



Combining online and in-person methods to evaluate the content validity of PROMIS fatigue short forms in rheumatoid arthritis

S. J. Bartlett^{1,2} · A. K. Gutierrez^{2,3} · A. Butanis² · V. P. Bykerk⁴ · J. R. Curtis⁵ · S. Ginsberg⁶ · A. L. Leong⁷ · A. Lyddiatt⁸ · W. B. Nowell⁶ · A. M. Orbai² · K. C. Smith⁹ · C. O. Bingham 3rd²

Accepted: 8 May 2018 / Published online: 24 May 2018
© Springer International Publishing AG, part of Springer Nature 2018

Abstract

Purpose Fatigue is frequent and often severe and disabling in RA, and there is no consensus on how to measure it. We used online surveys and in-person interviews to evaluate PROMIS Fatigue 7a and 8a short forms (SFs) in people with RA.

Methods We recruited people with RA from an online patient community ($n = 200$) and three academic medical centers ($n = 84$) in the US. Participants completed both SFs then rated the comprehensiveness and comprehensibility of the items to their fatigue experience. Cognitive debriefing of items was conducted in a subset of 32 clinic patients. Descriptive statistics were calculated, and associations were evaluated using Pearson and Spearman correlation coefficients.

Results Mean SF scores were similar ($p \geq .61$) among clinic patients reflecting mild fatigue (i.e., 54.5–55.9), but were significantly higher ($p < .001$) in online participants. SF Fatigue scores correlated highly ($r \geq 0.82$; $p < .000$) and moderately with patient assessments of disease activity ($r \geq 0.62$; $p = .000$). Most (70–92%) reported that the items “completely” or “mostly” reflected their experience. Almost all ($\geq 94\%$) could distinguish general fatigue from RA fatigue. Most ($\geq 85\%$) rated individual items questions as “somewhat” or “very relevant” to their fatigue experience, averaged their fatigue over the past 7 days (58%), and rated fatigue impact versus severity (72 vs. 19%). 99% rated fatigue as an important symptom they considered when deciding how well their current treatment was controlling their RA.

Conclusions Results suggest that items in the single-score PROMIS Fatigue SFs demonstrate content validity and can adequately capture the wide range of fatigue experiences of people with RA.

Keywords Rheumatoid arthritis · Fatigue · PROMIS · Content validity

Background

In people with rheumatoid arthritis (RA), fatigue is a primary symptom that is frequent, highly variable, and often severe and disabling which impacts multiple aspects of

health, work and social participation, and quality of life [1–5]. Fatigue in RA often results from systemic inflammation, limitation in joint mobility, and other factors including excess weight, poor or interrupted sleep, anxiety, depression, and stress. The experience of fatigue is often variable among

✉ S. J. Bartlett
Susan.bartlett@mcgill.ca

¹ Center for Health Outcomes Research, McGill University, 5252 de Maisonneuve West, #3D-57, Montreal, QC H4A 3S5, Canada

² Division of Rheumatology, Johns Hopkins School of Medicine, Mason F Lord Tower, 5200 Eastern Avenue, Rm 404, Baltimore, MD 21224, USA

³ Ateneo School of Medicine and Public Health, Pasig City, Philippines

⁴ Hospital for Special Surgery, Weill Cornell Medical College, 525 East 71st St, 7th floor, New York, NY 10021, USA

⁵ Division of Rheumatology, University of Alabama at Birmingham, Birmingham, AL, USA

⁶ Global Healthy Living Foundation, Upper Nyack, NY, USA

⁷ Healthy Motivation, Santa Barbara, CA, USA

⁸ Patient Partners, London, ON, Canada

⁹ Johns Hopkins Bloomberg School of Public Health Center for Qualitative Studies in Health and Medicine, Baltimore, MD, USA

and even varies from one person with RA to another. Thus, there are different approaches to conceptualizing fatigue, and little consensus on how to measure it [6–9].

The Patient-Reported Outcome Measurement Information System (PROMIS) was developed by the National Institutes of Health (NIH) to enhance measurement of physical, emotional, and social health across chronic conditions [10, 11]. PROMIS developers defined fatigue as an overwhelming, debilitating, and sustained sense of exhaustion that interferes with daily activities, work, and family or social roles [12]. The PROMIS Fatigue item bank includes > 90 items that ask about fatigue severity and its impact on day-to-day function. PROMIS Fatigue tools can be administered using computer adaptive testing (CAT) or with fixed-item Short Forms (SFs) containing 4, 7, or 8 items. The 7 and 8-item SFs contain non-overlapping items; those in the 4-item version are also contained in the 8 item SF. Items have been calibrated in the U.S. general population and some health conditions [13, 14]. Scores are reported using a T-score metric, where the population mean is 50 and the standard deviation is 10. We previously reported that the Fatigue CAT scores correlated moderately strongly with other indicators of disease activity, a fatigue visual analogue scale (VAS), and increased with worsening disease [15]. As many settings have practical constraints limiting widespread implementation of CATs, SFs may be preferable to administer in some situations [16, 17].

