AMERICAN COLLEGE
of RHEUMATOLOGY
Empowering Rheumatology Professionals

Patient-Reported Outcome Information Collected from Lupus Patients Using a Mobile Application: Compliance and Validation

Kristy Bell, [iD] Claire Dykas, Bridget Muckian, Brooke Williams, Hope Rainey, Maggy Comberg, Mary Mora, Katherine A. Owen, [iD] and Peter E. Lipsky [iD]

Objective. Patient-reported outcomes (PROs) can provide critical information concerning the impact of a disease on an individual. Mobile technology to collect PRO data in an electronic format (ePRO) allows for frequent assessment in the person's regular environment. The goal of this study was to assess the compliance with a phone application (app) and validate ePRO information in individuals with systemic lupus erythematosus (SLE).

Methods. A smartphone app that collects ePRO data from various clinical instruments was developed. Information was collected by both an ePRO and a paper-administered instrument as part of a multicenter randomized interventional clinical trial of patients meeting American College of Rheumatology (ACR) criteria for the classification of SLE. To determine agreement between PRO information collected in the different formats, intraclass correlation coefficients (ICCs), paired Student's t tests, and Bland–Altman plots were evaluated. Compliance and Cronbach's alpha were also assessed as a measure of survey reliability.

Results. For the 62 subjects from diverse ancestral backgrounds, compliance with ePRO completion was high (more than 75%). Cronbach alpha values for PROs indicated moderate to high survey reliability. The vast majority (73.4%) of ICC values were indicative of good to excellent reliability between measurement methods. Bland–Altman plots verified method agreement, and 87% of pairwise t tests yielded an insignificant difference between information collected with the different administration methods.

Conclusion. The excellent compliance and the high level of consistency between data collected by paper and that collected by electronic methods indicate that the app provides a reliable means of cataloging real-time changes in PROs in SLE patients.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease characterized by diverse manifestations and clinical heterogeneity (1). Patients with active SLE experience a range of clinical manifestations, and lupus is often complicated by flares of varying severity, followed by periods of clinical quiescence (2). Even during times of lesser inflammatory activity, lupus patients frequently experience varying levels of symptoms such as daily fluctuations in fatigue or pain (1). Consequently, individuals living with SLE face a lifetime of symptomatic burdens, including fatigue, pain, sleep disturbance, and neuropsychiatric manifestations, that impair their ability to carry out normal daily activities

and contribute to a reduction in health-related quality of life (HRQoL) (3). The detrimental impacts of SLE on HRQoL are often undervalued in physician assessments of disease activity and damage, causing frequent discordance between physicians' and SLE patients' estimations of disease burden (4). As a result, patient-reported outcomes (PROs) in general consist of structured feedback directly from patients regarding their symptoms, which can be used to supplement other more standard clinical measures such as the physician-reported SLE Disease Activity Index (SLEDAI) (5). PRO instruments capture critical information uniquely known to the patient, such as fatigue, pain, memory loss, emotional well-being, and anxiety level, and have been shown to

provide insight regarding treatment effectiveness and mortality prediction in SLE patients (6). Both disease agnostic HRQoL tools (such as the Medical Outcomes Short Form 36 [SF-36]) and SLE-specific PRO instruments (such as the LupusPRO questionnaire) (7) have been developed to evaluate the impact of disease on an individual patient.

One issue with current instruments is that PROs are often recorded intermittently, typically during an in-person clinic visit, and require patients to recall a period of several weeks or months. Consequently, important PRO information may not always be accurate or representative of the complete recall period. Administration of paper-and-pencil questionnaires require data to be collected, recorded, and computerized manually, which limits the ability to perform timely analysis and may lead to secondary data entry errors (8,9).

Clinically focused mobile health applications (apps) have been developed previously that use ePRO surveys to monitor other inflammatory diseases such as rheumatoid arthritis and report high (79%) median patient adherence (10,11). We have developed a custom-designed smartphone app for the purpose of dense PRO monitoring to facilitate the analysis of real-time trends in SLE patient self-assessment. The current report includes a relatively long-duration follow-up (6 months) of SLE patients to assess the utility of our app in the remote symptom reporting of various PRO instruments. This study sought to evaluate patient compliance with mobile app PRO completion and determine the variability and/or equivalency of measurements derived from digital measures compared with traditional paper PROs. Taken together, these analyses demonstrate that PRO data collected via the mobile app are reliable and suggest that this information can be used within the context of a clinical trial or in clinical practice as a means to catalog real-time changes in disease status and support timely therapeutic interventions.

MATERIALS AND METHODS

PRO measurement tools. Participants were instructed to use a 100-mm visual analog scale (VAS) daily with the app and at every site visit to measure patient global assessment (PtGA), fatigue (Fatigue), and pain (Pain). VASs allow continuous scaling of disease severity, directly grounded in clinical observation at the time of scoring (12).

Subjects were prompted to report the duration of morning stiffness each day by noon via smartphone and at clinic visits on paper. Although this is not a standard PRO measure, the morning stiffness ePRO was created for this trial because morning symptoms are typical of an inflammatory disease and are improved by RAYOS (Horizon Pharma) in rheumatoid arthritis (13). On weekdays and during clinic visits, patients completed the following fatigue assessments: the Fatigue Severity Scale (FSS; 9 questions) (14) and the Functional Assessment of Chronic Illness Therapy-Fatigue Scale (FACIT-F Version 4; 13 questions) (15). The SF-36 (36 questions; 8 scored domains) (16), the Patient Reported Outcome Measurement Information System (PROMIS-29 Profile Version 1.0; 29 questions; 7 scored domains) (17), and the LupusPRO (Version 1.8) survey (49 questions; 12 scored domains; 2 constructs) (7) were also completed once a week and on paper during in-clinic visits to evaluate effect of disease burden on quality of life. The Systemic Lupus Activity Questionnaire (SLAQ) (5), which is not a PRO but rather a personal assessment of SLE disease activity, was also completed weekly via the app and at clinic visits on paper. For FACIT-F, SF-36, and LupusPRO, higher scores indicate better health; for PtGA, Fatigue, Pain, Morning Stiffness, FSS, and SLAQ, higher scores indicate a negative impact on health.

eLuPRO development. The eLuPRO mobile device app was designed with input from both physician and patient focus groups. The mobile PRO app (hereafter referred to as "eLuPRO") was prepared in JavaScript for use on the android platform. The eLuPRO app was mounted on a Galaxy S7 smartphone (Samsung Electronics), which was provided to each subject for the duration of the study with entries uploaded daily to a secure database. The app data were stored in an independent database contained on a Health Insurance Portability and Accountability Act (HIPAA)-compliant secure cloud-based server. The app was created with content from the validated PRO instruments described subsequently. The individual PRO instruments were reproduced in English identically for the app except that the wording of several PRO instruments (FSS, FACIT-F, SF-36, LupusPRO Version 1.8, and SLAQ) was modified during eLuPRO development to account for the revised recall periods used for this study (ie, "the past 24 hours or the past week"). Questions using the VAS were oriented in landscape when displayed on the phone such that the scale was 100 mm in length as per the standard paper version. For every survey, one question was asked per screen, and a green check mark appeared on the eLuPRO home screen once a PRO was completed. Patients were instructed to bring the smartphone to all activities, including walks, errands, and trips. Daily reminders were set on the smartphone and via a paired Samsung Gear S2 Smartwatch to prompt PRO data entry. The reminder for the morning stiffness questionnaire was sent daily at noon; however, all other PROs were completed in the evening to capture the full day's variations. Patients were able to

customize the exact time at which they were reminded to complete evening surveys.

