Original article

The FATIGUE-PRO: a new patient-reported outcome instrument to quantify fatigue in patients affected by systemic lupus erythematosus

Thomas Morel [1], Stefan Cano[2], Susan J. Bartlett [3], Caroline Gordon[4], Birgit Haier[5], Antoine Regnault[2], Matthias Schneider[6], Christian Stach[5] and Sophie Cleanthous[2]

Abstract

Objectives. This study aimed to implement a patient-centred and evidence-based approach to develop a novel patient-reported outcome (PRO) instrument to measure fatigue in patients with SLE.

Methods. A three-step mixed methods psychometric (MMP) approach was followed. Steps comprised first draft item generation and review using interview data; evaluation and refinement of second draft items using mixed methods data, including interview and quantitative data from a phase 2 clinical study in SLE analysed using Rasch Measurement Theory (RMT) analysis; and evaluation of the final FATIGUE-PRO items using RMT and complementary Classical Test Theory (CTT) analyses. Guided by MMP criteria, a team of clinicians and outcome-measurement experts assessed evidence to inform instrument development.

Results. Step 1 culminated in 55 items ($n = 39$ patients interviewed). Their refinement in step 2 using mixed methods evidence led to the final FATIGUE-PRO instrument comprising 31 items across three scales of fatigue: physical fatigue (9 items), mental and cognitive fatigue (11 items) and susceptibility to fatigue (11 items). Qualitative ($n = 43$ patients) and quantitative ($n = 106$ patients) evidence strongly supported the scales' content comprehensiveness and targeting, item quality and fit, conceptual uniqueness and appropriateness of the response scale. The FATIGUE-PRO further benefited from excellent reliability (RMT: 0.92–0.94 and CTT: 0.95–0.96) and supportive evidence of construct validity from assessments against other PROs.

Conclusion. The conceptual advances, comprehensive coverage and strong psychometric properties of the FATIGUE-PRO will significantly advance the measurement and management of fatigue in SLE, both in clinical trials and routine practice.

Trial registration. ClinicalTrials.gov (https://clinicaltrials.gov), NCT02804763

Key words: lupus, fatigue, PRO, patient, mixed methods psychometric evaluation, validation

<div style="color: green; writing-mode: vertical">CLINICAL SCIENCE</div>

> **Rheumatology key messages**
>
> - The FATIGUE-PRO is a novel patient-reported outcome instrument to measure fatigue in systemic lupus erythematosus.
> - The instrument comprises three scales: physical fatigue, mental and cognitive fatigue, and susceptibility to fatigue.
> - FATIGUE-PRO is comprehensive, valid, and reliable and should improve the recording and management of fatigue.

---

[1]UCB Pharma, Brussels, Belgium, [2]Modus Outcomes, Letchworth Garden City, UK, [3]Department of Medicine, McGill University, Montreal, Quebec, Canada, [4]Rheumatology Research Group, Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK, [5]UCB Pharma, Monheim am Rhein and [6]Poliklinik und Funktionsbereich für Rheumatologie, Heinrich-Heine-Universität, Düsseldorf, Germany

Correspondence to: Thomas Morel, UCB Pharma, Allée de la Recherche 60, 1070 Brussels, Belgium. E-mail: thomas.morel@ucb.com

Introduction

SLE is an autoimmune disease characterized by inflammation and damage in multiple organs and a variable disease course including flares, periods of remission and/or more persistent active disease [1]. Fatigue is one of the most common and burdensome symptoms experienced by SLE patients [2–7]. The unpredictable nature of the disease and the severity of fatigue can heavily disrupt patients' daily lives [4], with patients reporting an emotional burden as well as an influence on cognition, work, leisure and social and family activities [8].

Numerous generic and SLE-specific patient-reported outcome (PRO) instruments are available for measuring fatigue [9–14]. Although some evidence supports their use in SLE [e.g. the relevance of Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) content in SLE] [15], potentially significant limitations constrain these instruments' measurement potential, such as issues with content validity [3, 16, 17]. Research has demonstrated that patients' experiences of fatigue are not comprehensively represented in PRO instruments used in SLE and other conditions [18, 19]. For example, FACIT-F [20], the Lupus Quality of Life (LupusQoL) Fatigue domain [10], and the 36-item Short Form (SF-36) [14] vitality scale do not adequately cover the breadth of fatigue experienced by SLE patients [3, 21]. Additionally, test design issues and issues with conceptual clarity and fit have been identified with fatigue-specific instruments such as the FACIT-F [3, 20]. These issues may stem from the lack of a clear and comprehensive conceptualization of the symptom experience in SLE to support the instruments—a key component in rigorous PRO development [22–24].

Until recently, a limited understanding of fatigue in SLE has hindered further instrument development, which led the US Food and Drug Administration (FDA) to acknowledge fatigue as a measurement challenge in clinical research [25]. To address this challenge, a recent qualitative study investigated patients' direct and explicit descriptions of fatigue in SLE and conceptualized the symptom experience in three overarching domains: physical fatigue, including concepts of general energy, stamina and physical manifestations of fatigue on the body and limbs; mental and cognitive fatigue, covering concepts such as mental energy, motivation, 'brain fog' and manifestations of fatigue on cognitive functioning; and susceptibility to fatigue, comprising concepts linked to the unpredictable course and onset of fatigue, as well as non-restorative sleep and rest [21]. This model has extended existing knowledge through its more comprehensive description of the fatigue experience in SLE and through establishment of the new concept of susceptibility to fatigue.

Using this conceptual model as a foundation and by implementing a patient-centred and evidence-based approach, the objective of this study was to develop a new PRO instrument to measure fatigue in SLE that addresses the conceptual limitations of existing PRO instruments. Here we describe the formulation, refinement and first evaluations of this new instrument, the FATIGUE-PRO.

Methods

Overview

Instrument development followed a three-step mixed methods psychometric (MMP) approach, enabling integrated interpretation of quantitative and qualitative findings to inform decisions on item selection and refinement (Fig. 1) [26, 27]. Instrument design reflected the three previously established conceptual domains of fatigue, from which corresponding scales were generated: physical fatigue (PF), mental and cognitive fatigue (MF) and susceptibility to fatigue (SF) [21].