Before widespread use of Fatigue SFs can be recommended in specific patient populations, evidence is required that items are relevant and can capture the full range of patient experiences [18, 19]. Although items in the Fatigue item bank were debriefed according to PROMIS standards in 22 patients with a range of medical conditions [12], the degree to which they are comprehensive and easily comprehended is unstudied. To address this gap, we evaluated content validity by reviewing items in the PROMIS Fatigue SFs 7a and 8a with a diverse group of people with RA using online surveys and in-person cognitive debriefing interviews.

Methods

Study design

To assemble a sample with diverse sociodemographic and RA characteristics, we recruited participants from several sources. We invited RA patients receiving treatment and enrolled in ongoing observational studies at academic medical centers as well as individuals with self-identified RA affiliated with an online arthritis community to complete an online survey. The survey included SF items and additional questions about the comprehensiveness and comprehensibility of the items in relation to their fatigue experience. We conducted cognitive debriefing interviews

in a separate group recruited from three academic arthritis centers. The survey was approved by the Johns Hopkins Institutional Review Board (IRB; 00059765). The debriefing was conducted with central oversight from the Johns Hopkins IRB (00059930), and at each site. Informed consent was obtained from all participants.

Participants

Online survey

A total of 57 adults, aged 18 years and older, who were fluent in English, had been diagnosed with RA, and were enrolled in an observational trial at the Johns Hopkins Arthritis Center, were consecutively approached by study staff; 52 agreed to participate. We also partnered with an online arthritis community (<https://CreakyJoints.org/>) to recruit participants with RA (hereafter referred to as the “online” participants) via email. Online participants were screened using inflammatory arthritis questions from the Connective Tissue Screening Questionnaire [20] which we adapted to include the 2010 ACR/EULAR RA criteria [21]. They also completed a DMARD checklist and answered questions to help identify and exclude those with a personal or family history of psoriasis or psoriatic arthritis (PsA) [22]. We tested our two-step screening approach using medical records of a convenience sample of 52 patients receiving care for RA and PsA from our general arthritis clinic and found this approach had 100% specificity for identifying people with RA.

The survey was conducted from April to September 2015. On the welcome page, we stated that the survey was voluntary and anonymous, and that completion was interpreted as providing informed consent. After providing information about age, sex, race/ethnicity, education, employment, year of diagnosis, and type of arthritis care provider (rheumatologist vs. other), online participants completed a self-assessment of their current disease activity and the two PROMIS Fatigue SFs (v1.0). They then rated how easy the questions were to understand and answer. We asked if the SF items covered the type of fatigue they experienced (current or past), and if additional questions were needed to fully capture their experience. To better understand how participants conceptualized their fatigue and selected responses, we asked if they distinguished overall fatigue from fatigue due to their RA; how much of their fatigue they attributed to RA (all, most, some, a little, none, can't tell); and whether their answers would be different if asked to rate only RA-related fatigue (yes/no). We also asked if their fatigue level provided important information about how effectively their current treatment was controlling their RA.

Cognitive debriefing interviews of individual items

Cognitive debriefing participants were recruited from three academic arthritis clinics in Baltimore MD, New York NY, and Birmingham AL. Interviews were conducted from September through December 2015. Debriefing participants had the same sociodemographic and RA information described above and completed both SFs. They then were randomly assigned to be debriefed on items in either SF using a random numbers generator. Two trained interviewers (AB, KG) conducted face-to-face or phone interviews where they followed a script to guide participants to “talk-through” how they interpreted and answered each of the items [19]. They answered questions about how they thought about their fatigue when selecting responses whether they mostly considered the intensity or impact of their fatigue, or both, affected their response selection; and if their answers reflected their fatigue at its worst or its average over the past 7 days. Conversations were audiotaped and transcribed. We conducted a targeted and pragmatic qualitative analysis of the interviews to descriptively and thematically summarize the information.

Statistical approach

Descriptive statistics were calculated and t-tests and Chi-square were used to compare groups. Pearson and Spearman’s correlations were calculated to evaluate the relationships. Free text responses were summarized. The PROMIS Assessment Center Scoring Service was used to obtain IRT-calculated scores. As characteristics were similar among participants from the three academic centers, data were collapsed for subsequent analysis. Statistical analysis was performed using SPSS V24.0 and a $p < .05$ was considered statistically significant.

Results

Participants were from regions across the US, and were mostly female, white, and middle aged (Table 1). Most had attended or completed college and had lived with RA for a decade on average. Among online participants, 305 unique individuals were screened, and 95 (31%) were excluded mostly (58%) due to a personal/family history of psoriasis. Of 210 (69%) who met eligibility criteria, 200 (66%) completed the survey. Almost all online participants (193/200) reported their RA was managed by a rheumatologist.