Study design. The eLuPRO app was evaluated as part of an exploratory study conducted within the completed phase 4 RIFLE trial (RAYOS Inhibits Fatigue in Lupus Erthymatosus; www.ClinicalTrials.gov identifier NCT03098823), which was a multicenter, randomized, double-blind, double dummy crossover study comparing the effect of delayed release prednisone (RAYOS) on fatigue in SLE with the effect of immediate-release (IR) prednisone on fatigue in SLE. This study recruited 62 SLE patients aged 18 years or older between September 12, 2017, and May 28, 2019. Participants were required to meet SLE classification criteria defined by either the American College of Rheumatology (ACR) or the Systemic Lupus International Collaborating Clinics Classification (SLICC), to have increased fatigue as assessed by a FACIT-F score of less than 25, and to be on a stable regimen of IR prednisone before screening. All patients were either English-speaking or had a caregiver who spoke English. During the 26-week trial, participants were instructed to use the custom-built mobile app eLuPRO to complete PRO surveys daily, weekly, or 5 days a week according to a provided PRO schedule. eLuPRO tracking additionally included a 14-day lead-in period to establish baseline disease activity and confirm eHealth literacy. Patients unable to use the eLuPRO app during baseline were removed. In-clinic visits occurred at two baseline visits and monthly for the duration of the study, during which patients completed both paper and eLuPRO versions of all PRO instruments separated by a distraction (participants were given lunch). Paper responses were manually entered into an electronic database, and data were securely stored in the study electronic data collection system (iMedNet), whereas PRO responses were entered into a database via the smartphone daily. The study was approved by the Institutional Review Board (IRB) at each clinical site, and patients agreed to participate by signing an IRB-approved informed consent form.

Health literacy was assessed using a validated 3-item measure developed by Chew et al [18]. For our analyses, participants

Table 1. Demographic details of enrolled study participants

Demographics and other characteristics		Count	Mean
Sex	Female (enrolled/completed)	57/41	N/A
	Male (enrolled/completed)	5/5	
Age at baseline, y	20-29	7	45.7
	30-39	13	
	40-49	20	
	50-59	13	
	60-69	8	
	70-89	1	
Ethnicity	Hispanic	11	N/A
	Non-Hispanic	51	
Race	American Indian/Alaska native	0	N/A
	Native Hawaiian/Pacific Islander	0	
	Asian	3	
	Black or African ancestry	18	
	White	37	
	Mixed race or other	4	
Education in years	12 (H.S.)	8	15.6
	13-16 (college)	35	
	17-20 (post-graduate)	21	
Health literacy	**Response**		
Question 1: "How often do you have problems learning about your medical condition because of difficulty understanding written information?"	Always	0	N/A
	Usually	0	
	Sometimes	4	
	Rarely*	8	
	Never*	50	
Question 2: "How often do you have someone like a family member, friend, hospital or clinic worker, or caregiver help you read health plan materials, such as written information about your health or care you are offered?"	Always	0	N/A
	Usually	1	
	Sometimes	4	
	Rarely*	7	
	Never*	50	
Question 3: "How confident are you filling out medical forms by yourself?"	Not at all	0	
	A little bit	1	
	Somewhat	1	
	Quite a bit*	6	
	Extremely*	54	

Abbreviations: H.S., high school; N/A, not applicable.
* Response indicates adequate health literacy.
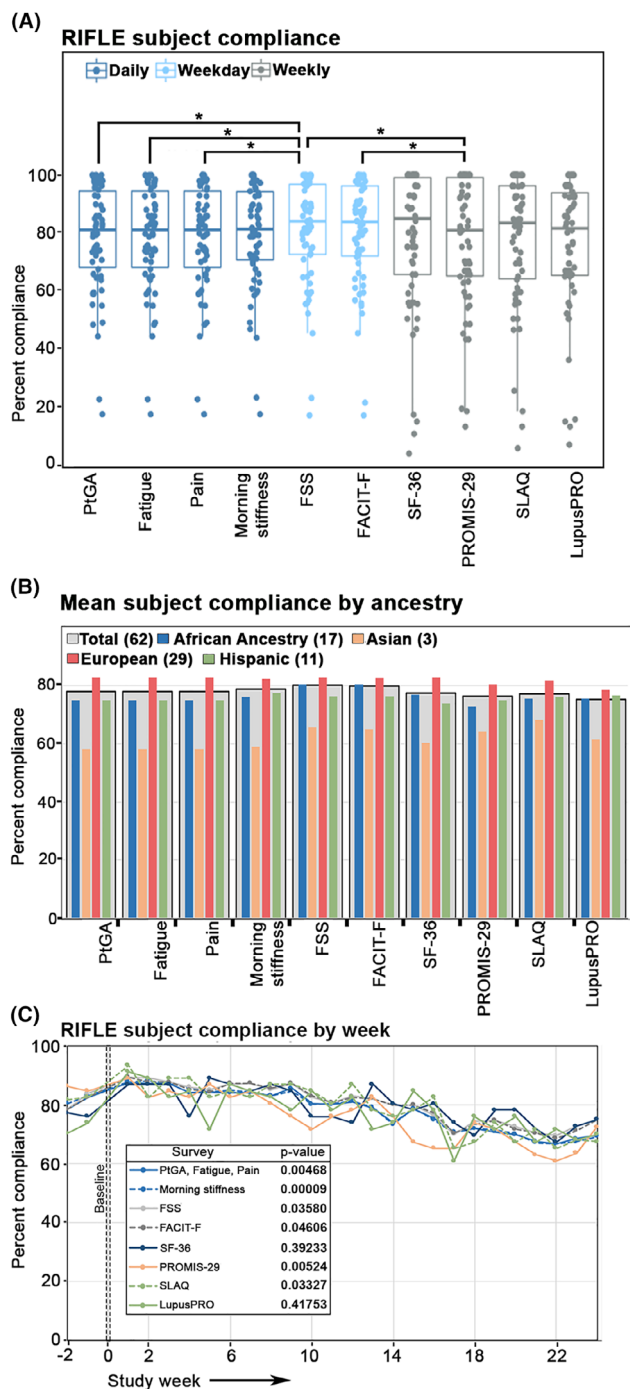
Figure 1. Patient compliance with application (app)-based electronic patient-reported outcome (ePRO) completion. **A**, Boxplots showing the compliance summary according to the survey schedule. Each point represents the compliance of one of the 62 patients. Asterisks (*) indicate a significant ($P < 0.05$) difference between surveys according to a Wilcoxon signed rank test with a Bonferroni P value adjustment. **B**, Each bar represents a group of subjects based on self-reported ancestry. Differences in the mean rank of compliance across ancestries were insignificant for each survey type (Kruskal-Wallis analysis of variance; $P > 0.05$). **C**, Mean compliance is detailed weekly for each questionnaire during the 24-week trial and 2 weeks of baseline measures. For each ePRO, P values from a Wilcoxon signed rank test compare compliance between the 2-week baseline and weeks 23-24.

who responded "sometimes," "usually," or "always" for questions 1 or 2, or "somewhat," "a little bit," or "not at all" to question 3, were classified as having limited health literacy, as described in Katz et al (19).

