Clinical Outcome Assessments (COA) Qualification Program
DDT COA #000015: PROMIS® Short Form Fatigue 10a in Rheumatoid Arthritis
August 16, 2017 Update

Section 1: Proposed Plan for COA Qualification

1.1 Introduction and overview

This should include a concise description of the disease and the clinical trial setting in which the COA would be used, the limitations of existing assessments, a brief description of the existing or planned COA, and the rationale for use in drug development.

Rheumatoid Arthritis (RA) is the most common form of inflammatory arthritis and is associated with fluctuating debilitating symptoms that confer considerable decrements to patients' longevity and quality of life.[1-5] In addition to symptoms such as pain, impaired physical function, stiffness, sleep disturbance, and emotional distress,[6-14] patient and clinician input reveals that fatigue has been identified as a common, persistent, disabling, and high-priority symptom in RA.[15-21] Findings suggest that after pain, fatigue represents one of the most important aspects of RA,[8,9,11,13,16,22,23] and a top priority for RA patients seeking treatment.[4,10] Research examining the fatigue experience of RA patients suggests that RA-associated fatigue differs from "normal" fatigue, impacts multiple domains of patients' lives, and is under-recognized by clinicians.[9,24,25] RA research typically characterizes patients' fatigue in terms of experience (e.g., severity, frequency) and impact (e.g., social functioning, physical function). These findings suggest that RA patients not only consider fatigue as one of the most important aspects of their disease experience, but as an important outcome when evaluating the effectiveness of interventions. Multiple factors have been proposed as contributing to fatigue in RA, including disease-related (e.g., inflammatory pathways), cognitive-behavioral (e.g., depression, sleep behavior), and personal (e.g., social activities and support).[26-29] Findings also indicate the heterogeneous experience of fatigue in RA, and its persistence even when traditional indicators suggest well-controlled RA activity.[16]

In response to the growing body of evidence regarding the centrality of fatigue among RA symptoms, several organizations have called for the broader inclusion of measures of fatigue in RA clinical trials. Currently, the American College of Rheumatology (ACR) core set of recommended outcome measures for use in RA clinical trials does not include a measure of fatigue (though it does include patient-reported assessments of physical function, pain, and global assessments of disease activity). Several groups have called for the addition of a fatigue measure to the ACR core set.[6,8,9,15,30,31] Moreover, in response to the increasingly feasible goal of achieving low disease activity and/or remission through intervention,[32] the ACR, European League Against Rheumatism (EULAR), and the Outcome Measures in Rheumatology (OMERACT) group recognized the importance of ensuring that remission is properly defined in RA.[33-36] Their efforts to redefine remission in RA were limited by the fact that only the domains of physical function, pain, and global assessment of disease activity were available in the ACR core set of measures used in clinical trials.[6] Input from patients and healthcare professionals emphasized the importance of patient's perspectives when studying RA remission to ensure

optimization of targeted therapy[20] and the inclusion of domains of importance in clinical trials, especially fatigue. Given that patient-reported outcome (PRO) measures serve as the best method for assessing symptoms that are only known to the patient, this input highlights the need to include fatigue PRO measures when evaluating RA remission.

To date, fatigue has not been assessed as often as physical function, pain, and disease activity in RA clinical trials and it has rarely been considered an independent treatment target. While this is consistent with the absence of a fatigue measure in the ACR core set (and perhaps a result of this absence),[6] it is inconsistent with the broader recognition of fatigue as an important RA symptom among patients, providers, and researchers. Given the importance patients place on fatigue and its resolution as part of remission, the persistence of fatigue despite well-controlled disease activity as defined by traditional indicators,[37] and the benefits of monitoring subtle symptom changes in the context of low disease activity/remission,[14,38,39] broader inclusion of fatigue measures is needed in RA clinical trials in order to better understand patients' responses to treatment. This includes determination of whether interventions can provide overall benefit to patients above and beyond the existing indicators for disease progression, symptom maintenance, and clinical remission. Precise and valid measurement of fatigue is required to fully evaluate the effects of RA interventions in clinical trials.

In 2012, the Patient-Reported Outcome (PRO) Consortium's Rheumatoid Arthritis (RA) Working Group held a workshop titled, "Toward Consensus Development: Qualifying Endpoint Measures for Rheumatoid Arthritis Clinical Trials."[15] The workshop's objective was to identify RA-related symptoms and RA-defining decrements in physical function that could be explored as potential PRO-based endpoints in clinical trials to support label claims for RA drugs. Along with the RA Working Group members and C-Path personnel, participants included RA patients, representatives from the US Food and Drug Administration (FDA), National Institute for Arthritis and Musculoskeletal and Skin Diseases (NIAMS), ACR, OMERACT, and EULAR. The conclusion of the workshop was that the greatest gap in terms of PRO measure qualification was demonstration of the incremental value of fatigue as a key symptom beyond the ACR criteria for assessing treatment benefit.

Several independent reports were subsequently produced, reviewing the available data concerning the measurement of patient-reported fatigue in RA.[15,37,40,41] These reports identified considerable limitations among several of the existing fatigue PRO measures for use in RA, including inadequate qualitative work with RA patients, variable quality in terms of psychometric properties, and a lack of validation in RA samples.[15,37,40,41] Additionally, there was insufficient evidence that frequently used fatigue PRO measures captured the full range of the fatigue experience in RA.[41]

These reports concluded that the *Patient-Reported Outcome Measurement Information System*® (*PROMIS*®**)** Fatigue metric was an appropriate candidate for inclusion in RA clinical trials, citing its broad coverage of fatigue concepts relevant to RA, the well-documented and rigorous methods used during its development, the precision through its use of item response theory (IRT)-based methods, and the considerable evidence of its reliability and validity in RA populations. **The ten items that comprise the *PROMIS*® *Short Form Fatigue 10a*[42] are propose d he re as the clinical outcome assessment (COA) tool to be qualifie d for use in the**

targeted concept of interest (i.e., fatigue severity assessed via patient-reported fatigue experience and impact) as a secondary efficacy endpoint measure in RA treatment trials.

Although the *PROMIS Fatigue Short Form 7a* was the measure originally suggested in the RA Working Group's Letter of Intent submitted on June 29, 2016, the Working Group's goal was to identify an optimized subset of items from the PROMIS Fatigue item bank to move forward for qualification. Further examination of the emerging literature led to the decision to focus on the *PROMIS Short Form Fatigue 10a*. The ***PROMIS Short Form Fatigue 10a*** is a fixed length short form derived from the PROMIS Fatigue Item Bank assessing fatigue experience and impact. It has a recall period of seven days and includes a 5-point verbal rating scale ranging from "Not at all" to "Very much." The process for selecting items for inclusion in the ***PROMIS Short Form Fatigue 10a*** is described in section 1.5. The ***PROMIS Short Form Fatigue 10a*** is comprised of 10 of the 13 items in the FACIT-Fatigue, which has been used extensively in RA research and RA clinical trials.[41] There is considerable published evidence from these studies to support their content validity, reliability, convergent validity, concurrent validity, predictive validity, responsiveness to ACR clinical classification, and minimally important differences (MID) in RA patients.[43] The ***PROMIS Short Form Fatigue 10a*** also enables the assessment of both fatigue experience and impact within a single brief measure, producing a score that locates the respondent on a unidimensional fatigue T score metric. This metric is linked to all the items in the PROMIS Fatigue Item Bank (and other PROMIS Fatigue short forms).

The PROMIS Fatigue Item Bank was developed as part of a broader initiative to develop and evaluate publicly available, efficient, and precise PRO measures for individuals across a wide variety of health conditions.[44] The development of the PROMIS Fatigue Item Bank utilized a rigorous, multi-step process involving comprehensive literature searches, patient focus groups, qualitative item review, and IRT analysis.[45,46] Given its intended use across a wide variety of patient populations, the item bank was developed to measure the full range of both fatigue experience and impact; these correspond to the sub-domains of fatigue which have emerged from qualitative research with RA patients.[42] In total, the PROMIS Fatigue Item Bank includes 38 fatigue experience and 57 fatigue impact (i.e., interference) items.[47] Although RA patients were among the patient groups included during the development of the PROMIS Fatigue Item Bank,[44] the items in the bank were not developed specifically for use in RA. Subsequent evaluation of the item content in the PROMIS Fatigue Item Bank in the context of fatigue concepts qualitatively elicited from RA patients suggests that the items in the ***PROMIS Short Form Fatigue 10a*** provide good coverage of the experience and impact of fatigue in RA.[42] Additional evidence (reviewed in Sections 2.2 – 2.4) supports the validity and reliability of the broader PROMIS Fatigue metric (and item bank) in RA[37,44,47] and its responsiveness to change in RA disease activity (assessed via patient perceptions and clinical indicators) over time.[37]

This Initial Briefing Package summarizes the evidence supporting the use of the ***PROMIS Short Form Fatigue 10a*** to address the need for a reliable, valid, and precise PRO measure of fatigue as a secondary efficacy endpoint in RA clinical trials.