Table 1 Participant characteristics

Characteristics [mean (SD) or <i>n</i> (%)]	Surveys		Debriefing interviews**
	Arthritis clinic*	Arthritis community	
<i>N</i>	52	200	32
Age (years)	53 (14)	51 (12)	54 (13)
Female	45 (87%)	167 (84%)	23 (72%)
White	46 (89%)	186 (94%)	18 (56%)
Hispanic/latino ethnicity	2 (4%)	11 (6%)	3 (9%)
Education > high school	36 (69%)	185 (93%)	31 (97%)
Disabled due to RA	8 (15%)	63 (32%)	5 (16%)
Use of aids or disability-related devices	26 (50%)	136 (68%)	15 (50%)
Region***			
Northeast	100%	34%	24 (75%)
South		18%	8 (25%)
Midwest		13%	
West		27%	
Outside US		8%	
Urban	33 (64%)	147 (74%)	23 (72%)
RA disease duration (years)	15 (11)	10 (10)	13 (10)
Patient global disease activity (0–100)	34 (30)	57 (23)	34 (26)
PROMIS fatigue 7a <i>T</i> -score	54.8 (13.6)	65.6 (8.1)	53.2 (9.9)
PROMIS fatigue 8a <i>T</i> -score	54.6 (11.2)	66.0 (7.8)	55.3 (10.3)

*Recruited from Johns Hopkins Arthritis Center, Baltimore, MD

**Recruited from inflammatory arthritis clinics at Johns Hopkins, Baltimore, MD ($n = 12$); Hospital for Special Surgery, New York, NY ($n = 12$); and University of Alabama at Birmingham ($n = 8$)

***Based on U.S. census regions

As compared with participants recruited from clinics ($N=82$), online participants had more years of education, a shorter disease duration, were more likely to report they were disabled due to RA and used disability aids and devices, and had higher disease activity scores.

Survey responses

Mean PROMIS Fatigue SF scores were similar ($p=.93$) among clinic patients and reflected mild levels of fatigue (i.e., 53.2–54.8), but were significantly higher (i.e., > 1 SD; $p=.000$) in online participants (Table 1). Scores on the two SF were highly correlated (r values from 0.82 to 0.94; p values < 0.000) and the mean difference between the 7a and 8a was -0.5 (95% CI $-1.04, 0.04$). However, in 25% of cases, the difference between the two SFs exceeded 5 points (0.5 SD; range 12–25 points). Among those for whom the two fatigue SF scores differed by at least 5 points (i.e., discrepant

scores), in 64% of individuals, the 7a score was lower than the 8a. Individuals with discrepant scores did not differ from those with similar scores by any sociodemographic or RA characteristic examined. The mean standard error was higher in the 7a than 8a (2.8 vs. 2.1, respectively). Notably, self-assessments of disease activity were also similar among clinic patients and significantly higher ($p=.000$; reflecting worse RA) in the online community.

Almost all (99%) respondents rated fatigue as an important symptom they considered when deciding how well their current treatment was controlling their RA (Table 2). The majority of participants (70–92%) reported that the 15 SF items in the 7a and 8a “completely” or “mostly” reflected their experience of fatigue. About 1 in 8 suggested that it would be useful to ask about other factors including sleep (“my fatigue causes me to sleep more”), the impact of fatigue on tasks requiring attention and memory (“brain fog”), and how fatigue affected work, social, and intimate

Table 2 Participant responses to questions about fatigue by source

	Arthritis clinic patients $N=84$	Online patient community $N=200$
How important is it to consider fatigue when deciding how well your RA treatment is working?		
Very important	69 (82%)	178 (89%)
May be important	12 (14%)	21 (11%)
Not important	3 (4%)	1 (<1%)
Do these questions generally reflect your experience of fatigue? ^a		
Completely	18 (35%)	88 (44%)
Mostly	18 (35%)	95 (48%)
Moderately	13 (25%)	13 (7%)
A little	2 (4%)	4 (2%)
Not at all	1 (2%)	0 (0%)
Are there other aspects of fatigue we should ask you about?		
Yes	11 (13%)	29 (15%)
How much of your fatigue is due to your RA? ^b		
All	3 (6%)	70 (35%)
A lot	14 (27%)	94 (47%)
Some	10 (19%)	18 (9%)
A little	8 (15%)	4 (2%)
None	11 (21%)	2 (1%)
Don't know	6 (12%)	13 (7%)
When you answered the questions, were you thinking about your...		
Fatigue in general	19 (23%)	15 (8%)
Fatigue from RA	36 (43%)	109 (55%)
Both RA fatigue and general fatigue	24 (29%)	64 (32%)
Unsure	5 (6%)	12 (6%)
If asked to rate only your fatigue due to RA, would you have answered differently? ^c		
Yes	9 (11%)	6 (3%)

^a $N=52$

^b $N=52$

^c $N=80$

relationships. One participant noted that the fear of becoming tired often caused them to limit activities.

Almost all ($\geq 94\%$) reported they were able to distinguish a general sense of fatigue from the fatigue they attributed directly to RA. Most (83%) of the online respondents (who reported significantly greater fatigue) as compared with 33% of clinic participants attributed “a lot” or “all” of their fatigue to their RA. When selecting a response, 72% of clinic patients and 87% of the online participants indicated that

they were specifically describing fatigue they attributed to their RA or a combination of RA and general fatigue. Further, when asked to rate only the fatigue they attributed to their RA, few (up to 11%) indicated they would have provided a different response. Correlations between patient self-assessments of disease activity and scores on the SFs were moderate (r values from 0.64 to 0.65; $p = .000$; Fig. 1).

Item-level debriefing

A diverse sample of participants was recruited from academic arthritis centers in Baltimore ($n = 12$), New York ($n = 12$), and Birmingham ($n = 8$) (Table 1). Interviewees also indicated that SF items were easily understood and no specific concerns were raised regarding question structure, stems, or recall period (data not shown).