Statistical analysis

Compliance. Compliance (completing PROs according to the survey schedule) for all surveys was computed by expressing the number of PROs completed on the specified days as a percentage of how many should have been completed given the subjects' enrollment and completion/withdrawal date. A Friedman's analysis of variance (ANOVA) test was employed to evaluate significant differences in the mean rank of compliance across surveys. Application of the Wilcoxon signed rank test additionally evaluated pairwise significance using the Bonferroni P value adjustment. Compliance of patients who completed the trial was also assessed weekly to determine whether app use fluctuated over time.

Cronbach alpha. For multiquestion PRO tools, Cronbach alpha coefficients were computed as a measure of intersurvey reliability. Alpha coefficients for each PRO tool were calculated separately for the paper and electronic modes using all available data between baseline and trial completion. Results for paper and ePROs were tabulated and compared with previously published Cronbach coefficients for each PRO instrument in order to assess similarity. No direct statistical comparisons of coefficients were made between measurement methods; however, we examined these values to assess whether using eLuPRO distorted the internal consistency of the PROs.

eLuPRO and paper PRO comparability. To assess the equivalence between administration methods, we used data from the 8 clinic visits for which same-day electronic and paper PRO responses were recorded. Domains for each measurement tool were assessed independently, and all comparisons were performed on a by-visit basis as well as were summarized on an overall, combined-visit basis. The strength of association was tested using multiple statistical approaches, including pairwise Student's t tests, intraclass correlation coefficients (ICCs), Pearson's correlation coefficients, and Bland–Altman plots displaying agreement and bias. Pairwise Student's t tests were performed to examine whether there was a statistically significant difference between the mean score of the two administration methods. Level of agreement was evaluated statistically by ICCs, with the absolute agreement denoted at each of the 8 study visits (20). Box-plots were produced to compare the distribution of PRO scores. Scatterplots with a fitted least squares regression line were created, and Pearson's correlation coefficients were calculated to evaluate the linear relationship between paper-based and mobile-app–based PRO

**Table 2.** Cronbach alpha coefficients for ePROs and paper PROs

Survey	Target concept	Cronbach alpha		
		ePRO	Paper PRO	Literature review
FSS	Fatigue	0.96	0.96	0.95 (ref. 20)
FACIT-F	Fatigue	0.93	0.88	0.88 (ref. 21)
SF-36	Bodily Pain	0.86	0.87	0.88 (ref. 21)
	General Health	0.74	0.73	0.70 (ref. 21)
	Mental Health	0.87	0.85	0.46 (ref. 21)
	Physical Functioning	0.91	0.91	0.94 (ref. 21)
	Role Emotion	0.94	0.90	0.86 (ref. 21)
	Role Physical	0.91	0.88	0.92 (ref. 21)
	Social Functioning	0.89	0.84	0.27 (ref. 21)
	Vitality	0.79	0.77	0.70 (ref. 21)
PROMIS-29	Anxiety	0.93	0.93	0.92 (ref. 22)
	Depression	0.91	0.92	0.94 (ref. 22)
	Fatigue	0.91	0.92	0.95 (ref. 22)
	Pain Interference	0.95	0.96	0.97(ref. 22)
	Physical Function	0.91	0.90	0.92 (ref. 22)
	Social Satisfaction	0.95	0.95	0.97 (ref. 22)
SLAQ	SLE Disease Activity	0.88	0.88	0.87 (ref. 23)
LupusPRO	HRQOL	0.95	0.95	0.96 (ref. 21)
	Non-HRQOL	0.73	0.73	0.81 (ref. 21)

Abbreviations: ePRO, electronic patient-reported outcome; FACIT-F, Functional Assessment of Chronic Illness Therapy-Fatigue Scale; FSS, Fatigue Severity Scale; HRQOL, health-related quality of life; PRO, patient-reported outcome; PROMIS-29, Patient Reported Outcome Measurement Information System; SF-36, Medical Outcomes Short Form 36; SLAQ, Systemic Lupus Activity Questionnaire.
Breakdown of the internal consistency of all multiquestion PRO tools quantified by Cronbach alpha coefficients. Cronbach alpha coefficients range from 0 to 1, with higher values indicating greater reliability in measuring the targeted concept (column 2, "Target concept") of every questionnaire. Alpha coefficients from this study were comparable to those of previously reported coefficients for each survey from a comprehensive literature review.
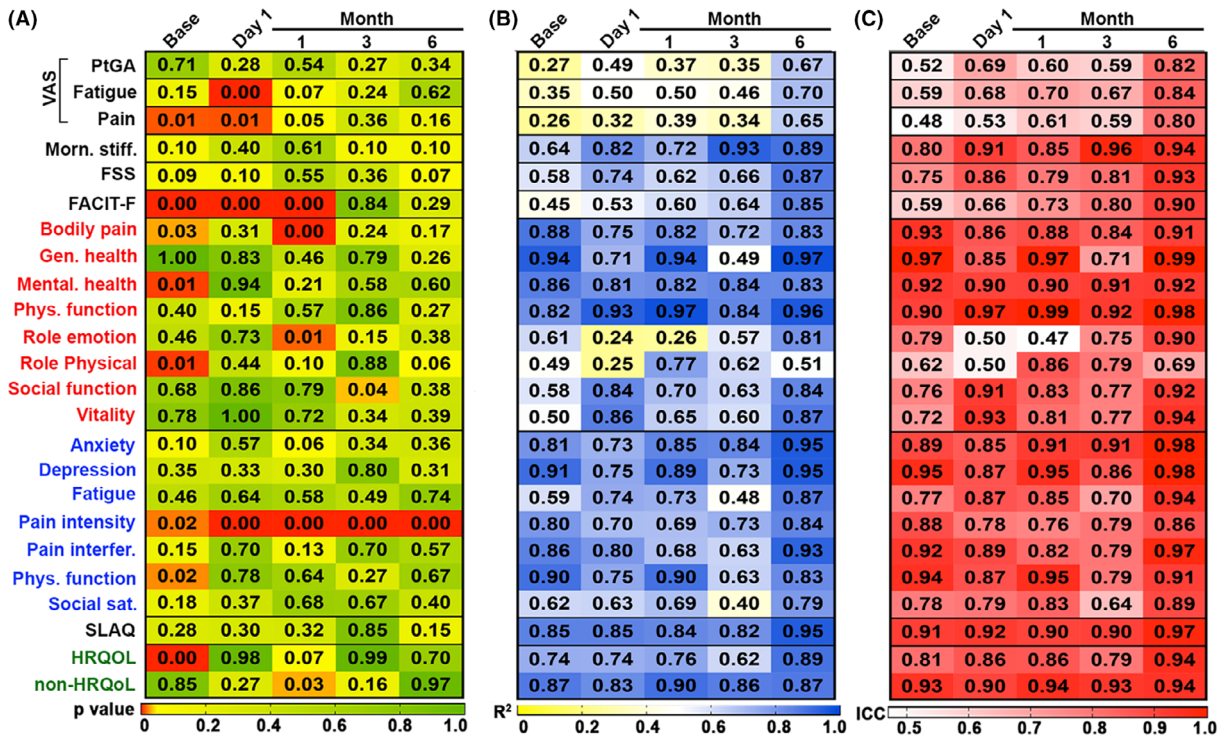