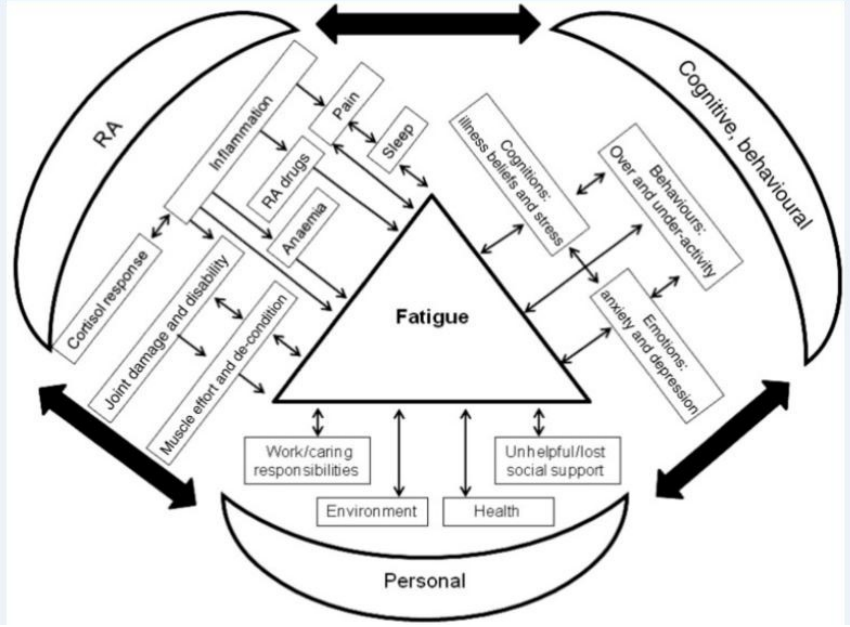
1.2 Concepts of interest for meaningful treatment benefit

Describe the meaningful aspect of patient experience that will represent the intended benefit of treatment (e.g., the specific symptom presence or severity or limitations in performance or daily activities relevant in the targeted COU)

The concept proposed as an indicator of treatment benefit in randomized clinical trials (RCTs) is the concept of fatigue severity among adults with RA. Fatigue is defined as "an overwhelming, debilitating, and sustained sense of exhaustion that decreases one's ability to carry out daily activities, including the ability to work effectively and to function at one's usual level in family or social roles."[45] Patients with chronic disease frequently identify fatigue as one of the key factors affecting their quality of life[48] and the experience of fatigue can confer decrements to multiple domains of quality of life, including physical, emotional, social, and cognitive well-being.[37] Further, the fatigue experienced by patients with chronic health conditions often differs from the experience of acute fatigue (such as that experienced after heavy work or exercise) in that it is not solely associated with overexertion and does not resolve following periods of rest.[49]

Fatigue represents a prominent and common symptom experienced by patients with RA, affecting an estimated 88-93% of patients.[49] The etiology of fatigue in RA is multifactorial, with inflammatory processes, pain, anemia, sleep quality, and psychosocial factors all playing potential roles.[27, 50-53] Hewlett and colleagues[27] have developed a multidimensional conceptual model of fatigue in RA (see Figure 1) that attempts to describe the disease-related factors (e.g., inflammation), cognitive-behavioral factors (e.g., anxiety, depression, stress, activity levels), and personal factors (e.g., work responsibilities, social support) that contribute to fatigue in RA.

Figure 1: Conceptual model of fatigue in RA[27]



Recent research on the etiology of fatigue in RA has focused particularly on associations with inflammatory processes. Evidence suggests positive relationships between fatigue and levels of the pro-inflammatory cytokines tumor necrosis factor (TNF), interleukin 1(IL-1), and interleukin 6 (IL-6),[54] as well as between fatigue and C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), and disease activity score (DAS28).[54] Further support for the relationship between inflammatory processes and RA fatigue is evidenced by findings demonstrating improvements in fatigue associated with disease-modifying therapies.[40,55-57]

Research on the relationship between fatigue and various indices of RA disease activity has been somewhat less consistent. For example, recent evidence suggests a dose-response relationship between fatigue and levels of the Clinical Disease Activity Index (CDAI), with significant differences between groups with high disease activity, low/moderate disease activity, and in remission.[58] This is consistent with significantly higher fatigue scores among RA patients experiencing an inflammatory flare as compared to those not in flare.[37] However, other evidence illustrating the high dispersion of fatigue scores within CDAI disease activity levels suggests the experience of fatigue is quite heterogeneous in RA, with a substantial number of patients reporting high levels of fatigue even when CDAI disease activity level appears well-controlled.[37] These findings are aligned with previous reports that, in spite of achieving clinical RA remission, many patients do not experience remission of their fatigue.[55]

There is also a growing body of research on the experience of fatigue from the RA patient perspective. In qualitative studies conducted with RA patients both in Europe and the United States, fatigue has consistently emerged as one of the most important symptoms to patients. Patients typically prioritize fatigue as secondary only to pain, though at least one study has reported that fatigue emerged as the highest priority for improvement[59] by RA patients. The importance of fatigue among RA patients is further illustrated by findings revealing that RA patients seeking pharmacologic intervention consider the elimination of fatigue as one of their top priorities.[4,10] When asked to consider the factors that would constitute remission, RA patients reported that, in addition to reduced pain and stiffness, fatigue would have to be either reduced or eliminated.[60] Further, when patients were queried about how much fatigue would need to improve in order to reflect their perception of remission, 23% indicated fatigue would need to be "less," 40% indicated it would need to be "almost gone," and 37% reported it would need to be "gone."[60] These findings suggest that RA patients consider fatigue one of the most important aspects of their disease experience, as well as an important outcome when evaluating the effectiveness of interventions.

In addition to highlighting the relative importance RA patients place on fatigue, recent research also elucidates unique aspects of the RA fatigue experience from the patient perspective. The difference between "normal" fatigue and RA-related fatigue emerges as a consistent theme across studies. RA patients differentiate between "tiredness" and the systemic fatigue they experience as part of RA[9,27,61] by describing the latter as overwhelming, difficult to resolve, and under-treated in clinical settings.[9]

A recent series of qualitative studies examined patients' perceptions of symptoms in the context of the fluctuating nature of RA. Patients indicated that when their symptoms were at their worst, their predominant symptom was pain, but as disease activity and pain improved, symptoms such

as fatigue were more apparent and had a more prominent impact on their daily functioning and well-being.[37] Similarly, when patients were asked to describe their experience as their disease transitioned from well-controlled to worsening, increased fatigue emerged as a prodromal symptom indicative of worsening disease activity[16] prior to the experience of other symptoms such as swelling, pain, and stiffness. RA patients also described experiencing persistent fatigue even when their pain and joint swelling were well-controlled.[16]

RA patients report that the experience of fatigue impairs their quality of life in many domains. These include their physical function, ability to participate in social roles and activities, emotional well-being, cognitive functioning, interpersonal relationships, ability to participate in rehabilitation treatment, and overall well-being.[27,61] Recent qualitative findings indicate that the negative impact of fatigue on physical function and ability to participate in activities is independent from the impact of pain and other RA symptoms.[37]

As the above findings suggest, the concepts of both *experience* and *impact* of fatigue are considered important when characterizing RA-related fatigue from the patient perspective. A recent report investigating the measurement of fatigue in RA further described sub-dimensions of fatigue experience and impact.[40] Sub-dimensions of fatigue experience in the RA literature include intensity, frequency, duration, variations in fatigue (e.g., unpredictable, irregular), and differentiation by cause.[40] Sub-dimensions of fatigue impact include impact/consequences, sleep/rest, requirements/problems, physical ability, cognition, emotions, and coping.[40] Taken as a whole, these findings suggest that both the experience and the impact of fatigue are important to include when assessing fatigue severity in RA intervention trials.

Identify targeted labeling or promotional claims based on the COA (i.e., proposed claim wording).