Across items in both SFs, $\geq 85\%$ rated the individual questions as “somewhat” or “very relevant” to their fatigue experience (Table 3). In the Fatigue 7a, 25% rated one item “not at all” relevant (FATIMP21; “How often were you too tired to take a bath or shower?”). From 19 to 25% of participants rated three items that asked about being bothered, having to push oneself, and trouble finishing things due to fatigue as “not at all relevant” to their experience of fatigue.

To gain insight on fatigue attributions, we asked participants how they thought about their fatigue. On average, 59% of participants rated their average fatigue over the past 7 days, while 19% rated their fatigue at its worst. The

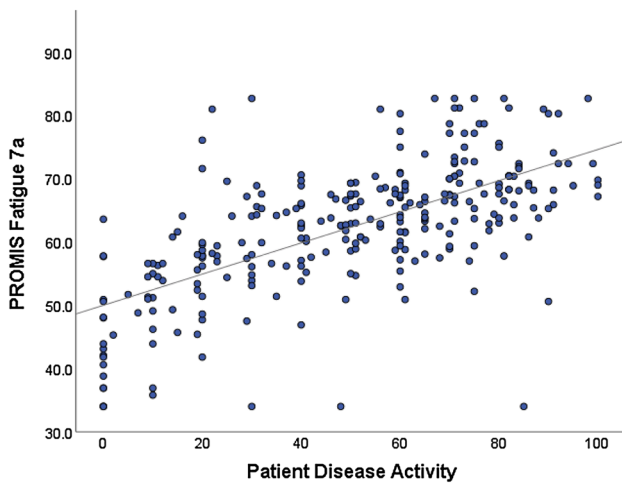


Fig. 1 Relationship between patient assessment of rheumatoid arthritis disease activity and PROMIS Fatigue 7a short form scores

Table 3 RA patient perceptions of relevance of PROMIS Fatigue short form items ($N = 32$)

Identifier	Question	Relevance		
		Very	Somewhat	Not at all
Fatigue 7a	How often...	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
FATEXP20	...did you feel tired	13 (81%)	2 (13%)	1 (6%)
FATEXP5	...did you experience extreme exhaustion	6 (38%)	8 (50%)	2 (13%)
FATEXP18	...did you run out of energy	10 (63%)	5 (31%)	1 (6%)
FATIMP33	...did your fatigue limit you at work (include work at home)	12 (75%)	4 (25%)	0
FATIMP30	...were you too tired to think clearly	8 (50%)	7 (44%)	1 (6%)
FATIMP21	...were you too tired to take a bath or shower	7 (44%)	5 (31%)	4 (25%)
FATIMP40	...did you have enough energy to exercise strenuously	9 (56%)	7 (44%)	0
	Proportion	58%	34%	8%
Fatigue 8a	During/In the past 7 days...			
HI7	...I feel fatigued	10 (63%)	4 (25%)	2 (13%)
AN3	...I have trouble starting things because I am tired	11 (69%)	3 (19%)	2 (13%)
FATEXP41	...how run down did you feel on average	11 (69%)	4 (25%)	1 (6%)
FATEXP40	...how fatigued were you on average	11 (69%)	4 (25%)	1 (6%)
FATEXP35	...how much were you bothered by your fatigue on average	9 (56%)	5 (31%)	2 (13%)
FATIMP49	...to what degree did fatigue interfere with your physical functioning	13 (81%)	0	3 (19%)
FATIMP3	...how often did you have to push yourself to get things done because of your fatigue	9 (56%)	4 (25%)	3 (19%)
FATIMP16	...how often did you have trouble finishing things because of your fatigue	9 (56%)	3 (19%)	4 (25%)
	Proportion	65%	21%	14%

remainder were not sure how they thought about their fatigue (13%) or used other heuristics (9%). Most (72%) reported they considered how their fatigue interfered with day-to-day life; 19% considered the intensity/severity of the fatigue they were experiencing, and one person (3%) said they considered both impact and severity. The mean (SD) scores of participants who anchored their ratings on fatigue impact/interference were significantly lower than those anchoring on severity/intensity or both [46.8 (9.2), 57.4 (9.0), 60.7 respectively; $p = .01$].

Discussion

A growing body of evidence suggests that fatigue in RA is a prevalent and debilitating symptom of RA which significantly impacts social and work participation, well-being, and quality of life. As definitions of fatigue vary, it remains unclear how to precisely and reliably measure this symptom in RA [23–26]. To our knowledge, this study is the first to assemble diverse samples and compare the responses of individuals recruited from an online community with those from specialty clinics to assess the relevance and representativeness of items in the two PROMIS Fatigue SFs currently available. Individual interviews were conducted in a separate sample to gain greater insight into how patients conceptualize and report on their experience of fatigue. Our results suggest that the contents of both the PROMIS Fatigue 7a and 8a SFs are relevant and representative of the full range of fatigue that people with RA experience. Almost all (87%) participants indicated that each PROMIS Fatigue SF asked about relevant aspects of fatigue. A similar proportion (88%) indicated that either version can capture the full spectrum of fatigue associated with RA. In one-on-one interviews, $\geq 75\%$ judged the items as “very” or “somewhat” relevant to their experience of fatigue.