Figure 2. Assessment of electronic patient-reported outcome (ePRO) versus paper PRO administration methods. **A**, Heatmap displays paired Student's t test computations for the indicated timepoints. Computations were made between either domain scores (red text for Medical Outcomes Short Form 36 [SF-36], blue text for Patient Reported Outcome Measurement Information System [PROMIS-29]), construct scores (green text for LupusPRO), or global scores (black text). Insignificant ($P > 0.05$) and significant ($P < 0.05$) differences between electronic and paper scores reported on the same day are indicated by color. **B**, Coefficients of determination (R^2) are reported for each survey at each site visit indicated. High R^2 values (blue) indicate a strong linear relationship between administration methods, whereas low R^2 values (yellow) suggest more scatter. **C**, Intraclass correlation coefficients (ICCs) were computed to assess reliability between measurement methods. All ICCs were statistically significant ($P < 0.001$). VAS, visual analog scale.
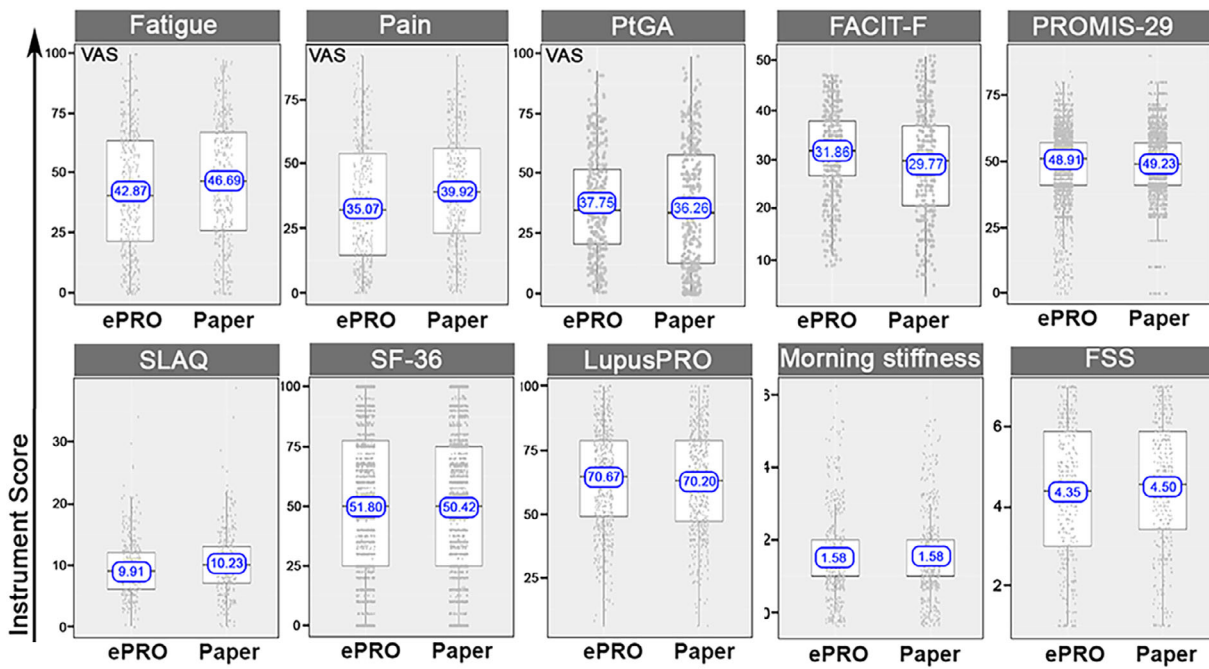
Figure 3. Distribution of electronic patient-reported outcome (ePRO) and paper PRO results for each survey. Boxplots showing the distribution and mean scores (blue text) for the indicated instrument. For multidomain PROs (Medical Outcomes Short Form 36 [SF-36], Patient Reported Outcome Measurement Information System [PROMIS-29], LuPRO), each point indicates an individual domain score for a particular patient.