The following are examples of potential labeling claims for this measure:

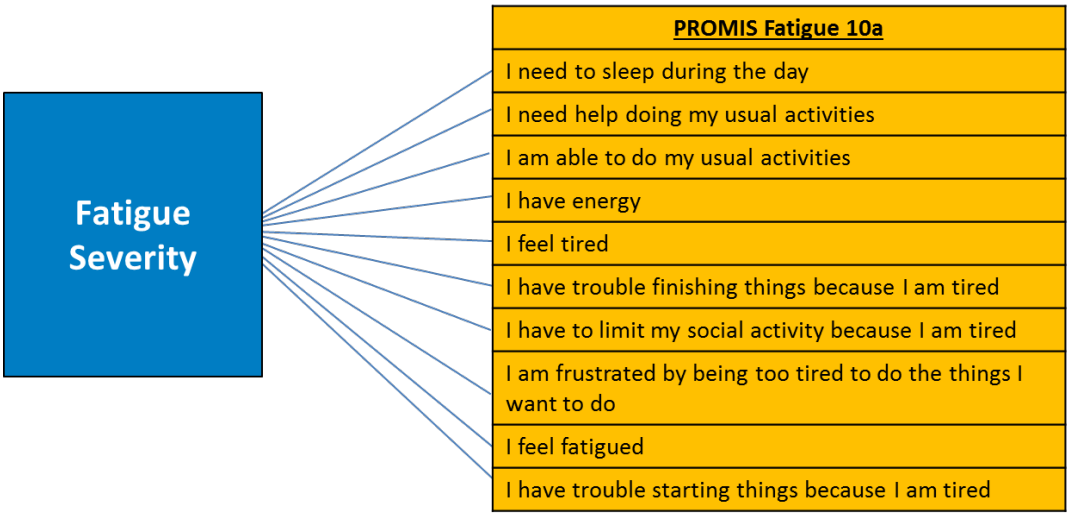
"A higher percentage of patients treated with [X] achieved meaningful reduction in fatigue severity than in the comparison group (placebo or alternative treatment)."

"The mean change (e.g., reduction) in fatigue severity from baseline was greater in patients treated with [X] than in the comparison group (placebo or alternative treatment)."

Provide a hypothesized conceptual framework for the outcome assessment(s).

The hypothesized conceptual framework for the assessment of fatigue in RA based on the ***PROMIS Short Form Fatigue 10a*** is shown in Figure 2 below.

Figure 2: Conceptual Framework



1.3 Context of use

Identify the targeted study population, including a definition of the disease and selection criteria for clinical trials (e.g., baseline symptom severity, patient demographics, language/culture groups)

The targeted study population is males and females 18 years of age and older with a definite diagnosis of RA based on a score of ≥ 6 on the American College of Rheumatology/European League Against Rheumatism 2010 Rheumatoid Arthritis Classification Criteria. This classification is based on the extent of joint involvement, serology, the results of acute-phase reactant tests, and the duration of symptom(s) (see Table 1 for more detail). It is anticipated that the largest proportion of patients in the targeted study population will be between 40 and 70 years old. The targeted study population will be without limitation regarding language, geography, or background/culture of the patient.

Identify the targeted study design. Most commonly the COA will be used to assess the change (compared to a control) induced by a medical treatment.

The targeted study design will be a longitudinal comparison of an experimental treatment to a control treatment (placebo control or active comparator) lasting a minimum of 12 weeks in length. We expect that the proposed fatigue outcome measure will typically be used to assess a secondary endpoint in randomized, double blind clinical trials to support the expected primary endpoint (e.g., the American College of Rheumatology [ACR] responder index,[62] the Disease

Activity Score for 28 joints [DAS28[63]]). In the anticipated study design, fatigue would be assessed along with other measures of RA symptoms and activity (e.g., those currently included in the ACR's recommended core set[6]) at baseline and then repeatedly, but no more often than weekly, while on study. Optional fatigue assessment at screening should be conducted when possible as this would allow for test-retest reporting. It is anticipated that fatigue assessment intervals would likely vary across trials; however, fatigue would be assessed concurrently with the other COAs in the trial.

Table 1: The 2010 American College of Rheumatology/European League Against Rheumatism classification criteria for Rheumatoid Arthritis

Criteria	Score
A. Joint Involvement	
1 large joint	0
2-10 large joints	1
1-3 small joints (with or without involvement of large joints)	2
4-10 small joints (with or without involvement of large joints)	3
>10 joints (at least 1 small joint)	5
B. Serology (at least 1 test result is needed for classification)	
Negative RF and negative ACPA	0
Low-positive RF or low-positive ACPA	2
High-positive RF or high-positive ACPA	3
C. Acute-phase reactants (at least 1 test result is needed for classification)	
Normal CRP and normal ESR	0
Abnormal CRP or abnormal ESR	1
D. Duration of symptoms	
< 6 weeks	0
≥ 6 weeks	1
Patient scores are derived by adding their scores from categories A-D. A score of 6 or more is needed for classification of a patient as having definite RA.	

Note: This table is for illustrative purposes only.

Identify the targeted study objectives and endpoint positioning (i.e., planned set of primary and secondary endpoints with hierarchy). Usually, the COA will serve as a primary or secondary study endpoint.

It is anticipated that the endpoint model would have the ACR Responder Index as the primary endpoint. The fatigue endpoint would be analyzed as a key secondary endpoint, to appropriately control for multiplicity as needed, aid in the interpretation of trial results in which RA disease activity is improved across treatment arms, with differential effects on fatigue. Thus, the fatigue endpoint can be used to interpret trial results by comparing the extent to which fatigue is affected by treatment and whether treatments are associated with benefits to patients in terms of fatigue severity. It is also possible that worsening fatigue could be an outcome of interest in tapering studies and those oriented around flare.

1.4 Critical details of the measure

Reporter, if applicable

PROMIS Short Form Fatigue 10a is a self-report assessment tool. Individual patients rate themselves on statements ("items") about fatigue. Patients choose the response option that most accurately describes their experience of fatigue and its impact.

Item Content or Description of the Measure

Adapting the World Health Organization's (2007) tripartite framework of physical, mental, and social health, PROMIS researchers developed multiple item banks for each domain,[44] including one for fatigue.[64-66] The WHO's International Classification of Functioning, Disability and Health included the minimization of fatigue among its stated aims,[67] highlighting the importance of regular assessment of fatigue in both research and clinical contexts. To this end, the PROMIS investigators used a multistep, mixed-methods approach to develop a fatigue item bank which can be used as an assessment tool as either a computerized adaptive test (CAT) or a fixed-length short form.[64-66] Given that the 10 items contained within the ***PROMIS Short Form Fatigue 10a*** are drawn from the larger 95-item calibrated PROMIS Fatigue Item Bank, a description of the item bank is provided here as background for description of the short form. It should be noted that several additional short forms (e.g., the PROMIS Fatigue 7a and 8a) are also drawn from the larger calibrated item bank.

The PROMIS Fatigue Item Bank v1.0 comprises 95 items that assess a range of self-reported symptoms, from mild subjective feelings of tiredness to an overwhelming, debilitating, and sustained sense of exhaustion that likely decreases one's ability to execute daily activities and function normally in family or social roles. Fatigue is divided into the experience of fatigue (frequency, duration, and intensity) and the impact of fatigue on physical, mental, and social activities. The fatigue items and short forms are universal rather than disease-specific. All assess fatigue over the past seven days using 5-level verbal rating scales.

PROMIS Fatigue is scored using the T-score metric, centered on the US general population.[66,68] This means that all PROMIS Fatigue measures – including the ***PROMIS Short Form Fatigue***

10a – have a mean score of 50 and a standard deviation of 10. Thus, a T-score of 60 is one standard deviation higher (more fatigue) than the "average person" in the US. The theoretical range of scores for the ***PROMIS Short Form Fatigue 10a*** extends infinitely in both directions, but in practice the range is from 30-85 T units.[69] This covers 5.5 standard deviation units, quite a broad range, in the general population.

As mentioned above, the PROMIS Fatigue Item Bank was designed to allow users to administer the measure in a number of ways. One method of administration is CAT; all of the items in the item bank have been calibrated to a mathematical model based on IRT, allowing for concise and reliable assessment across the full range of fatigue with just a few items. With a CAT, participant responses guide the system's choice of subsequent items from the full 95-item bank. Although items differ across respondents taking a CAT, scores are comparable across participants.

Fixed-length short forms have also been developed for PROMIS-Fatigue. The ***PROMIS Short Form Fatigue 10a*** is one of several available short forms made from the PROMIS Fatigue Item Bank. The content of the items in the ***PROMIS Short Form Fatigue 10a*** is also listed in Figure 2.

Mode of administration and data collection method

All of the items in the ***PROMIS Short Form Fatigue 10a*** and the PROMIS Fatigue Item Bank are intended to be self-administered (i.e., they do not require an interviewer).