The development process comprised the following steps: (1) first draft item generation and review, (2) evaluation and refinement of the second draft item sets, and (3) evaluation of the final FATIGUE-PRO instrument (Fig. 1). In step 1, qualitative interview data informed first draft item generation and refinement, culminating in a second draft item set within each of the three scales. In step 2, qualitative interview data and quantitative item response data from the SL0023 clinical study (NCT02804763) [28], alongside consultations with SLE patient focus groups, informed refinement of the second draft item sets to achieve the final FATIGUE-PRO. In step 3, the final FATIGUE-PRO item sets were assessed quantitatively using the same SL0023 study data used in step 2 in order to assess how the refinements made in step 2 affected the performance of the scales (Fig. 1).

A multidisciplinary research team comprising clinicians (S.J.B., C.G., M.S.) with expertise in SLE and researchers with expertise in SLE provided input throughout instrument development. Specifically, clinicians contributed to decisions during first draft item generation and to the interpretation of findings to inform item revisions and finalization of the FATIGUE-PRO. Clinician input ensured that items were not only grounded in patient experience, but also clinically meaningful from a physician perspective.

Instrument development aligned with best practice recommendations detailed in FDA PRO Guidance [22, 23].

Patients

Initial patient interviews (step 1)
Patients were recruited for concept elicitation interviews through MyLupusTeam US (www.myLupusTeam.com; a freely accessible social network for people with SLE in the US) and for cognitive debriefing interviews via a market research company (ZS Associates, Evanston, IL, USA). Eligible patients were ≥18 years of age, had SLE and were currently receiving prednisone and hydroxychloroquine treatment. Eligible patients could also be receiving immunosuppressants, although patients

Fɪɢ. 1 Overview of the three-step MMP approach to instrument development
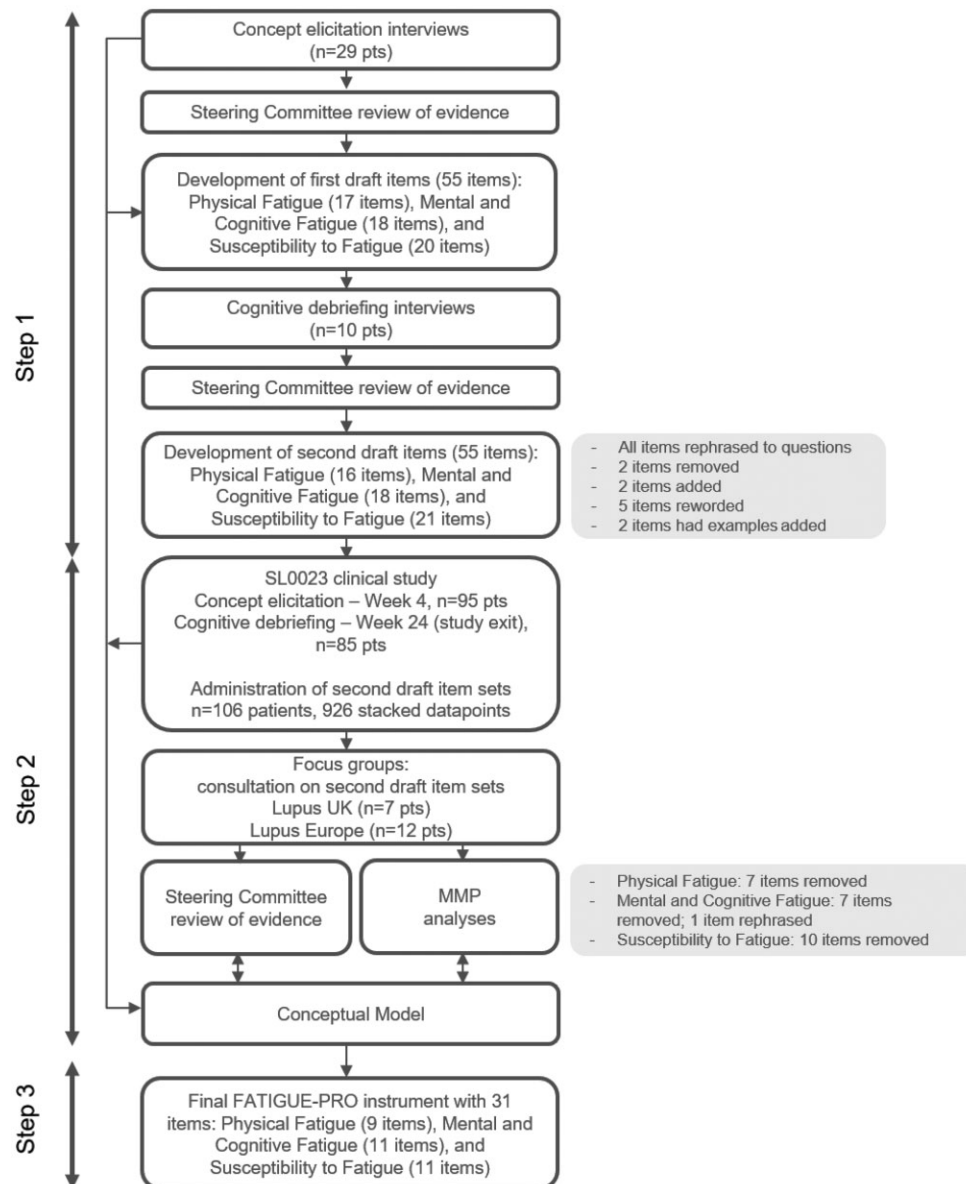
pts: patients.

receiving belimumab, rituximab, cyclophosphamide, warfarin or heparin were excluded (aligning with the SL0023 study eligibility criteria and to exclude patients with antiphospholipid syndrome). Eligibility was determined using a self-reported screening form. Patient data protection was ensured through the double-blinding of interviews, alongside other standard data protection processes.