scores. Agreement and bias between collection methods were shown graphically by Bland–Altman plots, which plot the score difference (electronic minus paper) against the mean paper and electronic score for each individual (21). The Bland–Altman plots include horizontal reference lines for the mean of the difference in the modes, the mean plus and minus twice the standard deviation for 95% limits of agreement, and a zero-reference line (21).
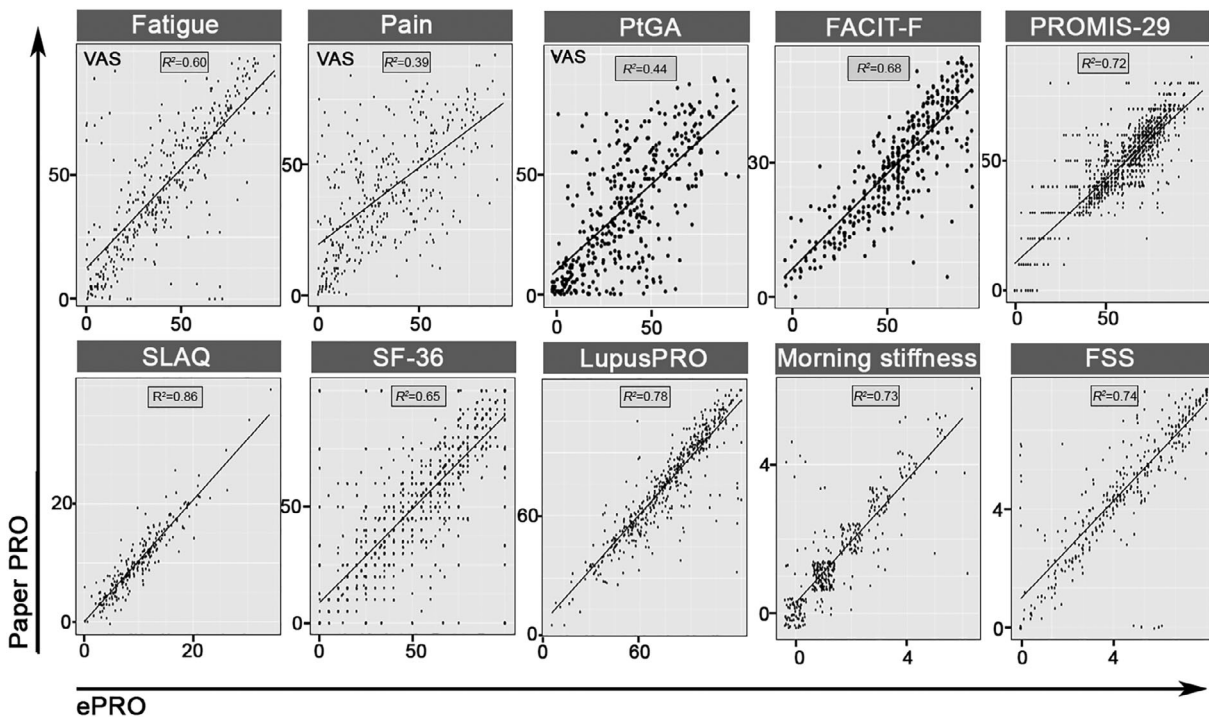
Figure 4. Correlation between collection methods. Scatter plots showing the correlation between scores recorded by both collection methods. Pearson coefficients (R^2) are shown in each plot. For multidomain patient-reported outcomes (PROs) (Medical Outcomes Short Form 36 [SF-36], Patient Reported Outcome Measurement Information System [PROMIS-29], LuPRO), each point indicates an individual domain score for a particular patient.
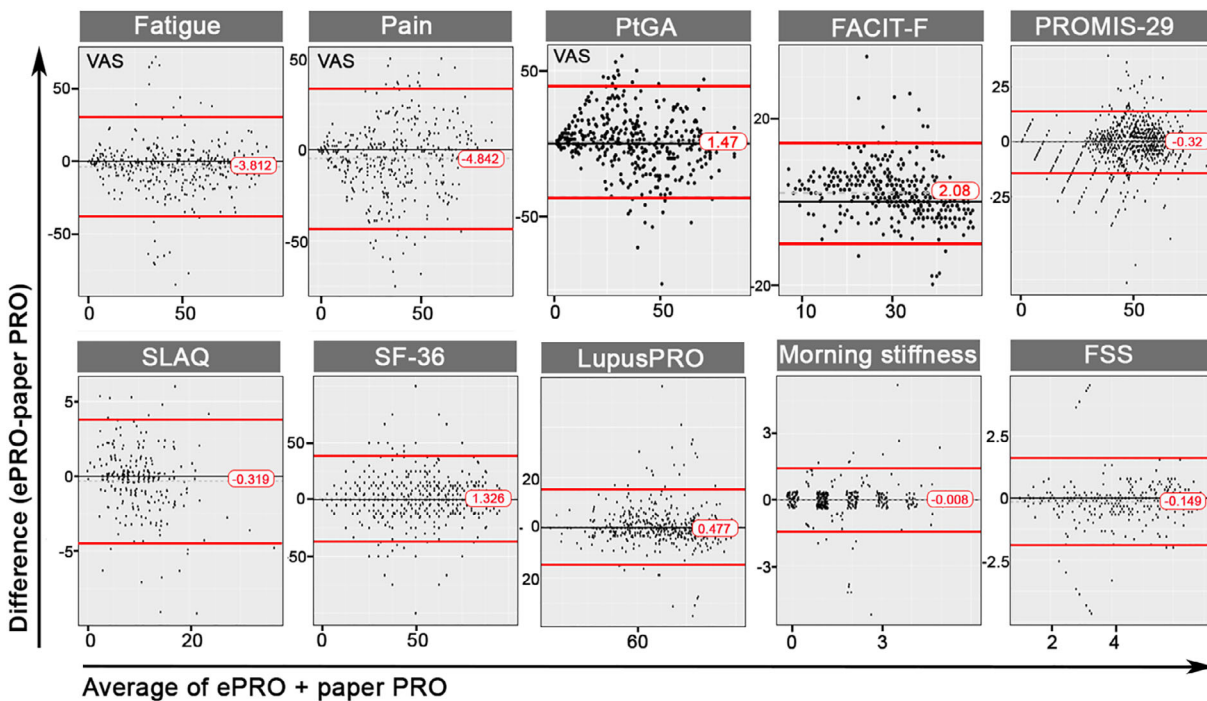
Figure 5. Bland–Altman plots to assess agreement between collection methods. The difference between electronic patient-reported outcome [ePRO] score and the paper PRO score (ePRO – paper PRO) and the average of the patient global assessment (PtGA) scores (ePRO score + paper PRO score divided by 2) are represented on the y-axis and the x-axis, respectively. The red lines represent the 95% confidence interval; the mean difference is in red text. For multidomain PROs (Medical Outcomes Short Form 36 [SF-36], Patient Reported Outcome Measurement Information System [PROMIS-29], LuPRO), each point indicates an individual domain score for a particular patient.

RESULTS

Data collection. PRO data were collected from the 62 SLE patients enrolled in the study from 21 sites across the United States (www.ClinicalTrials.gov identifier NCT03098823). A total of 46 subjects completed the entire 6-month study. Of the 16 subjects withdrawn, 11 did so within the first 3 months of the trial. Over the duration of the study, 58,173 PROs were collected through the eLuPRO app, along with 4,374 paper surveys from all clinic visits. This included 263 instances in which paper and eLuPRO versions were completed at the clinic site separated by a distraction (usually lunch), according to protocol.

Patient demographics. Enrolled subjects included 57 females (91.9%) and 5 males (8.1%) from diverse self-reported ancestral backgrounds, as detailed in Table 1. The subject population had a mean age of 45.7 years and had attained an average of 15.6 years of education where 16 years represents an individual who completed a 4-year college degree. Based on the responses to three validated health literacy questions (18), patients were adequately health literate, with 93.5% (58/62) of patients "rarely" or "never" experiencing problems learning about their condition because of difficulty understanding the information, 91.9% (57/62) of patients "rarely" or "never" receiving assistance reading health plan materials, and 97% (60/62) of patients being "extremely" or "quite a bit" confident in filling out medical forms without further assistance.