Several modes of data collection are possible using the ***PROMIS Short Form Fatigue 10a***, including paper-and-pencil; electronic data collection via tablet computer, smartphone, or similar device; and data capture via telephone using interactive voice response. With regards to the evidence regarding differences across modes, Bjorner et al.[70] compared four methods of data collection using two non-overlapping parallel forms that included items from the PROMIS Fatigue Item Bank (along with additional PROMIS items relating to Physical Function and Depression). The items were administered to 923 adults with chronic obstructive pulmonary disease (COPD), depression, or RA. The four methods included 1) paper-and-pencil, 2) automated data collection by telephone using interactive voice response, 3) computer connected to the internet, and 4) personal digital assistant (PDA). In difference score analyses, no significant method differences were found, and all confidence intervals were within the pre-specified minimally important difference threshold of 0.2 SD. Parallel forms reliabilities were very high: intraclass correlation coefficient (ICC) range = 0.85-0.93. Only one across-method ICC (between interactive voice response and computer administration) was significantly lower than the same-method ICC. Tests of validity showed no differential effect by method of data collection though participants preferred screen interface over the other methods.

1.5 Overview of current COA development status

The items of the ***PROMIS Short Form Fatigue 10a*** were initially developed as part of the 13-item Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue) questionnaire for use in the assessment of fatigue associated with anemia.[71] In addition to subsequent validation for use in RA populations and use as a secondary endpoint in several RA clinical trials,[71-76] these 13 items have also been evaluated for content validity using semi-structured interviews with patients who have moderately to highly active RA.[42] During this evaluation, three of the original 13 items were identified as containing content of low to moderate relevance during the concept elicitation interviews, and were removed from the FACIT-Fatigue to create the ***PROMIS Short Form Fatigue 10a***. All of the items in the ***PROMIS Short Form Fatigue 10a*** were deemed to have high content validity and these 10 items were further shown to cover all of the fatigue-related concerns identified in the RA population.

In 2005, the 13 items in the FACIT-Fatigue were included among the 129 items of the initial PROMIS Fatigue item pool tapping two conceptual areas of fatigue: an individual's fatigue experience and the impact of fatigue on an individual's daily living.[64] All 13 items were among the 95 items retained in the final item bank after following the instrument development and validation standards described in Section 2.1. The PROMIS Fatigue Item Bank was one of several domain-based PRO measures developed as part of the PROMIS initiative. This initiative began in late 2004 when scientists from the National Institutes of Health (NIH) and several academic institutions formed a cooperative group that was funded under the NIH Roadmap for Medical Research Initiative. The primary goal of this group was to develop a set of publicly-available PRO measures that provide efficient and flexible tools in a wide range of health domains, including physical (e.g., fatigue, pain, physical function), social (e.g., social satisfaction, ability to participate in social roles) and mental health (e.g., depression, anxiety).

Initial development of the PROMIS Fatigue Item Bank concluded in 2008 and no subsequent revisions have been needed to date for any of the items, including those in the ***PROMIS Short Form Fatigue 10a***. Future updates will be undertaken as warranted based on novel research findings and previously unidentified needs in the assessment of fatigue (if any), though consideration of these updates will be offset by the need for consistency in the metric over time. All updates to PROMIS item banks are documented and labeled with version control procedures established by members of the PROMIS staff.

1.6 Description of the involvement of external expertise, including scientific communities or other international regulatory agencies, if applicable

The original PROMIS Fatigue Item Bank was developed by the PROMIS Fatigue working group, chaired by Dr. Arthur Stone from Stony Brook University (now at the University of Southern California). This group developed the domain framework protocol (Stone, Lai, Moul), and qualitative item review protocol (Stone, Yorkston, Moul, Guess). PROMIS investigators reviewed literature as well as existing instruments measuring fatigue, conducted binning exercises to enable the identification of redundant items, and winnowing exercises to reduce the large item pool down to a representative set of items.[44,69] All these results were reviewed and discussed by the PROMIS Steering Committee, composed of investigators from Duke University, Northwestern University, Stanford University, Stony Brook University, University of

North Carolina – Chapel Hill, University of Pittsburgh, University of Washington, and representatives from NIH.

The following external experts and agencies provided input and expertise for the development and validation of the PROMIS Fatigue Item Bank: The US Food and Drug Administration (through two meetings held at FDA in 2006-2008), OMERACT, the National Cancer Institute (through science officer participation at steering committee meetings and CaPS project meetings at NCI in 2006-2008), the World Health Organization and its International Classification of Functioning, Disability, and Health (coordinated through Bedirhan Ustun), and many consultants with considerable expertise in the measurement of fatigue and health.

Several individuals and organizations have provided input at various stages of this submission through membership in the PRO Consortium's RA Working Group or through non-member affiliation.

Member representatives of the RA Working Group include: Pam Berry, MSc (GlaxoSmithKline); Kate Burslem, MSc (Boehringer Ingelheim); Carol Gaich, PharmD, RPh (Eli Lilly and Company); Tristan Gloede, PhD (Boehringer Ingelheim); Kristina Harris, PhD (UCB Pharma); April Naegeli, DrPH, MPH (Co-Chair of the Working Group; Eli Lilly and Company); and Enkeleida Nikai, MSc, MB (Co-Chair of the Working Group; Eli Lilly and Company).

Non-member participants include: Susan J. Bartlett, PhD (McGill University); Clifton O. Bingham III, MD (Johns Hopkins University); David Cella, PhD (PROMIS, Northwestern University); David M. Condon, PhD, MBA (Northwestern University); San Keller, PhD (PROMIS, American Institutes for Research [AIR]); Amye Leong, MBA (Patient Representative); Lee S. Simon, MD (OMERACT); Vibeke Strand, MD (OMERACT); James Witter, MD, PhD (PROMIS, NIH/NIAMS).

Section 2: Summaries of Completed Studies

2.1 *Evidence of content validity*

Kaiser et al.[42] conducted content validity evaluation of the items in the PROMIS Fatigue 13a (aka the FACIT-Fatigue) in a sample of 16 patients with moderately to highly active RA (mean CDAI score for the sample = 22.68), and found strong support for both the coverage and relevance of 10 of the 13 items, resulting in the proposed ***PROMIS Short Form Fatigue 10a***. Support was low-to-moderate for the other 3 (of 13) items. These analyses were based on semi-structured interviews with each patient that began by inquiring about the importance, experience, and impact of fatigue in daily living (i.e., concept elicitation). Then, after completing the 13 items, patients were further interviewed about the extent to which the items captured the experience of RA-related fatigue and the extent to which each participant found the items relevant and comprehensible. Interviews were conducted until evidence of saturation was reached (i.e., when three consecutive interviews occurred without producing a new, relevant

concept). In addition to providing notable feedback from participants in response to the content in each of the items in the ***PROMIS Short Form Fatigue 10a,*** Kaiser et al.[42] concluded that all of the items related to aspects of fatigue that were relevant and important in the experience of fatigue among RA patients. Additional qualitative information from content debriefing of the items in PROMIS Short Forms Fatigue 7a and 8a will also be available for inclusion in the full qualification package.

Evidence for content validity of the broader PROMIS Fatigue Item Bank stems from both the methods used to develop the bank and the procedures used for validation. Developers of the PROMIS Fatigue Item Bank had several aims. These included the intention to develop a measure that (1) could be used to assess a wide range of fatigue across many disease states, (2) capitalized on the benefits of inclusion in the PROMIS measurement framework (particularly standardization and uniformity across measures), and (3) reflected cutting edge methods for instrument development and administration. Experts in the assessment of fatigue collaborated to achieve these aims by following the protocols for PROMIS measurement development.[44] These procedures (detailed below) included construct elicitation and definition, the identification of a large pool of candidate items, extensive cognitive interviewing,[65] binning and winnowing of the item pool in order to reduce it down to a more manageable item set, review and revision of the reduced item set, large-scale data collection based on data collection from a representative sample, and intensive psychometric analyses.[64]

Construct Definition and Item Identification

The team of experts responsible for developing the PROMIS Fatigue item bank began by conducting a literature review[64] aimed at describing the fatigue domain and identifying its subordinate concepts. While fatigue is a familiar experience for almost all people and is relevant to a wide variety of situations, the definition generated for the creation of a PROMIS measure is focused only on medically-relevant pathological fatigue.[65] Fatigue was defined as "an overwhelming, debilitating, and sustained sense of exhaustion that decreases one's ability to carry out daily activities, including the ability to work effectively and to function at one's usual level in family or social roles."[65] Based on this literature review, the team identified more than 80 fatigue questionnaires containing over 1,000 fatigue items that were at least partially related to the construct definition for fatigue.