Clinical trial patients (steps 2 and 3)
SL0023 was a multicentre, phase 2, randomized, double-blind, placebo-controlled study with an observational follow-up period. Eligible patients were ≥18 years of age, had a clinical diagnosis of SLE confirmed by SLICC classification criteria [29] and had moderate to severe SLE disease activity. Patients were required to be receiving stable SLE standard-of-care medication. Full eligibility criteria have been reported previously [28, 30].

Focus groups (step 2)
Two patient support groups, Lupus Europe (www.lupus-europe.org) and Lupus UK (www.lupusuk.org.uk), invited SLE patients to join focus groups in Belgium and the UK, respectively. The focus groups were run as consultations and patient data were not collected.

Ethical approval

Ethical approval was granted by the Copernicus Group institutional review board (IRB) and written informed consent was obtained prior to interviewing patients outside of the SL0023 study. The SL0023 study protocol, amendments and patient written informed consent were reviewed by a national, regional or independent ethics committee or IRB. This study was conducted in accordance with the current version of the applicable regulatory and International Council for Harmonization Good Clinical Practice requirements, the ethical principles that have their origin in the principles of the Declaration of Helsinki and the local laws of the countries involved. Focus groups consisted of a consultation by patient representatives without the collection of any data and therefore did not require ethical approval.

Analytical techniques

Qualitative methods

Interviews were conducted via telephone, following a semi-structured interview guide comprising both open-ended questions and specific probes. Concept elicitation interviews explored the experience of living with SLE; specifically, all fatigue-related symptoms and impacts. Cognitive debriefing interviews involved each item, its response scale and the item instructions being rated for clarity, relevance and/or interpretation. The interview design is described fully elsewhere [21].

Quantitative analyses

Rasch measurement theory (RMT) analyses were used to examine whether item response data achieved the requirements specified by the Rasch model in relation to targeting, response thresholds, item fit, item dependency and person separation index (PSI; Supplementary Table S1, available at *Rheumatology* online). RMT analyses were conducted using RUMM2030 software (RUMM Laboratory, Perth, WA, Australia).

Complementary psychometric evaluation using classical test theory (CTT) was used to evaluate instrument reliability (test–retest and internal consistency) and construct validity. For test–retest reliability, intraclass correlation coefficients were calculated between assessments 4 weeks apart (baseline–week 4 and week 4–week 8) in the subgroup of SL0023 patients whose overall SLE severity was reported as stable [change in Physician Global Assessment (PGA) <10 mm]. Internal consistency reliability coefficients were estimated using Cronbach's α. Construct validity was assessed by the association (using Spearman rank-order correlation coefficient) of instrument scores with measures of SLE disease activity [SLEDAI-2K score [31], PGA and Patient Global Assessment (PtGA)] and the LupusQoL fatigue domain score [10]. CTT analyses were conducted using SAS version 9.4 (SAS Institute, Cary, NC, USA).

Step 1: item generation and review

Step 1 is summarized in Fig. 1. First draft item generation was based on the patients' own experiences of fatigue, identified through concept elicitation interviews with SLE patients recruited by MyLupusTeam. Patient quotations were coded using thematic analysis in ATLAS.ti (https://atlasti.com/). Items utilized patients' own words wherever possible, while aiming for brevity. In consultation with the clinician experts, items were generated within the conceptual subdomains and domains of the recently established qualitative conceptual model for fatigue to form three distinct scales (Supplementary Fig. S1, available at *Rheumatology* online) [21]. Items were framed with a 7-day lookback period; a 5-point Likert frequency scale was used for item response options.

First draft items were reviewed in cognitive debriefing interviews by the SLE patients recruited by ZS Associates. The multidisciplinary research team reviewed cognitive debriefing findings to determine necessary item refinements, leading to a second draft of each fatigue scale. The second drafts were pilot tested among patients recruited in the SL0023 study, as described below.

Step 2: evaluation and refinement of the second draft item sets

The second draft item sets were evaluated through qualitative and quantitative data collected from a subset of English-speaking SLE patients enrolled in the SL0023 study (Fig. 1). Qualitative data were gathered through concept elicitation interviews at week 4 of the SL0023 study and cognitive debriefing interviews with patients on exiting the double-blind period at week 24. Interviews conducted with the clinical trial patients addressed the limitations of interviewing patients in step 1 whose eligibility was self-reported.

As a prespecified study endpoint, a subset of patients from the SL0023 study who spoke English or Spanish completed the second draft item sets at 10 study time points: every 4 weeks from week 1 (baseline)–week 24 (seven visits during the double-blind period) and every 8 weeks from weeks 32–48 (three visits during the observational follow-up period). Responses collected during the double-blind period were stacked and assessed using RMT analyses.

In addition to SL0023 study patients, SLE patients were consulted on the second draft item sets during two sequential focus groups to assess face and content validity outside of a clinical trial setting. Discussions covered the conceptual content coverage of the fatigue scales, as well as the relevance, clarity and potential conceptual overlap of items and the ordering of items within each scale to reflect fatigue severity.

The MMP approach underpinned second draft refinement, whereby qualitative and quantitative data were assessed against five prespecified MMP measurement principles: comprehensiveness of the item set, targeting of the item set for the context of use, conceptual

TABLE 1 Patient characteristics

Baseline characteristics	Step 1		Step 2	Steps 2 and 3
	Concept elicitation interviewees (*n* = 29)[a]	Cognitive debriefing interviewees (*n* = 10)[b]	SL0023 concept elicitation interviewees (*n* = 43)[c]	SL0023 quantitatively analysed patients (*n* = 106)
Female, *n* (%)	27 (93.1)	8 (80.0)	37 (86.0)	94 (88.7)
Age, years, mean (S.D.)	47.2 (12.3)	Not available[d]	44.2 (11.7)	40.8 (12.1)
Minimum–maximum	27–84		25.0–69.8	21.5–69.8
Disease duration, years, mean (S.D.)	9.9 (10.1)	Not available	9.9 (8.3)	8.2 (7.2)
Minimum–maximum	0–46		0.5–27.8	0.3–27.8
SLICC damage, *n* (%)	Not available	Not available		
0			23 (53.5)	73 (68.9)
1			9 (20.9)	18 (17.0)
>1			11 (25.6)	15 (14.2)
SLEDAI-2K, mean (S.D.)	Not available	Not available	10.9 (3.7)	11.3 (3.7)
Minimum–maximum			6–22	6–30