Most participants indicated they could distinguish between fatigue they attributed primarily to their RA from a general tiredness resulting from other causes such as interrupted sleep or tending to young children. Among those with more active RA, fatigue was worse, and participants were more likely to attribute their fatigue directly to their RA. The attribution of symptoms such as fatigue is influenced by sociodemographic and psychological factors and disease knowledge [27], and often varies between patients and providers [28, 29]. In turn, symptom attributions influence coping, medication concerns, adherence to treatment, treatment response, and reporting of side effects to treating physicians [29, 30]. RA patients often attribute symptoms such as fatigue to less serious and non-modifiable causes, especially in the absence of joint swelling, and in turn are less likely to seek medical attention [27, 31]. In a recent study in the Netherlands, in an open-label transition to a biosimilar,

one quarter of patients who voluntarily switched asked to return to the originator mainly due to subjective experiences, including fatigue, that they attributed to the new drug [32].

When assessing the fatigue, most of the participants indicated that they conceptualized it in terms of its impact rather than severity suggesting it is the interference in day-to-day life more than the symptomatic experience that may be most salient and disabling. Interestingly, those who rated impact vs. severity also had significantly lower mean fatigue scores. The two Fatigue SFs that are currently available contain non-overlapping items querying both severity and impact to produce a single score [13, 33].

Although fatigue has been recommended as a core outcome measure in RA trials [34], there has been little consensus on how to measure it. Measures commonly used that conceptualize fatigue as a unidimensional factor include the fatigue severity VAS, the 4 SF-36 [35] vitality items, and the 13-item Functional Assessment of Chronic Illness Therapy (FACIT) Fatigue Scale [36]. Cella et al. compared the psychometric properties of the SF-36 Vitality subscale and FACIT in > 600 RA patients and found robust FACIT was better able to discriminate across the range of fatigue [37]. When comparing the fatigue VAS with longer scales in nearly 8000 people with RA, Wolfe concluded the VAS showed similar or better responsiveness than longer scales [38]. While unidimensional scales are generally brief and easy to complete, a single-item fatigue VAS is less reliable than multi-item measures [26] and may be too general. Although concerns have been raised about conceptualizing fatigue and energy as a single dimension, as in the SF36 items, a recent report using sophisticated bi-factor modeling supports a factor structure of one general (vitality) and two group (energy and fatigue) factors [23]. Examples of scales that assess fatigue as multiple domains in RA include the Multidimensional Assessment of Fatigue (MAF) scale, the Bristol Rheumatoid Arthritis Fatigue Multidimensional Questionnaire (BRAFM-DQ) [39], and the BRAF-NRS which consists of 3 single-item scales that assess severity, impact, and coping ability. BRAF is a multidimensional RA-specific fatigue measure developed from patient focus groups thus having strong content validity [39]. Using item response theory analysis, Oude Voshaar and colleagues [25] compared the psychometric characteristics of the BRAF-NRS, the SF36, and the BRAFM-DQ and concluded that while all measured a common underlying domain of fatigue severity, they differed considerably in precision and targeting. Whereas the SF36 items provided optimal information in individuals with mild fatigue, the BRAFM-DQ offered precision among those with high levels of fatigue [25]. The three single-item BRAF-NRS scales were not recommended due to the restricted measurement range. In the general population, the PROMIS Fatigue SFs offer maximum information in scores from 45 to 75 reflecting none to very mild

feelings of tiredness (-.5 SD) to very severe and sustained exhaustion (+ 2.5 SD) [40]. This range is highly relevant to people with RA. Tables are available that link scores on PROMIS Fatigue with FACIT-Fatigue and SF36 Vitality at prosettastone.org.

The PROMIS family of measures were developed to reliably and precisely assess a broad range of health domains that directly impact quality of life across chronic conditions [10]. We previously reported evidence of construct validity and relevance of several CATs assessing symptoms and impacts that people with RA have identified as important to them [15]. Our results from the survey and cognitive debriefing interviews suggest that the Fatigue SF contains items that are easy to understand and relevant to people with RA. Scores were similar on both SFs, and the primary difference between them appears to be length. PROMIS developers suggest that the 8a is more precise, whereas the 7a optimizes measurement across the full range of the domain [40]. Our results support the use of both versions in people with RA. Notably, the Fatigue 7a was recently used to evaluate overall symptom burden and quality of life in patients with myelofibrosis treated with ruxolitinib [41] with favorable results about fatigue included on the US product label. Rigorous methodology and a best practices approach [42] has been used to translate the PROMIS Fatigue SFs into more than 20 languages, with additional efforts underway [43].

Strengths of this study include the use of qualitative and quantitative methods to examine the content validity of the SF items in a diverse sample of patients with a wide range of disease symptoms and levels of fatigue. Purposive sampling was used to ensure good representation across age, sex, disease duration, disability, education, and geographical residence. There are also limitations. We were unable to confirm the diagnosis of RA in online participants but our screening approach increases confidence that these individuals had inflammatory arthritis that required DMARD therapy. Only participant perceptions of disease activity were available for the online group. Most participants were female, white, and well educated. Individual who agreed to participate may not have similar fatigue experiences with other RA patient populations. We did not specifically ask if patients attributed RA-related fatigue to their disease and/or medications used to control inflammation and pain.