Cognitive Interviewing

After paring down the items based on highly redundant content, the PROMIS team conducted extensive cognitive interviewing on the remaining 136 potential items.[65] The participant sample for the cognitive interviewing was designed to represent a diverse range of chronic health conditions (e.g., arthritis, pain, heart conditions) and participants were recruited from the North Carolina Musculoskeletal Health Project and the UNC General Internal Medicine Practice. Each participant ($N = 19$) responded to a series of open-ended questions about approximately one-quarter of the candidate items. Issues of concern for each item were rated as mild or serious by at least two coders and categorized using the QAS-99 coding system.[77] Seven items were eliminated based on the cognitive interviewing feedback, most often due to having lower ratings of clarity and/or applicability to respondents' lives.[65]

Binning and Winnowing

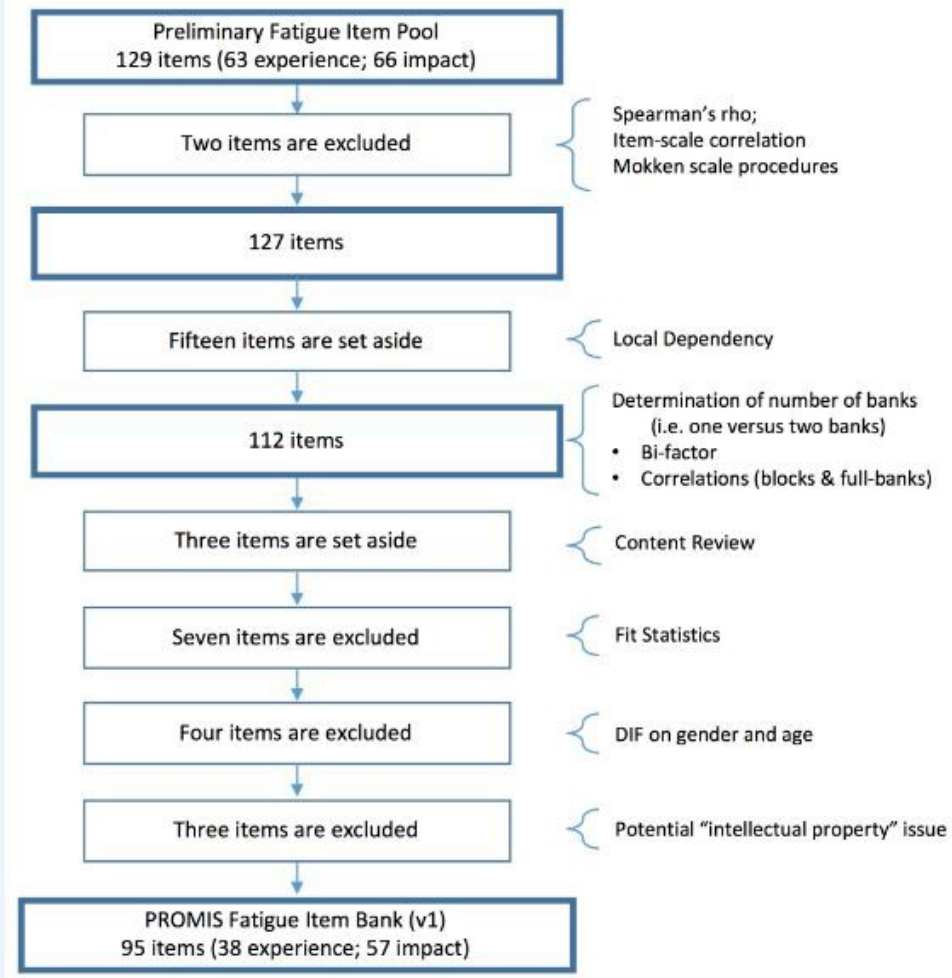
The remaining 129 fatigue items (including 17 items from two legacy measures – 4 items from the SF-36 Vitality Scale and 13 items from the FACIT-Fatigue) were then organized into two groups: 63 items relating to the experience (intensity, frequency, or duration) of fatigue; and 66 items relating to the impact of fatigue on physical function, emotional function or social function.[64]

Large-Scale Field Testing

The remaining candidate items were field tested by 21,133 participants.[44] The majority of the sample (93%; n = 19,601) was recruited by Polimetrix, a polling firm based in Palo Alto, California; the remainder was recruited by the Stanford PROMIS Research Site study cohort and the North Carolina PROMIS Research Site study cohort (n = 1,532). The Polimetrix sample was designed to reflect Year 2000 US Census demographics. Two-thirds of these participants (n = 13,250) were drawn from the general US population. The remaining Polimetrix participants were recruited from clinical samples of individuals with cancer (n = 1,754), COPD (n = 1,214), psychiatric illness (n = 1,193), heart disease (n = 1,156), osteoarthritis (n = 918), rheumatoid arthritis (n = 557), spinal cord injury (n = 531), and other conditions (n = 560). Overall, the field testing sample had a median age of 50 years, was 52% female, 82% white, and 97% had a high school education or more.

Data from the field-testing sample were used for a range of quantitative analyses[44,64] (see Figure 3). The "impact" and "experience" item sets were initially analyzed separately and then together to determine whether fatigue could be reported as a unidimensional construct using a single score.[64] In the first step, two items from the "impact" item set were removed due to low item-scale correlations. An additional five items in the "impact" set and 10 items in the "experience" set were subsequently removed due to evidence for local dependency (one item in each local dependency pair was retained). Bi-factor analyses of the remaining items indicated that all items loaded more highly on the general factor than their respective specific factors (impact or experience), suggesting unidimensionality in the full item set. This was supported by a correlation of 0.95 between the two sub-factors.[64]

Figure 3: Steps to develop the PROMIS Fatigue Item Bank [64]



At this stage, a post-hoc content review by the PROMIS Fatigue team led to the removal of three additional items and seven more were dropped based on low fit on the unidimensional factor. Analyses of differential item functioning (DIF) led to the exclusion of four items based on evidence of DIF across gender and age. Finally, three additional items were dropped due to potential intellectual property concerns. This left a total of 95 items in the item bank.[64]

These items were then calibrated and placed on the same metric using a graded response IRT model.[64] The resulting calibrations are used to score the general PROMIS Fatigue Item Bank and all PROMIS Fatigue short-forms, including the ***PROMIS Short Form Fatigue 10a***. IRT is a family of mathematical models that estimate unique properties for each item response category relative to the underlying (latent) dimension that is measured by each item.[78,79] These properties ("item parameters") are based on how likely people with different levels of the measured trait are to endorse each response option in an item. **The application of IRT permits users to administer any subset of items from a bank – including the items in any short form – and compare scores on a common metric**. Because the IRT assumptions for unidimensionality are stringent,[80,81] the items in PROMIS banks are highly inter-correlated, yielding a single dimension that explains the large majority of variance in person-level scores.

In sum, there is strong direct evidence for the content validity of the **PROMIS Short Form Fatigue 10a** among RA patients based on results from Kaiser et al.[42] There is also strong evidence for the validity of the broader PROMIS Fatigue metric in RA (and other disease states) based on the procedures and samples used during the development and validation of the PROMIS Fatigue Item Bank.

2.2 *Cross-sectional evaluation of measurement properties*

There is considerable evidence for evaluating the cross-sectional properties of the ***PROMIS Short Form Fatigue 10a*** and the larger PROMIS Fatigue Item Bank in RA populations. Evaluation of these properties is strengthened by the legacy status of the items in the ***PROMIS Short Form Fatigue 10a*** as part of the FACIT-Fatigue measure.

FACIT-Fatigue – which includes all 10 items in the ***PROMIS Short Form Fatigue 10a*** and is also known as the PROMIS Short Form v1.0 Fatigue 13a – was tested in a sample of 636 RA patients in the Safety Trial of Adalimumab in Rheumatoid Arthritis ("STAR").[68] The FACIT-Fatigue, the SF-36 Vitality scale, and the Multidimensional Assessment of Fatigue (MAF) were administered at baseline, week 12, and week 24 of the 24-week trial. Cronbach's coefficient alpha for internal consistency reliability for the ***PROMIS Short Form Fatigue 10a*** was 0.91-0.93 across all visits. Correlations among all three measures of fatigue were 0.69 – 0.86[68] providing strong evidence for concurrent validity. Table 2 demonstrates significantly different scores between known-groups of patients with different self-assessed levels of disease activity.