Overall and longitudinal patterns of patient compliance. Aggregate patient compliance was determined as the extent to which the ePRO requirements were fulfilled (ie, surveys were completed on time via eLuPRO). Mean compliance for mobile-app–based PRO completion was high for all surveys (more than 75.4%), with 75% of patients being at least 64.0% compliant with each measurement tool (Figure 1A). Mean rank of compliance across all study instruments was found to be statistically significant (Friedman's ANOVA; $P = 0.0071$) with significant differences also determined between individual instruments, such as FACIT-F and PROMIS-29 (Wilcoxon signed rank test, $P < 0.05$). Compliance varied slightly across ancestries; however, the difference between the mean ranks across ancestries was not significant ($P > 0.05$) (Figure 1B). The weekday surveys yielded the highest mean compliance when they included the FSS (80.2%) and FACIT-F (80.1%). Mean subject compliance for all ePRO surveys peaked at week 1 (89.4%), declining to 71.7% by week 24 (Figure 1C). Notably, the decline was significant for all but the two longest questionnaires (SF-36, LupusPRO). Nevertheless, mean compliance by week for all surveys remained high through trial progression (more than 60%), verifying the utility of mobile-app–based PRO reporting.
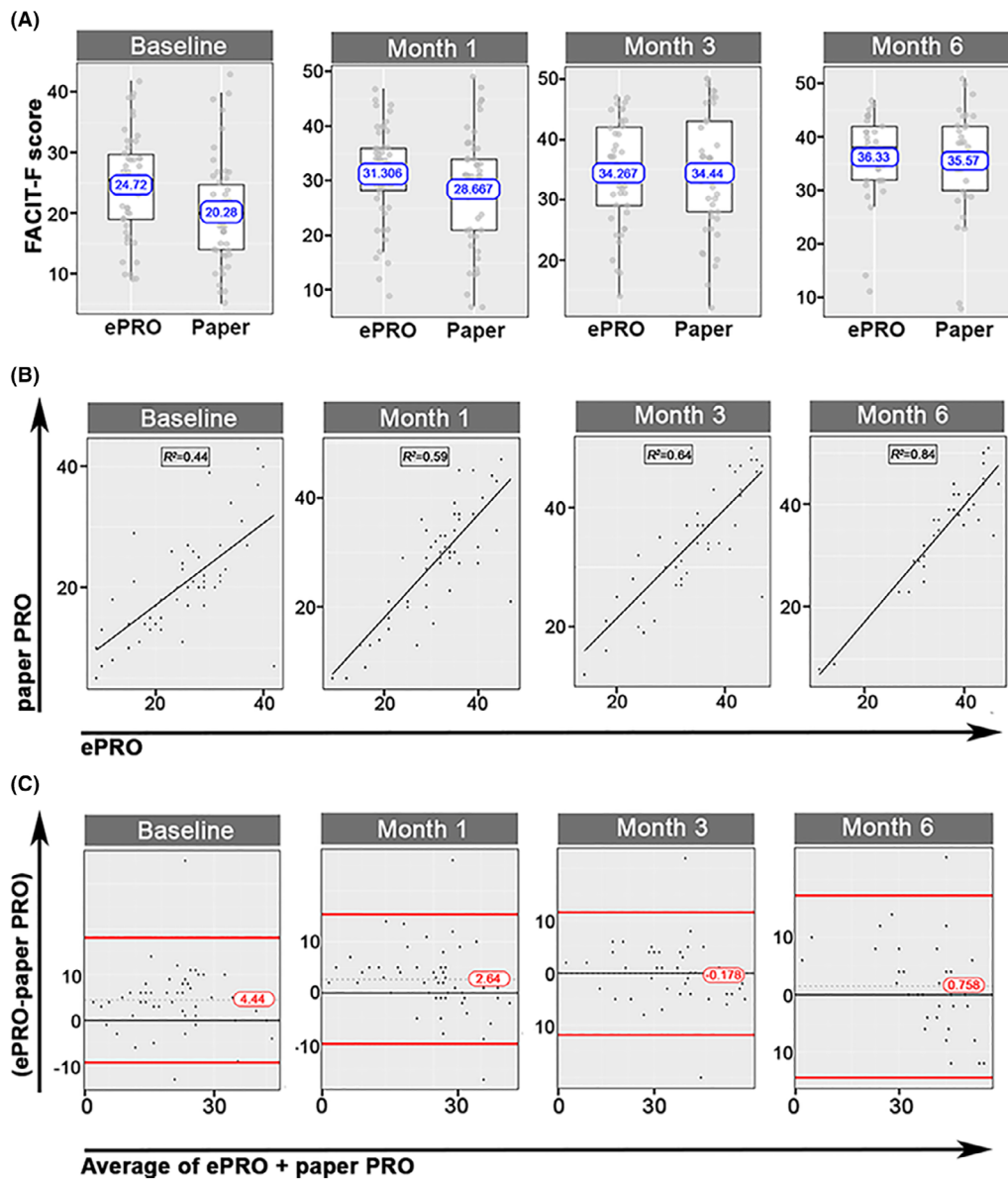
Figure 6. Agreement between collection methods for the Functional Assessment of Chronic Illness Therapy-Fatigue Scale (FACIT-F) survey at selected timepoints. **A**, Boxplots showing the distribution and mean scores (blue text) for each administration method at baseline and at months 1, 3, and 6. **B**, Scatter plots showing the correlation between scores recorded by both collection methods. Pearson coefficients (R^2) are shown in each plot. **C**, Bland-Atman plots were used to assess the agreement between each collection method. The difference between the electronic patient-reported outcome (ePRO) score and the paper PRO score (ePRO – paper PRO score) and the average (ePRO score + paper PRO score divided by 2) of the FACIT-F scores are represented on the y-axis and the x-axis, respectively. The red lines represent the 95% confidence interval; the mean difference is in red text.

Determination of internal consistency. To evaluate the robustness of instrument consistency and survey reliability, Cronbach alpha coefficients were calculated for each multiquestion PRO. Cronbach alpha coefficient computations used all available PRO data, including the 58,173 ePROs and the 4,374 paper PROs. PRO coefficients collected via eLuPRO and the corresponding paper versions ranged from 0.73 to 0.96, suggesting that both measurement methods yielded moderate to high inter-survey reliability in measuring targeted concepts (Table 2). In

addition, mobile-app–based PRO and paper PRO alpha coefficients were comparable and, in a few cases, greater than outcomes previously reported in the literature (Table 2) (22–25). This was particularly noted for the SF-36, for which alpha coefficients for the mental health and social functioning domains were greater than 0.84, whereas the literature reported outcomes were 0.27 and 0.46 for social functioning and mental health, respectively (23). Furthermore, alpha coefficients between the electronic and paper administration methods were highly similar (absolute

differences of less than 0.06), indicating that the within-survey question consistency is not lost with the use of the eLuPRO app.

Phone-app–based patient assessment is comparable to paper administration methods. We next sought to determine whether patient assessment data collected using the eLuPRO app was similar to data collected using traditional paper methods. Pairwise Student's t tests calculated at the monthly clinic visits for every PRO survey suggested insignificant differences ($P > 0.05$) between the app and paper-derived data in 167 of the 192 comparisons, representing 87% of all computations (Figure 2A; five representative time points are shown). Coefficients of determination (R^2) were computed at each clinic visit for every survey in order to measure the strength of the linear relationship between pairwise mobile-app–based PRO and paper PRO scores reported on the same day. Correlation coefficients ranged from 0.24 to 0.97 (Figure 2B), and 86.5% of the 192 coefficients computed indicated a strong relationship between modes ($r > 0.70$; $R^2 > 0.49$).

To further assess agreement between survey collection methods, 192 ICCs were calculated for each survey at each clinic visit (Figure 2C). Of the ICCs computed, 47 were indicative of moderate (0.5-0.75), 77 of good (0.75-0.9), and 64 of excellent (more than 0.90) reliability between measurement methods. All ICCs computed were significant ($P < 0.001$) and ranged from 0.47 to 0.99 with a median ICC of 0.85. Each survey exhibited a different level of variability in ICC values amid each of the 8 site visits, of which 5 are shown. The heatmap in Figure 2C reveals that Likert-scale PRO surveys appear more reliable between electronic and paper administration than the single-question VAS surveys (PtGA, Fatigue, Pain). The SLAQ survey emerged as the most reliable between methods, yielding ICC values above 0.9 at each site visit.

