Touma Z, Urowitz M and Gladman D. SLEDAI-2K for a 30-day window. *Lupus* 2010; 19(1): 49–51.

30. Castelino M, Abbott J, McElhone K, et al. 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(4): 684–696.

31. Koo TK and Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropractic Medicine* 2016; 15(2): 155–163.

32. Raykov T. Evaluation of convergent and discriminant validity with multitrait-multimethod correlations. *Br J Math Stat Psychol* 2011; 64(1): 38–52.

33. Streiner DL, Norman GR and Cairney J. *Health measurement scales: a practical guide to their development and use*. USA: Oxford University Press, 2015.

34. Chahal J, Whelan DB, Jaglal SB, et al. The multiligament quality of life questionnaire. *Am J Sports Med* 2014; 42(12): 2906–2916.

35. McHorney CA, Johne W Jr and Anastasiae R. The MOS 36-item short-form health survey (SF-36). *Med Care* 1993; 31: 247–263.

36. Sweet JJ, Doninger NA, Zee PC, et al. Factors influencing cognitive function, sleep, and quality of life in individuals with systemic lupus erythematosus: a review of the literature. *Clin Neuropsychol* 2004; 18(1): 132–147.

37. Hwang MC, Ogdie A and Puravath A. Reliability and validity of patient-reported outcomes measurement information system short forms in ankylosing spondylitis. *J Rheumatology* 2019; 47(8): 1182–1188

38. Jones JT, Carle AC, Wootton J, 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(1): 133–142.

39. Kasturi S, Szymonifka J, Burket JC, 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(3): 397–404.

40. Chu P, Hinze A, Mathis N, et al. *Systemic lupus erythematosus and the evaluation of poor sleep. Arthritis Rheumatol*; 2016; 68.

41. Ware JE Jr and Sherbourne CD. The MOS 36-ltem Short-Form Health Survey (SF-36). *Med Care* 1992; 30: 473–483.

42. Yazdany J. Health-related quality of life measurement in systemic lupus erythematosus: The LupusQoL, SLEQoL, and L-QoL. *Arthritis Care Res* 2011; 63(0 11): S413.

43. Ulusoy M, Sahin NH and Erkmen H. The Beck anxiety inventory: psychometric properties. *J Cognitive Psychotherapy* 1998; 12(2): 163–172.

44. Lam R, Lamy F-X, Danchenko N, et al. Psychometric validation of the Perceived Deficits Questionnaire-Depression (PDQ-D) instrument in US and UK respondents with major depressive disorder. *Neuropsychiatr Dis Treat* 2018; 14: 2861–2877.

45. Beck AT, Steer RA, Brown GK. Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation, 1996.

46. Tennant K. Assessment of fatigue in older adults: the FACIT Fatigue Scale (version 4). *Support Care Cancer* 2015; 23(5): 1355–1364.

47. Johns M. *The Epworth Sleepiness Scale, Lyon France:* MW Johns, 1997. Available from:, https://epworthsleepinessscale.com/about-the-ess/.