[a]Patients recruited by MyLupusTeam US. [b]Patients recruited by ZS Associates. [c]Cognitive debriefing interviewees (*n* = 38) were a subset of the 43 patients who took part in concept elicitation interviews at week 4 in the SL0023 study. [d]The number of patients in each of the following age groups was recorded: 18–25 years: 1 (10%), 26–40 years: 5 (50%), 41–50 years: 2 (20%), ≥51 years: 2 (20%).

uniqueness and singularity of the items, item quality and appropriateness of the response scale (Supplementary Table S2, available at *Rheumatology* online). Mixed methods evidence (MME) was reviewed by the multidisciplinary research team at a steering committee meeting to agree on the potential revision or deletion of items, resulting in the final FATIGUE-PRO scales taken forward for evaluation in step 3.

Step 3: evaluation of the final FATIGUE-PRO

RMT analyses conducted in step 2 were repeated on the final FATIGUE-PRO scales using the SL0023 study data (Fig. 1; Supplementary Table S1, available at *Rheumatology* online); complementary CTT analyses were also carried out on these data.

Results

Patient demographics and characteristics

During step 1, 29 patients with SLE completed concept elicitation interviews; another 10 patients completed cognitive debriefing interviews. For steps 2 and 3, 106 English- and Spanish-speaking patients in the SL0023 study (of 182 randomized patients) completed the second draft item sets across up to 10 time points, resulting in a total stacked sample of 926 quantitative responses. Additionally, in step 2, qualitative data were gathered from the English-speaking patients, including 43 who attended concept elicitation interviews at week 4, 38 of

whom attended cognitive debriefing interviews on exiting the double-blind period at week 24. Finally, in step 2, the Lupus Europe and Lupus UK focus groups were attended by 12 and 7 patients, respectively.

Demographics and baseline characteristics were comparable across the SL0023 study and interviewed patient groups (Table 1). Data were not collected for focus group attendees.

Step 1: first draft item generation and review

The first draft comprised 55 items across three scales reflecting the three conceptual domains of fatigue, including 17 PF items, 18 MF items and 20 SF items (Fig. 1). Feedback from the cognitive debriefing interviews and two clinicians led to the removal of two items and the addition of another two items to ensure all relevant concepts were covered, examples being added to two item stems to improve their interpretability, the rewording of five items to better reflect their intended meaning and improve clarity and the rephrasing of all items as questions prefixed by 'Over the past 7 days, how often did you…' (Fig. 1). The second draft taken forward to step 2 totalled 55 items (Supplementary Table S3, available at *Rheumatology* online): 16 PF items, 18 MF items and 21 SF items.

Step 2: evaluation and refinement of the second draft item sets

Qualitative analysis of SL0023 study interviews (Fig. 1) demonstrated that SLE patients generally found second

TABLE 2 The subset of items associated with issues identified through MMP analysis in step 2

Problematic second draft items	MMP criteria which items failed to achieve				
Over the past 7 days, how often did you...	1: Comprehensiveness of item set	2: Targeting of the item set for the COU	3: Conceptual uniqueness	4: Quality of item	5: Appropriateness of response scale
Physical fatigue					
PF01: Feel lightheaded?		MME	MME	MME	
PF02: Feel dizzy?		MME	MME	MME	
PF04: Feel that you had no strength in your muscles?	Quantitative				
PF12: Feel drained?			MME		
PF14: Feel like your body could not keep up with what you wanted to do?			Quantitative		
PF15: Feel that it was hard to move your body?			Quantitative		
PF16: Feel physically exhausted?			MME		
Mental and cognitive fatigue					
MF17: Feel less alert than you usually are?	MME			MME	
MF18: Feel that your thinking was slowed down?	MME		MME		
MF20: Feel like you could not gather your thoughts?			Quantitative		
MF21: Feel you were forgetful (for example, missed appointments)?			MME	MME	
MF22: Feel you could not remember things (for example, had trouble recalling information)?[a]				MME	
MF25: Feel you could not think clearly?			MME		
MF27: Feel it was hard for me to organize my thoughts?				MME	
MF33: Feel mentally overwhelmed?[b]			Quantitative		
MF34: Feel mentally exhausted?			MME		
Susceptibility to fatigue					
SF35: Go to bed earlier than you usually do?				MME	
SF36: Feel it took longer to fully wake up in the morning?				Multidisciplinary research team decision[c]	
SF38: Feel you needed more sleep than you usually do?			Qualitative	Qualitative	
SF41: Feel that everyday activities left you feeling tired?			MME		
SF44: Feel more tired early in the day than you usually do?			Qualitative	Qualitative	
SF47: Feel a sudden need for a short rest?			Qualitative	Qualitative	
SF48: Feel a sudden need to lie down?				Qualitative	
SF49: Feel so exhausted you could barely keep your eyes open?	MME		MME		
SF50: Feel exhausted even though you had not done much?	Qualitative				
SF51: Wake up feeling exhausted?	MME		MME		

Table details evidence indicating issue(s) with each item; empty cells indicate where items met MMP criteria. Of the listed items, only MF22 and MF33 were included in the final FATIGUE-PRO instrument; italicized items were included in a long-form item set. [a]Item MF22 was retained after the removal of the problematic element '(for example, ...)'. [b]Item MF33 was retained because patients identified the item as conceptually unique. [c]The issue of quality with item SF36 was identified solely by the multidisciplinary research team based on the problematic comparative term 'longer'. COU: context of use.