In summary, in a socio-demographically and geographically diverse sample of people with RA from across the United States, fatigue was a common and important concern that affected day-to-day function and quality of life. Our results suggest that PROMIS fatigue SFs are relevant and can measure across the continuum of fatigue experienced by people with RA. PROMIS Fatigue SFs generates a single summative score that can be easily interpreted and widely applied in clinical and research settings. These data contribute to growing evidence supporting the use of

PROMIS measures to reliably and precisely evaluate fatigue and other symptoms in people with RA in clinical trials and care settings.

Acknowledgements The authors would like to thank research teams at Johns Hopkins, Hospital for Special Surgery, and University of Alabama at Birmingham for their assistance (Michelle Jones, Bernadette Johnson, Kathleen Andersen, Jessica Ashley), and the rheumatologists and trainees who referred patients for participation in the studies. We gratefully acknowledge assistance from Global Healthy Living Foundation for access to an online arthritis patient community through CreakyJoints.org and technical support from David Curtis. We thank members of our external advisory group to this project, including our patient research partners Ms. Amye Leong and Ms. Anne Lyddiatt, in addition to Dr. James Witter and Dr. April Naegeli.

Disclosures Dr. Bartlett is an unpaid non-voting member of the RA Working Group of the C-Path PRO Consortium and is an unpaid member of the steering committees for PROMIS International, the OMERACT Drug Safety Group, the Scleroderma Patient-centered Interventions Network (SPIN), Cancer Care Ontario, and Canada-PRO. She has served as a consultant to Eli Lilly, Janssen, Pfizer, Merck, and UCB in the area of patient-reported outcomes. Dr. Bykerk has consulted for Amgen, BMS, Pfizer, Sanofi/Regeneron, Gilead, and UCB; she has participated in advisory boards for Amgen, BMS, Pfizer, Sanofi/Regeneron, Gilead. Amgen, BMS, UCB have funded studies that are being or were recently performed at HSS where Dr. Bykerk is or was an investigator or principle investigator or site principle investigator on these studies. Dr. Orbai received research funding to Johns Hopkins University from Celgene, Eli Lilly, Horizon, Janssen, Novartis, and Pfizer. Dr. Orbai received consulting fees from participating in advisory boards for Eli Lilly, Janssen, Novartis, Pfizer, and UCB. Dr. Bingham is an unpaid non-voting member of the RA Working Group of the C-Path PRO Consortium and is an unpaid member of the executive committee of OMERACT (Outcome Measures in Rheumatology) which has been funded by arm's length funding from 32 pharmaceutical companies and research organizations. He has served as a consultant to Eli Lilly, Janssen, Pfizer, and UCB in the area of patient-reported outcomes. All other authors declare that he/she has no conflict of interest.

Funding The primary research data included within this report were acquired as part of projects funded in part from the Patient-Centered Outcomes Research Institute (PCORI) through a PCORI Methods Award (SC14-1402-10818). This work was also supported by the National Institutes of Health (NIH) through the Rheumatic Diseases Resource-based Core Center (P30-AR053503 Cores A and D, and P30-AR070254, Cores A and B), a research agreement from the Critical Path Institute (C-Path) to Johns Hopkins, and the Camille Julia Morgan Arthritis Research and Education Fund. All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of PCORI, its Board of Governors or Methodology Committee, of the NIH or the National Institute of Arthritis Musculoskeletal and Skin Diseases (NIAMS), or of the PROMIS Health Organization (PHO).