Boxplots confirm the similarity between the distribution of scores and mean responses for each administration method (Figure 3). Additionally, scatterplots were created for each survey to visualize the correlation between PRO collection methods (Figure 4). The SLAQ patient estimate of disease activity and LupusPRO survey displayed the strongest overall combined-visit correlation of all measurement tools with Pearson's coefficients of $r = 0.93$ and $r = 0.89$, respectively (Figure 4).

Lastly, Bland–Altman plots assessed agreement between instrument implementation (electric vs. paper) by combining all pairwise data points (Figure 5). Average differences (biases) between measurement methods as well as confidence interval widths varied at each visit, with some visits having minor positive or negative biases for each PRO. Bland–Altman plots combining all time points revealed slight positive bias between electronic and paper methods in four PROs (PtGA, FACIT-F, SF-36, and LupusPRO; electronic scores were higher) and slight negative bias in six PROs (Fatigue, Pain, Morning Stiffness, FSS, PROMIS-29, and SLAQ; paper scores were higher). Nevertheless, the zero line

was always contained within the limits of agreement in all the by-visit and combined-visit Bland–Altman plots created; therefore, there is no evidence to suggest a significant, systematic difference between administration methods. No biases surpassed the minimum clinically important difference for each PRO survey, supporting a high level of agreement between electronic and paper-reported scores. Figure 6 provides the by-visit boxplots, scatterplots, and Bland–Altman plots generated to compare the ePRO and paper PRO scores for the FACIT-F survey.

DISCUSSION

SLE is a clinically heterogenous autoimmune disease with a wide array of symptoms that negatively impact an individual's quality of life. The electronic capture of clinical trial source data, including PRO endpoints, is increasingly used to assess the impact of medical treatment or intervention. In general, PROs assess a range of outcomes, including symptoms, functional health and well-being, and psychological issues, to provide a holistic view of daily disease burden from the patient's perspective (3). PRO questionnaires have been used extensively in clinical trials to supplement clinical measures and provide clinicians with more information that may aid in decision-making regarding treatments. For example, changes in PRO outcomes from the RIFLE trial (www.ClinicalTrials.gov identifier NCT03098823) showed that treating rheumatoid arthritis patients with upadacitinib can lead to clinically significant relief from symptoms (26).

Mobile-app–based PRO data collection in clinical trials offers many advantages over traditional paper-based methods: they are not location dependent, they can be conducted in an unsupervised manner, and, most importantly, they allow for accurate and real-time reporting of symptoms. Many other SLE-specific patient-centered apps, such as LupusTracker PRO (ToTheHand, LLC) and My Lupus Log (GlaxoSmithKline), have been developed in order to empower patients in the daily management of their disease and/or in order to reduce the communication gap between SLE patients and their providers (27). In addition, ePRO apps have been developed for other inflammatory diseases, including rheumatoid arthritis. Whereas these and other studies demonstrate app compliance, few, if any, provide validation analyses (ie, the app successfully measures the domain of interest). Here, the eLuPRO phone-based app was developed for a phase 4 clinical trial in order to examine real-time changes in multiple different PRO instruments during a period of therapeutic intervention. In addition to evaluation by PROs, the inclusion of SLAQ for the personal assessment of disease activity within the eLuPRO framework provides an additional tool for patients to judge the benefit of care received. It should be noted that RIFLE was biased toward subjects experiencing increased fatigue (FACIT-F score of more than 25), which may decrease its application to the greater lupus population. Nonetheless, our results indicate that eLuPRO was both functional and widely used by patients throughout the trial,

with several subjects continuing to use the eLuPRO tools beyond their enrollment in the trial. Our double baseline approach was useful in that it provided a period of app training and it allowed us to collect multiple data points before initiation of the intervention.

Patient demographics reveal a diverse range of ancestral backgrounds, with over half of enrolled subjects of non-European descent. This is important given that certain ancestral groups experience the disease more severely, such as those of African ancestry, who account for 43% of all SLE subjects yet typically represent a low proportion of trial participants (less than 14%) (28–30). Overall compliance with app usage was high (more than 75%) for most surveys across demographics, particularly the weekly FSS and FACIT-F surveys with 80% mean compliance, demonstrating the utility of electronic patient-directed data collection.

In order to validate the extensive PRO information collected via eLuPRO, we sought to verify the equivalence of paper and electronic administration methods for all surveys. There was remarkable comparability with a significant difference in only 13% of comparisons using a Student's t test to examine differences in mean scores between methods. Notably, the pain intensity domain of the PROMIS-29 instrument was the only survey that resulted in a consistently significant difference in method score at the in-clinic visits yet ICCs and R^2 values indicated excellent agreement. One reason for the discrepancy in results is that the t test does not consider patient bias (differences at the level of individual patient), but rather it compares the mean score for each administration method at each time point. Additionally, multiquestion surveys in which one score is reported showed greater correlation between paper and ePRO responses than that of single-question instruments. ICC analysis further revealed that reliability between administration methods was acceptable, and oftentimes high, for all PROs at every in-clinic visit. Administration agreement appears slightly lower in the VAS questions than in the Likert-scaled surveys. Compared with instruments using a Likert scale to obtain ordinal-level measurements, instruments using the 100-mm VAS scale allow for the collection of measurements with more variability. Although this produces more fine-grained responses based on a line continuum, data obtained using VAS are generally more variable because of the "unstructured" nature of the scale; it is therefore not surprising that the VAS questions performed less well. Nonetheless, the ICC values for the VAS PROs show an increasing trend over the course of the trial, suggesting that agreeability may improve as more surveys are taken. All Bland–Altman plots showed points that were roughly scattered evenly around the zero line, suggesting no consistent bias between paper-based and mobile-app–based PRO scores.

Limitations of the study include its relatively small size, made up for by the large number of PROs collected. In addition, the study inadvertently collected data from patients with high medical literacy in a structured academic setting; it is, therefore, uncertain whether the app will work comparably in general practice with patients of varying health literacy profiles (31). Although construct validity had not been demonstrated at the time this study was carried out, the acceptability of apps can now be assessed with rating scales, such as the Mobile Application Rating Scale (MARS) (32), that could be useful to evaluate the app more fully. However, patient feedback indicated frustration with the redundant nature of the selected PRO instruments, manifesting as "app fatigue" and likely contributing to declining compliance over time. Although there was some manifestation of app fatigue in this trial, in the future this might be mitigated by rewards, simplification, or providing patient access to their personal data. Despite these caveats, this study represents the first successful attempt to validate a wide range of PRO information from lupus patients with a mobile app. We found that collecting data via phone app is both feasible and valid and is likely to detect changes related to treatment and/or spontaneous fluctuations in disease. The collection of dense PRO data permits analysis of real trends rather than intermittent pools of information, allows for assessment in a patient's regular environment, and is resistant to obstructions in data quality and blank entries. Importantly, the use of the eLuPRO app permits real-time decision-making because data collection and entry into the database are automatically reported daily. The data suggest that PRO collection by app could replace that done in the clinic by paper or electronic methodology. Future analyses will expand on these current observations and will focus on identifying those health domains that best correlate with clinical changes in disease activity, reducing both redundancy and response burden.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

Ms. Bell and Drs. Owen and Lipsky drafted the manuscript, and all authors revised it critically for important intellectual content. All authors approved the final version to be published. Dr. Lipsky had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study conception and design. Bell, Dykas, Muckian, Williams, Rainey, Comberg, Mora, Owen, Lipsky.
Acquisition of data. Dykas, Muckian, Williams, Rainey, Comberg, Mora, Lipsky.
Analysis and interpretation of data. Bell, Owen, Lipsky.