TABLE 3 RMT analysis results for second draft item sets and the final FATIGUE-PRO item sets

Question		Physical fatigue		Mental and cognitive fatigue		Susceptibility to fatigue	
		Second draft item set (16 items)	Final FATIGUE-PRO (9 items)	Second draft item set (18 items)	Final FATIGUE-PRO (11 items)	Second draft item set (21 items)	Final FATIGUE-PRO (11 items)
Targeting: How adequate is the scale-to-sample targeting?[a]		93	89	84	83	85	86
Response thresholds: Do the response categories work as intended?		100	100	100	100	100	100
Item fit: To what extent do the items work together to define a single measurement construct?	Fit residuals	50	56	56	82	48	73
	Chi-square[b]	88	100	94	100	90	91
Item dependency: To what extent are the items locally independent?		93	92	92	98	97	98
PSI: Are patients in the sample separated by the scale items?[c]		0.95	0.92	0.95	0.93	0.97	0.94

Values reported as % success unless stated otherwise. Higher percentages indicate better findings. [a]Estimated using the percentage of individual sample measurements ($n = 916$) covered by the scale range. [b]Chi-squared estimates computed on an adjusted sample of $n = 500$. [c]PSI was reported on a scale from 0 to 1: 0 = all error, 1 = no error.

draft items clear, easy to complete and a relevant reflection of their fatigue. However, some issues were identified, including with clarity and conceptual overlap (Table 2). Similarly, quantitative RMT analyses of responses from SL0023 study patients (Fig. 1) showed that item sets had good psychometric properties (Table 3), although specific issues were identified for some items, most commonly in relation to item fit and dependency (Table 2).

Consultation with SLE patients at the focus groups endorsed the relevance and clarity of the items among SLE patients beyond a clinical trial setting. The clinicians at the steering committee meeting scrutinized and endorsed the clinical relevance of the MME of psychometric issues (Table 2; assessed against the MMP criteria in Supplementary Table S2, available at *Rheumatology* online), summarized as follows.

Within the PF scale, MME highlighted issues with four items—all four suffered from conceptual overlap, while items PF01 ('Feel lightheaded?') and PF02 ('Feel dizzy?') had issues relating to targeting/relevance and quality. Issues with three additional PF items were identified by quantitative evidence alone, including two items with conceptual overlap and one with an issue of item set comprehensiveness (Table 2). Although the separate items relating to the feeling of weakness or heaviness in each of the arms, legs and whole body (PF06–11) were specifically probed for potential overlap during cognitive debriefing, both the qualitative and quantitative data supported their retention as unique items. Qualitatively, patients provided descriptions of how the two concepts (weakness and heaviness) differ, and quantitatively, the items were located

on different points of the fatigue severity continuum, which suggested that heaviness is linked with higher levels of fatigue compared with weakness. No changes were made to items PF06–11 since MME indicated that they contributed unique information to the PF score.

Within the MF scale, MME demonstrated issues with seven items; conceptual overlap was the most common, followed by issues with item quality and item set comprehensiveness. For example, item MF21 ['Feel you were forgetful (for example, …)?'] had issues with conceptual overlap and quality demonstrated by MME. Quantitative evidence alone uncovered conceptual overlap with an additional two MF items (Table 2). However, qualitative evidence indicated that one of these items, MF33 ('Feel mentally overwhelmed?'), was conceptually unique from the patients' perspective; the steering committee therefore decided to retain this item.

Lastly, within the SF scale, MME identified issues with four items, while an additional five items were associated with issues through qualitative evidence alone (Table 2). Again, conceptual overlap was most frequent, followed by issues with item quality and item set comprehensiveness.

As well as endorsing the issues identified from the analysis of patient data, the multidisciplinary research team identified an issue with item SF36 ('Feel it took longer to fully wake up in the morning?'), due to the inclusion of the comparative term 'longer' (Table 2).

Following detailed assessment of this evidence at the steering committee meeting, the final FATIGUE-PRO item sets were formulated by removing 24 of the aforementioned items and rephrasing one, MF22 ['Feel you could

FIG. 2 Scale-to-sample targeting and threshold mapping for the final FATIGUE-PRO scales
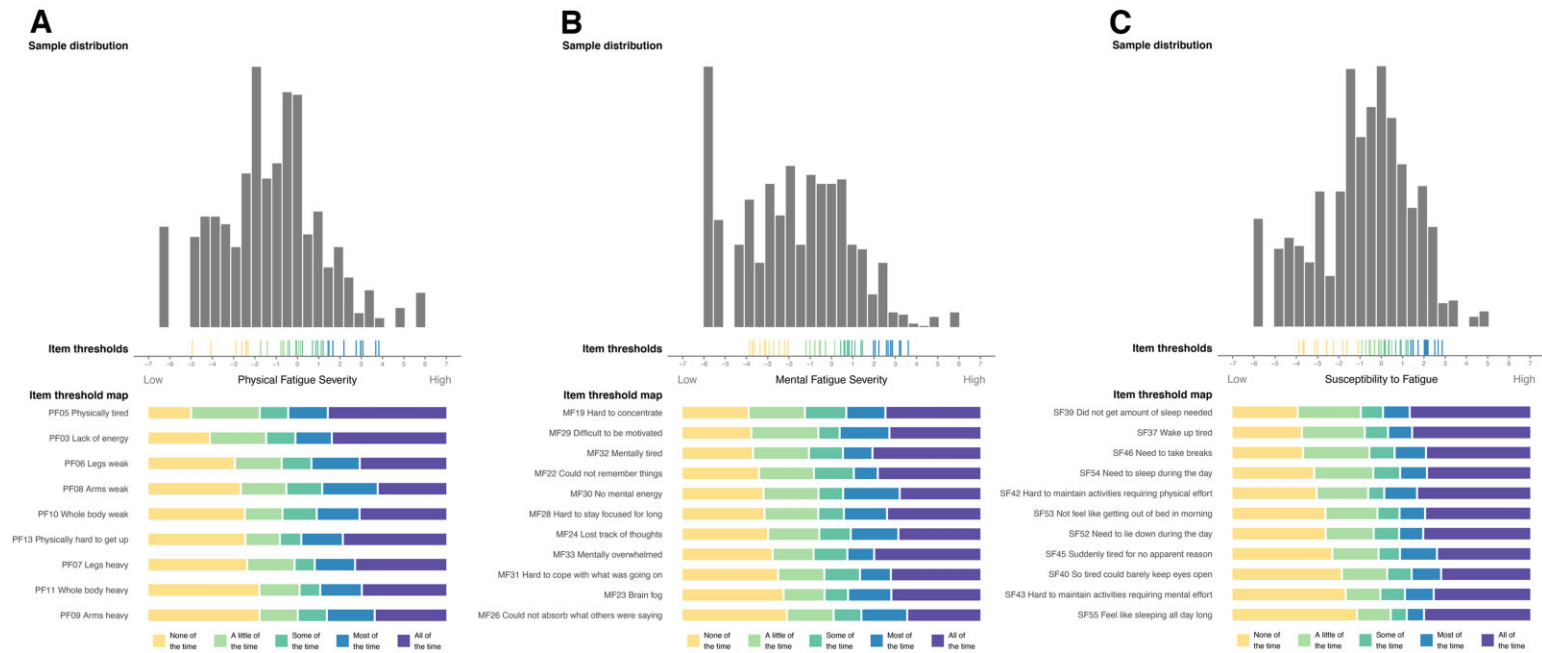
Sample distributions (grey blocks representing total score distributions for each FATIGUE-PRO scale) and item threshold distributions (coloured blocks) plotted on the same linear measurement continuum of fatigue severity. Item threshold maps show the most probable response category across the measurement continuum, with thresholds between coloured blocks reflecting locations where adjacent response categories are equally likely. Targeting is assessed by examining the relative range and coverage of the sample distribution by item thresholds. Items (abbreviated; full names listed in Supplementary Table S4, available at *Rheumatology* online) ordered top to bottom by increasing difficulty. RMT expects ordering of response categories to reflect intended severity.