The Johns Hopkins Arthritis Center initiated a prospective cohort study of patients with RA in 2012. Data from the first 177 patients enrolled in the study are described in Bartlett et al.[21] In addition to other PROMIS measures and clinical instruments, the PROMIS Fatigue CAT was administered. The CAT was programmed to administer 4-8 items until a standard error of 0.3 or less for respondent's scores was detected. Cronbach's coefficient alpha, a measure of internal consistency reliability, was 0.99. Thirty-four participants completed a test-retest assessment approximately 2 days later. Among these 34 participants, the Pearson correlation coefficient for reliability was 0.88. Test-retest reliability will be explored further for the qualification package of the ***PROMIS Short Form Fatigue 10a***.

PROMIS Fatigue scores in the Johns Hopkins cohort were moderately correlated ($0.4 < r < 0.7$) with the other PROMIS domains measured (physical function, pain intensity, pain interference, sleep disturbance, sleep impairment, anxiety, depression, anger, ability to participate, and satisfaction with social roles and activities).[37] Further evidence of construct validity is indicated in Table 3; PROMIS Fatigue scores were also moderately correlated with CDAI scores ($r = 0.60$) and patient global assessment of disease activity ($r = 0.68$). PROMIS Fatigue correlated strongly ($r = 0.86$) with a visual analog scale (VAS) rating of fatigue. Table 4 demonstrates significantly different scores between known-groups of patients with different levels of disease activity. Patients who reported that they were experiencing an inflammatory flare also had significantly worse PROMIS Fatigue scores than those who were not in a flare ($p=.03$), as shown in Table 5.

Table 2: PROMIS Fatigue T-scores among groups with differing levels of patient global assessment of disease activity in the STAR trial at baseline[68]

	Patient Global Assessment of Disease Activity (0-100)		
	0-39 (n=168)	40-59 (n=199)	60-100 (n=258)
PROMIS Fatigue, Mean (SD)	53.3 (7.8)	57.6 (6.9)	61.0 (8.0)

Higher scores on the Patient Global Assessment represent higher levels of disease activity

Table 3: Correlations of PROMIS Fatigue T-scores with related clinician- or patient-reported measures in an RA cohort[37]

Related Measures	*r*
Clinical Disease Activity Index (CDAI)	0.60
Patient Global Assessment of Disease Activity	0.68
Fatigue VAS	0.86
Modified Health Assessment Questionnaire (mHAQ)	0.51
PROMIS Physical Function	-0.64
Pain Visual Analog Scale	0.64

Table 4: PROMIS Fatigue T-scores among groups with differing levels of disease activity based on the Chronic Disease Activity Index (CDAI)[37]

	CDAI Level			
	Remission (n=56)	Low (n=67)	Moderate (n=39)	High (n=14)
PROMIS Fatigue, Mean (SD)	46.2 (8.6)	55.7 (8.3)	58.5 (8.9)	64.0 (9.6)

Table 5: PROMIS Fatigue T-scores and "flare" activity in RA patients [37]

	Flare (n=25)	Not in flare (n=100)
PROMIS Fatigue, Mean (SD)	58 (7)	53 (11)

Note: "Flare" in RA defined as inflammatory activity significantly beyond usual baseline levels after taking into account usual day-to-day activity

2.3 *Longitudinal evaluation of measurement properties*

Responsiveness

In the STAR sample,[68] patients were classified into groups according to their ACR clinical response category at week 24: those who failed to achieve ACR20; those who achieved ACR20 as their best response; those who achieved ACR50; and those who achieved ACR70. Table 6 demonstrates that greater improvements in ***PROMIS Short Form Fatigue 10a*** scores were seen with greater levels of ACR response. Similarly, Table 7 demonstrates that changes in ***PROMIS Short Form Fatigue 10a*** scores from baseline to week 24 were consistent with the direction and magnitude of the changes in defined groups based on the SF-36 Vitality scale and the MAF.

*Table 6: Change in **PROMIS Short Form Fatigue 10a** scores by ACR response*

ACR group	N	Week 24 Mean Change	Effect Size	SRM
Did not achieve ACR20	290	-1.9	-0.23	-0.24
ACR20	144	-5.0	-0.61	-0.65
ACR50	68	-8.8	-1.07	-1.14
ACR70	56	-10.6	-1.29	-1.38

Effect size = mean change / pooled baseline SD
SRM = standardized response mean / pooled change score SD

*Table 7: Change in **PROMIS Short Form Fatigue 10a** scores by SF-36 Vitality and Multidimensional Assessment of Fatigue (MAF) mean scores*

	Change in ***PROMIS Short Form Fatigue 10a***				
	≤ -0.5 SD	-0.5 to -0.2 SD	-0.2 to 0.2 SD	0.2 to 0.5 SD	> 0.5 SD
SF-36 Vitality					
N	48	72	51	131	257
Mean (SD)	3.2 (5.5)	-0.2 (5.3)	-1.5 (4.3)	-3.2 (5.4)	-8.2 (8.0)
MAF					
N	250	80	105	58	62
Mean (SD)	-9.1 (7.6)	-3.1 (4.0)	-1.4 (4.4)	0.7 (4.5)	3.1 (5.7)

Higher scores on the SF-36 Vitality scale indicate higher levels of vitality.
A 10-point change in ***PROMIS Short Form Fatigue 10a*** scores is equal to 1 SD

Evidence for ability to detect change

In the Johns Hopkins Arthritis Center cohort,[21] responsiveness was assessed by examining the change from baseline to the first follow-up visit. Patients completed a health transition question: "Compared to your last visit, how would you rate your RA? Much better, a little better, the same, a little worse, or much worse?" As Table 8 indicates, patients who rated their RA as much worse had scores that were 4.7 points worse than baseline. Those who reported that their RA was much better had a mean score improvement of 3.7 points.

Table 8: Change in PROMIS Fatigue CAT scores by RA improvement

	Much better (n=22)	A little better (n=22)	Same (n=66)	A little worse (n=34)	Much worse (n=10)
PROMIS Fatigue, Mean	-3.7	-0.6	0.3	1.0	4.7

In the subset of patients with moderate to severe disease activity at baseline, patients were classified as improved, the same, or worsened based on shifts in CDAI category. Patients who improved by one or more CDAI categories had a mean PROMIS Fatigue improvement of 5.1 points, while those who worsened by one or more levels had mean change of 4.6 points (see Table 9).

Table 9: Change in PROMIS Fatigue CAT scores by Chronic Disease Activity Index (CDAI) category

	Improved (n=19)	Same (n=18)	Worse (n=8)
PROMIS Fatigue, Mean (SD)	-5.1 (10.4)	0.3 (6.2)	4.6 (5.7)
	p=0.033	*p=0.822.*	*p*=0.059

The responsiveness of the PROMIS Fatigue CAT was also evaluated in a longitudinal, observational study of 521 RA patients with no intervention.[47] The follow-up assessment occurred approximately 12 months after study enrollment. Change in general health was rated and patients were classified as better, about the same, or worse. Results are shown in Table 10. Patients who rated their general health as improved had a mean PROMIS Fatigue CAT score improvement (i.e., reduction) of 2.8 points, while those who reported their general health worsened had a mean increase of 2.6 points on the PROMIS Fatigue metric.

Table 10: Change in PROMIS Fatigue scores by general health

		N	Mean change (SD)	SRM
General global change	Better	61	-2.8 (5.6)	-0.50
	About the same	297	0.5 (5.9)	0.09
	Worse	92	2.6 (6.1)	0.43

SRM = standardized response mean = mean change / SD of change

2.4 *Longitudinal evaluation to provide guidelines for interpretation of trial results*

The estimated low end of an important difference (ID) for the FACIT-Fatigue was confirmed in the STAR study[68] as a 3-4 point change. Based on the starting point of that study and observed standard deviations, this would translate to a change score of 2-4 points on the PROMIS metric, including the ***PROMIS Short Form Fatigue 10a***. In a cancer study, a suggested ID for the PROMIS Fatigue has been proposed as a change of 2.5-4.5 points.[82] In addition, a bookmarking study is currently being conducted by Bingham and Bartlett to generate an RA-specific fatigue guideline that will enable evaluation of within-patient change for the PROMIS Fatigue metric. Preliminary evidence from that work suggests, as with several other PROMIS T score domains, a change of 5 points in an individual person may be a reasonable starting point for a responder definition. Further examinations of ID will be evaluated in future RA clinical trials.

2.5 *Language translation and cultural adaptation*

Process for translation/adaptation of original version

The items in the ***PROMIS Short Form Fatigue 10a*** have been translated into 59 languages. Many of the remaining items in the PROMIS Fatigue Item Bank have also been translated into many languages as well, though none of the available short forms have been as widely translated as the ***PROMIS Short Form Fatigue 10a***.