REFERENCES

1. Kiriakidou M, Ching C. Systemic lupus erythematosus. Ann Intern Med 2020;172:ITC81–96.

2. Gensous N, Marti A, Barnetche T, Blanco P, Lazaro E, Seneschal J, et al. Predictive biological markers of systemic lupus erythematosus flares: a systematic literature review. Arthritis Res Ther 2017;19:238.

3. Schmeding A, Schneider M. Fatigue, health-related quality of life and other patient-reported outcomes in systemic lupus erythematosus. Best Pract Res Clin Rheumatol 2013;27:363–75.

4. Yen JC, Neville C, Fortin PR. Discordance between patients and their physicians in the assessment of lupus disease activity: relevance for clinical trials. 1999;8:660–70.

5. Romero-Diaz J, Isenberg D, Ramsey-Goldman R. Measures of adult systemic lupus erythematosus: updated version of British Isles Lupus Assessment Group (BILAG 2004), European Consensus Lupus Activity Measurements (ECLAM), Systemic Lupus Activity Measure, Revised (SLAM-R), Systemic Lupus Activity Questionnaire for Population Studies (SLAQ), Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K), and Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index (SDI). Arthritis Care Res (Hoboken) 2011;63.

6. Azizoddin DR, Jolly M, Arora S, Yelin E, Katz P. Patient-reported outcomes predict mortality in lupus. Arthritis Care Res (Hoboken) 2019; 71:1028–35.

7. 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.

8. Rolfson O, Salomonsson R, Dahlberg LE, Garellick G. Internet-based follow-up questionnaire for measuring patient-reported outcome after total hip replacement surgery-reliability and response rate. Value in Health 2011;14:316–21.

9. Muehlhausen W, Doll H, Quadri N, Fordham B, O'Donohoe P, Dogar N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. Health Qual Life Outcomes 2015;13:167.

10. Bingham CO 3rd, Gaich CL, DeLozier AM, Engstrom KD, Naegeli AN, de Bono S, et al. Use of daily electronic patient-reported outcome (PRO) diaries in randomized controlled trials for rheumatoid arthritis: rationale and implementation. Trials 2019;22;20:182 [erratum in: Trials 2019;20:322].

11. Richter JG, Nannen C, Chehab G, Acar H, Becker A, Willers R, et al. Mobile app-based documentation of patient-reported outcomes— 3-months results from a proof-of-concept study on modern rheumatology patient management. Arthritis Res Ther 2021;19;23:121.

12. Thanou A, James J, Arriens C, Aberle T, Chakravarty E, Rawdon J, et al. Scoring systemic lupus erythematosus (SLE) disease activity with simple, rapid outcome measures. Lupus Sci Med 2019;6: e000365.

13. Alten R, Holt R, Grahn A, Rice P, Kent J, Buttgereit F, et al. Morning stiffness response with delayed-release prednisone after ineffective course of immediate-release prednisone. Scand J Rheumatol 2015; 44:354–58.

14. Ad Hoc Committee on Systemic Lupus Erythematosus Response Criteria for Fatigue. Measurement of fatigue in systemic lupus erythematosus: a systemic review. Arthritis Rheum 2007;57:1348–57.

15. Kosinski M, Gajria K, Fernandes AW, Cella D. Qualitative validation of the FACIT-fatigue scale in systemic lupus erythematosus. Lupus 2013;22:422–30.

16. Stoll T, Gordon C, Seifert B, Richardson K, Malik J, Bacon PA, et al. Consistency and validity of patient administered assessment of quality of life by the MOS SF-36; its association with disease activity and damage in patients with systemic lupus erythematosus. J Rheumatol 1997;24:1608–14.

17. Katz P, Pedro S, Michaud K. Performance of the patient-reported outcomes measurement information system 29-item profile in rheumatoid arthritis, osteoarthritis, fibromyalgia, and systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2017;69:1312–21.

18. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004;36:588–94.

19. Katz P, Dall'Era M, Trupin L, Rush S, Murphy LB, Lanata C, et al. Impact of limited health literacy on patient-reported outcomes in systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2021;73: 110–19.

20. Mcgraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. American Psychological Association 1996;1: 30–46.

21. Giavarina D. Understanding Bland Altman analysis. Biochem Med (Zagreb) 2015;25:141–51.

22. Rosti-Otajärvi E, Hämäläinen P, Wiksten A, Hakkarainen T, Ruutiainen J. Validity and reliability of the Fatigue Severity Scale in Finnish multiple sclerosis patients. Brain Behav 2017;7:e00743.

23. Azizoddin DR, Weinberg S, Gandhi N, Arora S, Block JA, Sequeira W, et al. Validation of the LupusPRO version 1.8: an update to a disease-specific patient-reported outcome tool for systemic lupus erythematosus. Lupus 2018;27:728–37.

24. Schnall R, Liu J, Cho H, Hirshfield S, Siegel K, Olender S. A health-related quality-of-life measure for use in patients with HIV: a validation study. AIDS Patient Care STDs 2017;31:43–8.

25. Yazdany J, Yelin EH, Panopalis P, Trupin L, Julian L, Katz PP. Validation of the Systemic Lupus Erythematosus Activity Questionnaire in a large observational cohort. Arthritis Rheum 2008;59:136–43.

26. Strand V, Schiff M, Tundia N, Friedman A, Meerwein S, Pangan A, et al. Effects of upadacitinib on patient-reported outcomes: results from SELECT-BEYOND, a phase 3 randomized trial in patients with rheumatoid arthritis and inadequate responses to biologic disease-modifying antirheumatic drugs. Arthritis Res Ther 2019;21:263.

27. Dantas LO, Weber S, Osani MC, Bannuru RR, McAlindon TE, Kasturi S. Mobile health technologies for the management of systemic lupus erythematosus: a systematic review. Lupus 2020;29: 144–56.

28. Falasinnu T, Chaichian Y, Bass MB, Simard JF. The Representation of gender and race/ethnic groups in randomized clinical trials of individuals with systemic lupus erythematosus. Curr Rheumatol Rep 2018; 20:1–11.

29. Williams EM, Bruner L, Adkins A, Vrana C, Logan A, Kamen D, et al. I too, am America: a review of research on systemic lupus erythematosus in African-Americans. Lupus Sci Med 2016;3:e000144.

30. Anjorin A, Lipsky P. Engaging African ancestry participants in SLE clinical trials. Lupus Sci 2018;5.

31. Bakker MM, Putrik P, Rademakers J, Laar M van de, Vonkeman H, Kok MR, et al. Addressing health literacy needs in rheumatology— which patient health literacy profiles need the attention of health professionals? Arthritis Care Res (Hoboken) 2020;73:100–9.

32. Terhorst Y, Philippi P, Sander L, Schultchen D, Paganini S, Bardus M, et al. Validation of the mobile application rating scale. Plos One 2020; 15:e0241480.