TABLE 4 CTT analysis results for the final FATIGUE-PRO

Scales	Internal consistency reliability[a] (n = 104)	Test–retest reliability[b]		Correlation with SLEDAI-2K score[c] (n = 104)	Correlation with PGA (n = 104)[c]	Correlation with PtGA (n = 104)[c]	Correlation with LupusQoL fatigue domain score (n = 104)[c,d]
		Between baseline and week 4 (n = 41)	Between week 4 and week 8 (n = 58)				
Physical fatigue	0.95	0.80	0.71	0.02	0.44	0.41	−0.68
Mental and cognitive fatigue	0.96	0.82	0.79	−0.17	0.36	0.26	−0.57
Susceptibility to fatigue	0.95	0.71	0.85	−0.03	0.41	0.25	−0.63

[a]Cronbach's α. [b]Intraclass correlation coefficient. [c]Spearman correlation coefficient. [d]Negative correlation coefficients reflect the opposing directions of the LupusQoL and FATIGUE-PRO scales, i.e. higher scores represent better outcomes for LupusQoL but worse outcomes for FATIGUE-PRO.

not remember things (for example, …)?'] by removing '(for example, …)' to resolve the quality issue identified through MME (Table 2). Distinct long-form item sets were also generated, which included the final FATIGUE-PRO items and items for which issues were demonstrated through only qualitative or quantitative data.

In summary, 31 items were included in the final FATIGUE-PRO: 9 PF items, 11 MF items and 11 SF items (Supplementary Table S4, available at *Rheumatology* online).

Step 3: evaluation of the final FATIGUE-PRO

Within the RMT analyses of the FATIGUE-PRO, items demonstrated adequate targeting within each of the three scales, as item thresholds covered 89% of the sample measurements for PF, 83% of the sample measurements for MF and 86% of the sample measurements for SF (Table 3). Across the scales there were no substantial gaps in coverage and item bunching was limited (Fig. 2). The item response option thresholds were ordered for all items, indicating that the proposed response scale worked as intended (Fig. 2; Table 3). Item fit statistics were very good across the three scales and most items were found to be locally independent of each other (PF 92%, MF 98%, SF 98%; Table 3). PSI remained high (>0.90) across the three scales (Table 3).

Internal consistency reliability was excellent for the three scales (PF 0.95, MF 0.96, SF 0.95; Table 4). Test–retest reliability of the scales was adequate, ranging from 0.71 to 0.85 depending on the time period and scale considered (Table 4). FATIGUE-PRO scores consistently showed poor correlations with the SLEDAI-2K and moderate correlations with the PGA (Table 4). The highest correlations were observed with the LupusQoL fatigue domain score, ranging from −0.57 (MF) to −0.68 (PF) (Table 4).

The FATIGUE-PRO does not generate a total score; instead, scores are calculated for each scale on a 5-point frequency scale. The sums of item scores are linearly transformed to provide scores ranging from 0 to 100, with higher scores indicating higher levels of fatigue. Further details of the FATIGUE-PRO scoring are available in the scoring manual.

Discussion

The FATIGUE-PRO is a new PRO instrument for measuring fatigue in patients with SLE. The instrument takes 5–10 min to complete and generates separate scores for three fatigue scales—physical fatigue, mental and cognitive fatigue and susceptibility to fatigue—based on a recently established conceptual model of fatigue in SLE [21]. The FATIGUE-PRO benefits from this model, which provided a clear framework for instrument development, and from continual and extensive input from patients and clinicians to ensure meaningful content validity and clinical relevance. These features are essential components of PRO development [24], as specified in best practice guidance [22, 23], but they have been overlooked in many existing fatigue measures used in SLE that consequently lack conceptual clarity and suffer from psychometric limitations [2, 3]. Overall, the FATIGUE-PRO addresses an unmet need for an instrument that encapsulates the conceptually complex and multifaceted nature of fatigue in SLE while achieving good psychometric performance. The FATIGUE-PRO was primarily developed for use in clinical trials, with the intention to develop an electronic assessment system to better facilitate future use in clinical practice.