Compliance with ethical standards

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

- Bartlett, S. J., Hewlett, S., Bingham 3rd, C. O., Woodworth, T. G., Alten, R., Pohl, C., et al. (2012). Identifying core domains to assess flare in rheumatoid arthritis: An OMERACT International patient and provider combined Delphi consensus. *Annals of the Rheumatic Diseases*, *71*(11), 1855–1860.
- Hewlett, S., Cockshott, Z., Byron, M., Kitchen, K., Tipler, S., Pope, D., et al. (2005). Patients' perceptions of fatigue in rheumatoid arthritis: Overwhelming, uncontrollable, ignored. *Arthritis and Rheumatism*, *53*(5), 697–702.
- Katz, P. (2017). Fatigue in rheumatoid arthritis. *Current Rheumatology Reports*, *19*(5), 25.
- Kirwan, J. R., & Hewlett, S. (2007). Patient perspective: Reasons and methods for measuring fatigue in rheumatoid arthritis. *The Journal of Rheumatology*, *34*(5), 1171–1173.
- Kvien, T. K. (2004). Epidemiology and burden of illness of rheumatoid arthritis. *Pharmacoeconomics*, *22*(2), 1–12.
- Alten, R., Pohl, C., Choy, E. H., Christensen, R., Furst, D. E., Hewlett, S. E., et al. (2011). Developing a construct to evaluate flares in rheumatoid arthritis: A conceptual report of the OMERACT RA Flare Definition Working Group. *The Journal of Rheumatology*, *38*(8), 1745–1750.
- Kirwan, J. R., Minnock, P., Adebajo, A., Bresnihan, B., Choy, E., de Wit, M., et al. (2007). Patient perspective: Fatigue as a recommended patient centered outcome measure in rheumatoid arthritis. *The Journal of Rheumatology*, *34*(5), 1174–1177.
- Aletaha, D., Landewe, R., Karonitsch, T., Bathon, J., Boers, M., Bombardier, C., et al. (2008). Reporting disease activity in clinical trials of patients with rheumatoid arthritis: EULAR/ACR collaborative recommendations. *Arthritis Care & Research*, *59*(10), 1371–1377.
- Felson, D. T., Smolen, J. S., Wells, G., Zhang, B., van Tuyl, L. H., Funovits, J., et al. (2011). American College of Rheumatology/European League against rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Arthritis & Rheumatology*, *63*(3), 573–586.
- Khanna, D., Krishnan, E., Dewitt, E. M., Khanna, P. P., Spiegel, B., & Hays, R. D. (2011). Patient-Reported Outcomes Measurement Information System (PROMIS(R))—the future of measuring patient reported outcomes in rheumatology. *Arthritis Care Research (Hoboken)*, *63*(S11), S486–S490.
- Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., et al. (2010). The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology*, *63*(11), 1179–1194.
- Christodoulou, C., Junghaenel, D. U., DeWalt, D. A., Rothrock, N., & Stone, A. A. (2008). Cognitive interviewing in the evaluation of fatigue items: Results from the Patient-Reported Outcomes Measurement Information System (PROMIS). *Quality of life research: An international journal of quality of life aspects of treatment. Care and Rehabilitation*, *17*(10), 1239–1246.
- Lai, J. S., Cella, D., Choi, S., Junghaenel, D. U., Christodoulou, C., Gershon, R., et al. (2011). How item banks and their application can influence measurement practice in rehabilitation medicine: A PROMIS fatigue item bank example. *Archives of Physical Medicine and Rehabilitation*, *92*(10 Suppl), S20–S27.
- Cella, D., Lai, J. S., Jensen, S. E., Christodoulou, C., Junghaenel, D. U., Reeve, B. B., et al. (2016). PROMIS fatigue item bank had clinical validity across diverse chronic conditions. *Journal of Clinical Epidemiology*, *73*, 128–134.
- Bartlett, S. J., Orbai, A. M., Duncan, T., DeLeon, E., Ruffing, V., Clegg-Smith, K., et al. (2015). Reliability and validity of selected PROMIS measures in people with rheumatoid arthritis. *PLoS ONE*, *10*(9), e0138543.
- Bingham 3rd, C. O., Noonan, V., Auger, C., Feldman, D., Ahmed, S., & Bartlett, S. J. (2017). Montreal Accord on Patient-Reported Outcomes use series—paper 4: Patient Reported Outcomes (PRO) can inform clinical decision-making in chronic care. *Journal of Clinical Epidemiology*, *89*, 136–141.
- Bingham 3rd, C. O., Bartlett, S. J., Merkel, P. A., Mielenz, T. J., Pilkonis, P. A., Edmondson, L., et al. (2016). Using Patient-Reported Outcomes and PROMIS in research and clinical applications: Experiences from the PCORI pilot projects. *Quality of Life Research*, *25*(8), 2109–2116.
- Food and Drug Administration. (2009). Guidance for Industry—Patient-Reported Outcome measures: Use in medical product development to support labeling claims, Washington, DC: Department of Health and Human Services, Retrieved August 1, 2017, from <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>.
- Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., et al. (2011). Content validity—establishing and reporting the evidence in newly developed Patient-Reported Outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 2—assessing respondent understanding. *Value in Health*, *14*(8), 978–988.
- Karlson, E. W., Sanchez-Guerrero, J., Wright, E. A., Lew, R. A., Daltroy, L. H., Katz, J. N., et al. (1995). A connective tissue disease screening questionnaire for population studies. *Annals of Epidemiology*, *5*(4), 297–302.
- Aletaha, D., Neogi, T., Silman, A. J., Funovits, J., Felson, D. T., & Bingham 3rd, C. O. (2010). 2010 rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis & Rheumatology*, *62*(9), 2569–2581.
- Taylor, W., Gladman, D., Helliwell, P., Marchesoni, A., Mease, P., Mielants, H., et al. (2006). Classification criteria for psoriatic arthritis: Development of new criteria from a large international study. *Arthritis & Rheumatology*, *54*(8), 2665–2673.
- Deng, N., Guyer, R., & Ware, J. E. (2015). Energy, fatigue, or both? A bifactor modeling approach to the conceptualization and measurement of vitality. *Quality of Life Research*, *24*(1), 81–93.
- Hewlett, S., Dures, E., & Almeida, C. (2011). Measures of fatigue: Bristol rheumatoid arthritis fatigue multi-dimensional questionnaire (BRAFMQ), Bristol rheumatoid arthritis fatigue numerical rating scales (BRAFNRS) for severity, effect, and coping, Chalder fatigue questionnaire (CFQ), checklist individual strength (CIS20R and CIS8R), fatigue severity scale (FSS), functional assessment chronic illness therapy (fatigue) (FACIT-F), multi-dimensional assessment of fatigue (MAF), multi-dimensional fatigue inventory (MFI), pediatric quality of life (PedsQL) multi-dimensional fatigue scale, profile of fatigue (PROF), short form 36 vitality subscale (SF-36 VT), and visual analog scales (VAS). *Arthritis Care & Research*, *63* (Suppl 11), S263–S286.
- Oude Voshaar, M. A., Ten Klooster, P. M., Bode, C., Vonkeman, H. E., Glas, C. A., Jansen, T., et al. (2015). Assessment of fatigue in rheumatoid arthritis: A psychometric comparison of single-item, multi-item, and multidimensional measures. *Journal of Rheumatology*, *42*(3), 413–420.
- Wolfe, F. (2004). Fatigue assessments in rheumatoid arthritis: Comparative performance of visual analog scales and longer fatigue questionnaires in 7760 patients. *Journal of Rheumatology*, *31*(10), 1896–1902.
- Simons, G., Belcher, J., Morton, C., Kumar, K., Falahee, M., Mallen, C. D., et al. (2017). Symptom recognition and perceived urgency of help-seeking for rheumatoid arthritis and other diseases in the general public: A mixed method approach. *Arthritis Care & Research*, *69*(5), 633–641.