The PROMIS translation methodology is described in the Translation and Cultural Adaptation section of the PROMIS Instrument Development and Psychometric Evaluation Scientific Standards. It can be summarized as an iterative process of forward and back-translations, multiple item-level reviews, and cognitive interviews with a sample of native speakers of the target language. A universal approach to translation guides the translation process – that is, the goal is to produce, whenever possible, one language version suitable for the countries where the language is spoken, rather than country-specific versions of that language. This translation methodology is consistent with recognized consensus-derived best practice recommendations.[83,84]

In addition to adhering to the standard linguistic validation method of cognitive interviewing to ensure comprehension of the items and evaluate conceptual equivalence, most translations of the ***PROMIS Short Form Fatigue 10a*** items were also submitted to psychometric evaluation, to further assess measurement equivalence vis a vis the English source.[85]

Process for simultaneous development of versions in multiple language or cultures

The PROMIS Fatigue Item bank was developed first in English, then translated in its entirety (95 items) into Spanish. After that, multiple language translations were initiated. There have been instances when PROMIS Fatigue items (either all of the items in a specific fixed-length form or a subset of the items in the full item bank) have been translated into several languages at the same time for the same project. The procedures used for multiple concurrent translations are the same as that for single language translations. Each new translation is harmonized with the English source and all other languages already available or being translated at the same time. The

process consists of comparing the English back-translations for each item and, whenever possible, comparison with similar languages (for example, within romance languages, or within Nordic languages).

Evidence that content validity is similar for versions in multiple languages

Each of the available translations was linguistically validated based on administration and subsequent cognitive interviews with samples from various targeted language populations to ensure that the items are well understood and conceptually equivalent to the English source. This is the standard linguistic validation process, and it helps us to (1) assess the relevance of the items in the target language, (2) identify potentially problematic issues in the translated items, and (3) evaluate the content validity of the measures in the target language. We have not conducted concurrent qualitative assessments to identify, within the target languages, potential sub-concepts or constructs relevant to the target culture that may not be reflected in the English items. However, during the cognitive interviews, subjects have been asked if there was anything that should be added, anything missing or anything offensive. Nothing was identified.

The ***PROMIS Short Form Fatigue 10a*** items have been cognitively debriefed and tested with samples of 10 to 15 subjects in each language. Given that some of the items and/or concepts (e.g., "tired," "fatigued," "need to sleep during the day") overlap with other PROMIS, Neuro-QoL and FACIT measures, some of the items have been debriefed multiple times in the same language, thus accumulating qualitative data to confirm the appropriateness of the translated versions. In general, the results of cognitive interviews have demonstrated that the subjects understand the translated ***PROMIS Short Form Fatigue 10a*** items as they are intended and that the items are culturally appropriate. In most languages, subjects have reported no difficulty understanding the items and either found them relevant to their situation or described examples of situations when the items would apply. Based on subjects' feedback, wording revisions have been made where needed to increase comprehensibility and to ensure conceptual equivalence between the translated versions and the English source.

2.6 *User manual*

A user manual for the PROMIS Fatigue Item Bank and fixed-length short forms has been developed which outlines information relating to the qualitative and quantitative development and testing of the measures. Information on the administration procedures, methods and modes are outlined as well as study participant and investigator training processes. Scoring and interpretation procedures for the PROMIS Fatigue Item Bank are also included in the user manual to provide guidance to users of the PROMIS Fatigue Item Bank and fixed-length forms and to ensure consistent implementation in clinical studies. It will be updated to include the PROMIS Fatigue Short Form 10a as part of the qualification package submitted to FDA.

References:

[1]Hootman JM, Helmick CG. Projections of US prevalence of arthritis and associated activity limitations. *Arthritis and Rheumatism* 2006; 54(1):226–9. doi: 10.1002/art.21562 PMID:16385518.

[2]Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. *Arthritis and Rheumatism* 2008; 58(1):26–35. doi: 10.1002/art.23176 PMID: 18163497; PubMed Central PMCID: PMC3266664.

[3]Helmick CG, Felson DT, Lawrence RC, Gabriel S, Hirsch R, Kwoh CK et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. *Arthritis and Rheumatism* 2008; 58(1):15–25. doi: 10.1002/art.23177 PMID: 18163481.

[4]Sanderson T, Morris M, Calnan M, Richards P, Hewlett S. What outcomes from pharmacologic treatments are important to people with rheumatoid arthritis? Creating the basis of a patient core set. *Arthritis Care & Research (Hoboken)* 2010; 62(5):640–6. Epub 2010/05/13. doi: 10.1002/acr.20034 PMID: 20461785; PubMed Central PMCID: PMC2887082.

[5]Sanderson T, Kirwan J. Patient-reported outcomes for arthritis: time to focus on personal life impact measures? *Arthritis and Rheumatism* 2009; 61(1):1–3. Epub 2009/01/01. doi: 10.1002/art.24270 PMID: 19116964.

[6]Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. *Arthritis and Rheumatism* 1993; 36(6):729–40. doi: 10.1002/art.1780360601 PMID: 8507213

[7]Singh JA, Furst DE, Bharat A, Curtis JR, Kavanaugh AF, Kremer JM et al. 2012 Update of the 2008 American College of Rheumatology Recommendations for the Use of Disease-Modifying Antirheumatic Drugs and Biologic Agents in the Treatment of Rheumatoid Arthritis. *Arthritis Care & Research* 2012; 64(5):625–39. doi: 10.1002/acr.21641 PMCID: PMC4081542

[8]Kirwan JR, Minnock P, Adebajo A, Bresnihan B, Choy E, de Wit M et al. Patient perspective: fatigue as a recommended patient centered outcome measure in rheumatoid arthritis. *Journal of Rheumatology* 2007; 34(5):1174–7. PMID: 17477482

[9]Hewlett S, Cockshott Z, Byron M, Kitchen K, Tipler S, Pope D et al. Patients' perceptions of fatigue in rheumatoid arthritis: overwhelming, uncontrollable, ignored. *Arthritis and Rheumatism* 2005; 53(5):697–702 doi: 10.1002/art.21450 PMID: 16208668

[10]Sanderson T, Morris M, Calnan M, Richards P, Hewlett S. Patient perspective of measuring treatment efficacy: the rheumatoid arthritis patient priorities for pharmacologic interventions outcomes. *Arthritis Care & Research* 2010; 62(5):647–656. PMID: 20461786

[11]Bartlett SJ, Hewlett S, Bingham CO 3rd, Woodworth TG, Alten R, Pohl C et al. Identifying core domains to assess flare in rheumatoid arthritis: an OMERACT international patient and provider combined Delphi consensus. *Annals of the Rheumatic Diseases* 2012; 71(11):1855–60. doi: 10.1136/annrheumdis-2011-201201 PMID: 22772326.

[12]Gossec L, Dougados M, Rincheval N, Balanescu A, Boumpas DT, Canadelo S et al. Elaboration of the preliminary Rheumatoid Arthritis Impact of Disease (RAID) score: a EULAR initiative. *Annals of the Rheumatic Diseases* 2009; 68(11):1680–5. doi: 10.1136/ard.2008.100271 PMID: 19054825.

[13]Berthelot JM, De BM, Morel J, Benatig F, Constantin A, Gaudin P et al. A tool to identify recent or present rheumatoid arthritis flare from both patient and physician perspectives: The 'FLARE' instrument. *Annals of the Rheumatic Diseases* 2012; 71(7):1110–6. doi: ard.2011.150656 [pii]; doi: 10.1136/ard.2011.150656 PMID:22072015

[14]Bingham CO 3rd, Alten R, Bartlett SJ, Bykerk VP, Brooks PM, Choy E et al. Identifying preliminary domains to detect and measure rheumatoid arthritis flares: report of the OMERACT 10 RA Flare Workshop. *Journal of Rheumatology* 2011; 38(8):1751–8. doi: 10.3899/jrheum.110401 PMID: 21807797.