Psychometric analyses demonstrated the strong measurement properties of the FATIGUE-PRO scales for quantifying fatigue in SLE. Specifically, items were well targeted to SLE, with good coverage of the fatigue

issues relevant to patients. Additionally, items were clear, conceptually unique and cohesive within each of the three scales, and patients could accurately discriminate between the five frequency levels of the item response scale. High PSI and internal consistency values indicated that the FATIGUE-PRO can reliably separate patients with different levels of fatigue within the sample. Each FATIGUE-PRO scale was designed to cover a specific fatigue concept; their moderate correlations with the LupusQoL fatigue score (a summary assessment combining different components of fatigue) indicated that different, but related, constructs are measured, confirming the initial basis for the development of the FATIGUE-PRO. Qualitative interviews and consultations with patients demonstrated that the instrument had good face and content validity, was easily understood and accurately represented their symptom experience.

The FATIGUE-PRO benefits from a rigorous concept-driven and patient-centred development approach, enabling the instrument to be optimized for measuring fatigue in SLE. Specifically, the large number of SLE patients interviewed aided the conceptualization of fatigue and subsequent item generation and refinement and ensured the FATIGUE-PRO comprehensively and accurately represents the patient experience of fatigue in SLE [24]. Indeed, compared with existing PRO instruments, the FATIGUE-PRO captures additional but complementary concepts of fatigue. This study further benefitted from rigorous analytical techniques and an MMP approach that enabled the synthesis of qualitative and quantitative evidence, which subsequently informed decisions on item refinement [26]. Development was further strengthened by the multidisciplinary team of clinicians and researchers with expertise in SLE and outcome measurement, respectively, which ensured the clinical relevance and psychometric effectiveness of the FATIGUE-PRO.

In line with previous findings linked to both multi- and unidimensional PRO instruments that measure fatigue [2], the FATIGUE-PRO did not demonstrate an association with the SLEDAI-2K measure of disease activity. Although the link between fatigue and disease activity remains complex, this finding may reflect either a lack of or a weak association between levels of fatigue and disease activity, or specifically between fatigue and disease activity as measured by the SLEDAI-2K, as has been hypothesized previously [2]. This rationale may also explain the weak correlations between the FATIGUE-PRO (particularly the MF and SF scales) and PGA. Future research should aim to further explore the association between fatigue and disease activity as measured by these instruments.

While the SL0023 study provided a valuable opportunity to test item sets as a study endpoint, the large proportion of clinical trial patients involved in the FATIGUE-PRO development was also a limitation, since these patients may not be representative of the wider population of SLE patients. Another limitation was that instrument development was based on qualitative data from only English-speaking patients and quantitative data from only English- and Spanish-speaking patients from a limited number of countries.

Future research should aim to improve our understanding of score interpretation and the relationship between components of fatigue, partly through the identification of meaningful thresholds of fatigue severity. Evaluation of the FATIGUE-PRO in real-world settings is needed to further validate the applicability of the instrument in clinical practice. Additional research with non-English-speaking patients and patients from different countries and cultural backgrounds is also needed to validate the instrument cross-culturally. To date, the FATIGUE-PRO has been used in two clinical outpatient settings in the UK and Germany; it is also being used alongside FACIT-F in a phase 3 clinical study of patients with moderate–severe SLE (NCT04294667). Data from these studies will be utilized to confirm the FATIGUE-PRO measurement properties and determine thresholds for clinically meaningful within-patient changes in scores. Finally, we note the potential for the FATIGUE-PRO to be applied in other disease areas (e.g. other autoimmune conditions, neurological diseases or even 'long COVID') [32–38] as an exciting avenue of research warranting exploration. Initially, the relevance of the FATIGUE-PRO concepts and the usability of the instrument in other diseases will require validation.

Conclusion

In conclusion, the FATIGUE-PRO, comprising 31 items across the three scales of physical fatigue, mental and cognitive fatigue and susceptibility to fatigue, represents a significant advance for patient-centred care and research in SLE. Primarily the instrument will allow patients' fatigue to be better understood, more accurately quantified and effectively treated in clinical trials and clinical practice [39, 40]. Beyond the direct benefits to patients assessed using the FATIGUE-PRO, all SLE patients may benefit longer term from enhanced fatigue measurement capability in assessments of treatment efficacy, which may help researchers to better identify effective therapies to manage fatigue.

Acknowledgements

Data availability statement

The FATIGUE-PRO is copyrighted by UCB Pharma but will be made available for public use in the ePROVIDE database; it is currently available in >60 languages. Underlying data from this article may be requested by qualified researchers 6 months after study drug approval in the USA and/or Europe, or global development is discontinued, and 18 months after trial completion. Investigators may request access to anonymized IPD and redacted study documents that may include raw datasets, analysis-ready datasets, study protocol, blank case report forms, annotated case report forms, statistical analysis plan, dataset specifications and clinical study report. Prior to use of the data, proposals need to be approved by an independent review panel at www. Vivli.org and a signed data sharing agreement will need to be executed. All research documentation is available in English only, for a prespecified time, typically 12 months, on a password-protected portal.

Supplementary data

Supplementary data are available at *Rheumatology* online.

References

1 Durcan L, O'Dwyer T, Petri M. Management strategies and future directions for systemic lupus erythematosus in adults. Lancet 2019;393:2332–43.

2 Cleanthous S, Tyagi M, Isenberg DA, Newman SP. What do we know about self-reported fatigue in systemic lupus erythematosus? Lupus 2012;21:465–76.

3 Cleanthous S, Regnault A, Schneider M *et al.* AB0633 measuring what matters to lupus patients: translating patient views into novel patient-reported outcomes. Ann Rheum Dis 2018;77:1464.

4 Lupus and Allied Diseases Association, Lupus Foundation of America, Lupus Research Alliance. Lupus: Patient Voices. http://lupuspfdd.org/LupusPatientVoicesFINAL.pdf (17 December 2020, date last accessed).

5 Touma Z, Hoskin B, Atkinson C *et al.* Systemic lupus erythematosus symptom clusters and their association with patient-reported outcomes and treatment: analysis of real-world data. Arthritis Care Res (Hoboken) 2020; doi: 10.1002/acr.24546.