28. Shiyabola, O. O., & Farris, K. B. (2010). Variation in patients' and pharmacists' attribution of symptoms and the relationship to patients' concern beliefs in medications. *Research in Social and Administrative Pharmacy*, 6(4), 334–344.
29. Nijrolder, I., Leone, S. S., & van der Horst, H. E. (2015). Explaining fatigue: An examination of patient causal attributions and their (in)congruence with family doctors' initial causal attributions. *European Journal of General Practice*, 21(3), 164–169.
30. Horne, R., & Weinman, J. (1999). Patients' beliefs about prescribed medicines and their role in adherence to treatment in chronic physical illness. *Journal of Psychosomatic Research*, 47(6), 555–567.
31. Stack, R. J., Shaw, K., Mallen, C., Herron-Marx, S., Horne, R., & Raza, K. (2012). Delays in help seeking at the onset of the symptoms of rheumatoid arthritis: A systematic synthesis of qualitative literature. *Annals of the Rheumatic Diseases*, 71(4), 493–497.
32. Tweehuysen, L., van den Bemt, B. J. F., van Ingen, I. L., de Jong, A. J. L., van der Laan, W. H., van den Hoogen, F. H. J., et al. (2018). Subjective complaints as the main reason for biosimilar discontinuation after open-label transition from reference infliximab to Biosimilar Infliximab. *Arthritis & Rheumatology*, 70(1), 60–68.
33. Rose, M., Bjorner, J. B., Gandek, B., Bruce, B., Fries, J. F., & Ware, J. E. (2014). Jr. The PROMIS physical function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. *Journal of Clinical Epidemiology*, 67(5), 516–526.
34. Kirwan, J. R., Minnock, P., Adebajo, A., Bresnihan, B., Choy, E., de Wit, M., et al. (2007). Patient perspective: Fatigue as a recommended patient centered outcome measure in rheumatoid arthritis. *Journal of Rheumatology*, 34(5), 1174–1177.
35. Ware, J. E. Jr., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, 30(6), 473–483.
36. Webster, K., Cella, D., & Yost, K. (2003). The Functional Assessment of Chronic Illness Therapy (FACIT) measurement system: Properties, applications, and interpretation. *Health and Quality of Life Outcomes*, 1, 79–79.
37. Cella, D., Yount, S., Sorensen, M., Chartash, E., Sengupta, N., & Grober, J. (2005). Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. *The Journal of Rheumatology*, 32(5), 811–819.
38. Allaire, S., Wolfe, F., Niu, J., & Lavalley, M. P. (2008). Contemporary prevalence and incidence of work disability associated with rheumatoid arthritis in the US. *Arthritis Care & Research*, 59(4), 474–480.
39. Nicklin, J., Cramp, F., Kirwan, J., Greenwood, R., Urban, M., & Hewlett, S. (2010). Measuring fatigue in rheumatoid arthritis: A cross-sectional study to evaluate the Bristol rheumatoid arthritis fatigue multi-dimensional questionnaire, visual analog scales, and numerical rating scales. *Arthritis Care & Research*, 62(11), 1559–1568.
40. PROMIS Health Organization (2015) A brief guide to the PROMIS fatigue instruments: Health Measures, Updated 9/9/2015; cited 2018 from, <https://www.assessmentcenter.net/documents/PROMIS%20Fatigue%20Scoring%20Manual.pdf>.
41. Mesa, R. A., Gotlib, J., Gupta, V., Catalano, J. V., Deininger, M. W., Shields, A. L., et al. (2013). Effect of ruxolitinib therapy on myelofibrosis-related symptoms and other Patient-Reported Outcomes in COMFORT-I: A randomized, double-blind, placebo-controlled trial. *Journal of Clinical Oncology*, 31(10), 1285–1292.
42. Wild, D., Eremenco, S., Mear, I., Martin, M., Houchin, C., Gawlicki, M., et al. (2009). Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: The ISPOR Patient-Reported Outcomes translation and linguistic validation good research practices task force report. *Value Health*, 12(4), 430–440.
43. Alonso, J., Bartlett, S. J., Rose, M., Aaronson, N. K., Chaplin, J. E., Efficace, F., et al. (2013). The case for an international Patient-Reported Outcomes Measurement Information System (PROMIS®) initiative. *Health and Quality of Life Outcomes*, 11, 210.