6 Mertz P, Schlencker A, Schneider M *et al.* Towards a practical management of fatigue in systemic lupus erythematosus. Lupus Sci Med 2020;7:e000441.

7 Arnaud L, Mertz P, Amoura Z *et al.* Patterns of fatigue and association with disease activity and clinical manifestations in systemic lupus erythematosus. Rheumatology (Oxford) 2021;60:2672–7.

8 Sterling K, Gallop K, Swinburn P *et al.* Patient-reported fatigue and its impact on patients with systemic lupus erythematosus. Lupus 2014;23:124–32.

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

10 McElhone K, Abbott J, Shelmerdine J *et al.* Development and validation of a disease-specific health-related quality of life measure, the LupusQol, for adults with systemic lupus erythematosus. Arthritis Rheum 2007;57:972–9.

11 Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus. Arch Neurol 1989;46:1121–3.

12 Horisberger A, Courvoisier D, Ribi C. The Fatigue Assessment Scale as a simple and reliable tool in systemic lupus erythematosus: a cross-sectional study. Arthritis Res Ther 2019;21:80.

13 Barbacki A, Petri M, Aviña-Zubieta A, Alarcón GS, Bernatsky S. Fatigue measurements in systemic lupus erythematosus. J Rheumatol 2019;46:1470–7.

14 Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.

15 Raymond K, Park J, Joshi AV, White MK. Patient experience with fatigue and qualitative interview-based evidence of content validation of the FACIT-Fatigue in systemic lupus erythematosus. Rheumatol Ther 2021;8: 541–54.

16 Dittner AJ, Wessely SC, Brown RG. The assessment of fatigue: a practical guide for clinicians and researchers. J Psychosom Res 2004;56:157–70.

17 Cleanthous S, Bongardt S, Marquis P *et al.* Psychometric analysis from EMBODY1 and 2 clinical trials to help select suitable fatigue PRO scales for future systemic lupus erythematosus studies. Rheumatol Ther 2021;8:1287–301.

18 Hobart J, Cano S, Baron R *et al.* Achieving valid patient-reported outcomes measurement: a lesson from fatigue in multiple sclerosis. Mult Scler 2013;19:1773–83.

19 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;12:116.

20 Cella D, Yount S, Sorensen M *et al.* Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. J Rheumatol 2005;32: 811–9.

21 Cleanthous S, Strzok S, Haier B, Cano S, Morel T. The patient experience of fatigue in systemic lupus erythematosus: a conceptual model. Rheumatol Ther 2021;doi: 10.1007/s40744-021-00374-0.

22 U.S. Food and Drug Administration. Patient-reported outcome measures: use in medical product development to support labeling claims. 2009. https://www.fda.gov/media/77832/download (17 December 2020, date last accessed).

23 U.S. Food and Drug Administration. Roadmap to patient-focused outcome measurement in clinical trials. 2015. https://www.fda.gov/media/87004/download (17 December 2020, date last accessed).

24 Lasch KE, Marquis P, Vigneux M *et al.* PRO development: rigorous qualitative research as the crucial foundation. Qual Life Res 2010;19:1087–96.

25 U.S. Food and Drug Administration. Systemic lupus erythematosus—developing medical products for treatment. 2010. https://www.fda.gov/media/71150/download (17 December 2020, date last accessed).

26 Morse JM. Evolving trends in qualitative research: advances in mixed-method design. Qual Health Res 2005;15:583–5.

27 Fisher WP, Stenner AJ. Integrating qualitative and quantitative research approaches via the phenomenological method. Int J Mult Res Approaches 2011;5:89–103.

28 Furie RA, Bruce IN, Dörner T *et al.* Phase 2, randomized, placebo-controlled trial of dapirolizumab pegol in patients with moderate-to-severe active systemic lupus erythematosus. Rheumatology (Oxford) 2021;60: 5397–407.

29 Petri M, Orbai AM, Alarcón GS *et al.* Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64:2677–86.

30 Furie R, Bruce IN, Dörner T *et al.* FRI0195 Efficacy and safety of dapirolizumab pegol (DZP) in patients with moderately to severely active systemic lupus erythematosus (SLE): a randomised, placebo (PBO)-controlled study. Ann Rheum Dis 2019;78:775–6.

31 Gladman DD, Ibañez D, Urowitz MB. Systemic Lupus Erythematosus Disease Activity Index. J Rheumatol 2002;29:288–91.

32 Braley TJ, Chervin RD. Fatigue in multiple sclerosis: mechanisms, evaluation, and treatment. Sleep 2010;33: 1061–7.

33 Kluger BM, Herlofson K, Chou KL *et al.* Parkinson's disease-related fatigue: a case definition and recommendations for clinical research. Mov Disord 2016;31: 625–31.

34 Kluger BM, Krupp LB, Enoka RM. Fatigue and fatigability in neurologic illnesses: proposal for a unified taxonomy. Neurology 2013;80:409–16.

35 Penner IK, Paul F. Fatigue as a symptom or comorbidity of neurological diseases. Nat Rev Neurol 2017;13: 662–75.

36 Marshall M. The lasting misery of coronavirus long-haulers. 2020. https://www.nature.com/articles/d41586-020-02598-6 (22 March 2021, last accessed).

37 Davis HE, Assaf GS, McCorkell L *et al.* Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. EClinicalMedicine 2021;38: 101019.

38 Collins FS. NIH launches new initiative to study "long COVID". 2021. https://www.nih.gov/about-nih/who-we-are/nih-director/statements/nih-launches-new-initiative-study-long-covid (22 March 2020, date last accessed).

39 Strand V, Chu AD. Measuring outcomes in systemic lupus erythematosus clinical trials. Expert Rev Pharmacoecon Outcomes Res 2011;11:455–68.

40 Izadi Z, Gandrup J, Katz PP, Yazdany J. Patient-reported outcome measures for use in clinical trials of SLE: a review. Lupus Sci Med 2018;5:e000279.