[15]Critical Path Institute, Patient-reported Outcome Consortium, Rheumatoid Arthritis Working Group. Toward Consensus Development: Qualifying Endpoint Measures for Rheumatoid Arthritis Clinical Trials. 2012; Accessed February 1, 2016 at: https://c-path.org/towards-consensus-development-qualifying-endpoint-measures-for-rheumatoid-arthritis-clinical-trails/

[16]Hewlett S, Sanderson T, May J et al. "I'm hurting, I want to kill myself": rheumatoid arthritis flare is more than a high joint count—an international patient perspective on flare where medical help is sought. *Rheumatology (Oxford)* 2012; 51 (1): 69-76. doi: 10.1093/rheumatology/keq455 PMID: 21565901

[17]Orbai AM, Halls S, Hewlett S, Bartlett SJ, Leong AL, Bingham CO, 3rd. More than Just Minutes of Stiffness in the Morning: Report from the OMERACT Rheumatoid Arthritis Flare Group Stiffness Breakout Sessions. *Journal of Rheumatology* 2015; 42(11):2182-2184. doi: 10.3899/jrheum.141172 PMID: 25729035

[18]Orbai AM, Smith KC, Bartlett SJ, De Leon E, Bingham CO, 3rd. "Stiffness has different meanings, I think, to everyone": examining stiffness from the perspective of people living with rheumatoid arthritis. *Arthritis Care & Research (Hoboken)* 2014; 66(11):1662-1672. doi: 10.1002/acr.22374 PMID: 24891304

[19]Gossec L, Paternotte S, Aanerud GJ et al. Finalisation and validation of the rheumatoid arthritis impact of disease score, a patent-derived composite measure of rheumatoid arthritis: a EULAR initiative. *Annals of Rheumatic Diseases* 2011; 70(6):935-942. doi: 10.1136/ard.2010.142901 PMID: 21540201

[20]Sanderson TC, Hewlett SE, Flurey C, Dures E, Richards P, Kirwan JR. The impact triad (severity, importance, self-management) as a method of enhancing measurement of personal life impact of rheumatic diseases. *Journal of Rheumatology* 2011; 38(2):191-194.

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

[22]Ter Wee MM, van Tuyl LH, Blomjous BS, Lems WF, Boers M, Terwee CB. Content validity of the Dutch Rheumatoid Arthritis Impact of Disease (RAID) score: results of focus group discussions in established rheumatoid arthritis patients and comparison with the International Classification of Functioning, Disability and Health core set for rheumatoid arthritis. *Arthritis Research & Therapy* 2016; 18(1):22.

[23]van Tuyl LH, Sadbnova M, Davis B et al. Remission in rheumatoid arthritis: working toward incorporation of the patient perspective at OMERACT 12. *Journal of Rheumatology* 2016; 43(1):203-207.

[24]Hewlett S, Hehir M, Kirwan JR. Measuring fatigue in rheumatoid arthritis: a systematic review of scales in use. *Arthritis & Rheumatology* 2007; 57(3):429-439.

[25]Kirwan JR, Hewlett S. Patient perspective: reasons and methods for measuring fatigue in rheumatoid arthritis. *Journal of Rheumatology* 2007; 34(5):1171-1173

[26]Hewlett S, Dures E, Almeida C. Measures of fatigue: Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAF MDQ), Bristol Rheumatoid Arthritis Fatigue Numerical Rating Scales (BRAF NRS) 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* 2011; 63 Suppl 11:S263-286.

[27]Hewlett S, Chalder T, Choy E et al. Fatigue in rheumatoid arthritis: time for a conceptual model. *Rheumatology* 2011; 50(6):1004-1006.

[28]Zonna-Nacach A, Roseman JM, McGwin G Jr, Friedman AW, Baethge BA, Reveille JD et al. Systemic lupus erythematosus in three ethnic groups. VI. Factors associated with fatigue within 5 years of criteria diagnosis. *Lupus* 2000; 9: 101–9.

[29]Pollard LC, Choy EH, Gonzalez J, Khoshaba B, Scott DL. Fatigue in rheumatoid arthritis reflects pain, not disease activity. *Rheumatology* (Oxford) 2006; 45: 885–9.

[30]Nicklin J, Cramp F, Kirwan J et al. Collaboration with patients in the design of patient-reported outcome measures: capturing the experience of fatigue in rheumatoid arthritis. *Arthritis Care & Research* (Hoboken) 2010; 62:1552–8.

[31]Minnock P, Kirwan J, Bresnihan B. Fatigue is a reliable, sensitive and unique outcome measure in rheumatoid arthritis. *Rheumatology (Oxford)* 2009; 48: 1533–6.

[32]Hewlett S, Carr M, Ryan S et al. Outcomes generated by patients with rheumatoid arthritis: how important are they? *Musculoskeletal Care* 2005; 3:131–42.

[33]Kirwan JR, Newman S, Tugwell PS et al. Progress on incorporating the patient perspective in outcome assessment in rheumatology and the emergence of life impact measures at OMERACT 9. *Journal of Rheumatology* 2009; 36:2071–6.

[34]van Tuyl LH, Smolen JS, Wells GA et al. Patient perspective on remission in rheumatoid arthritis. *Journal of Rheumatology* 2011; 38:1735–8.

[35]Smolen JS, Aletaha D, Bijlsma JWJ et al. Treating rheumatoid arthritis to target: recommendations of an international task force. *Annals of Rheumatic Diseases* 2010; 69: 631–37.

[36]Arnett FC, Edworthy SM, Bloch DA et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis & Rheumatology* 1988; 31:315–24.

[37]Bingham CO and Bartlett SJ. Qualification of a measure of fatigue for use in the assessment of treatment benefit in rheumatoid arthritis clinical trials. A report prepared for the Critical Path Institute, PRO Consortium, RA Working Group. June 15, 2016.

[38]Alten R, Pohl C, Choy EH, Christensen R, Furst DE, Hewlett SE et al. Developing a construct to evaluate flares in rheumatoid arthritis: a conceptual report of the OMERACT RA Flare Definition Working Group. *Journal of Rheumatology* 2011; 38(8):1745–50. doi: 10.3899/jrheum.110400 PMID: 21807796.

[39]Bingham CO III, Alten R, de Wit MP. The importance of patient participation in measuring rheumatoid arthritis flares. *Annals of the Rheumatic Diseases* 2012; 71(7):1107–9. doi: 10.1136/annrheumdis-2011-200870 PMID: 22323439

[40]Keller S, Mangrum R, Yang M. The measurement of fatigue in rheumatoid arthritis: review and identification of gaps in evidence. *American Institutes for Research (AIR)* 2014. Accessed February 1, 2016, at http://w w w .nihpromis.com/Documents/FatigueMeasureRA2014_ES_508.pdf?AspxAutoDetectCookieSupport=1

[41]Strand V, Simon LS, Bingham III CO, Hewlett S. Gap analysis on measures used to assess fatigue in adult patients with rheumatoid arthritis. A report prepared for the Critical Path Institute, PRO Consortium, RA Working Group. August 18, 2014.

[42]Kaiser K, Shaunfield S, Clayman ML, Ruderman E, Cella D. Content validation of the Functional Assessment of Chronic Illness Therapy (FACIT)-Fatigue Scale in moderately to highly active rheumatoid arthritis. *Rheumatology (Sunnyvale) 2016;* 6:193. doi: 10.4172/2161-1149.1000193

[43]Cella D, Yount S, Sorensen M, Chartash E, Sengupta N, Grober J. Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. *Journal of Rheumatology* 2005; 32:811-819.

[44]Cella D, Riley W, Stone AA, Rothrock N, Reeve B, Yount S et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology* 2010; 63(11):1179–94.

[45]Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Medical Care* 2007; 45(5 Suppl 1):S3–S11.

[46]DeWalt DA, Rothrock N, Yount S, Stone AA, PROMIS Cooperative Group. Evaluation of Item Candidates: The PROMIS Qualitative Item Review. *Medical Care* 2007; 45(5 Suppl 1):S12-S21.

[47]Cella D, Lai JS, Jensen SE, Christodoulou C, Junghaenel DU, Reeve B, Stone AA. PROMIS Fatigue Item Bank has clinical validity across diverse chronic conditions. *Journal of Clinical Epidemiology* 2016; 73:128-134. doi:10.1016/j.jclinepi.2015.08.037

[48]Swain M. Fatigue in chronic disease. *Clinical Science* 2000; 9:1-8.

[49]Piper BF. Fatigue: current bases for practice. In: Funk SG, Tornquist EM, Champagne MT, Copp LA, Weise R, editors. Key Aspects of Comfort: Management of Pain, Fatigue and Nausea. New York: Spring Publishing; 1989:197-98.

[50]Wolfe F, Hawley DJ, Wilson K. The prevalence and meaning of fatigue in rheumatic disease. *Journal of Rheumatology* 1996; 23(8):1407-17

[51]Riemsma, RP, Rasker JJ, Taal E, Griep EN, Wouters JM, Wiegman O. Fatigue in rheumatoid arthritis: the role of self-efficacy and problematic social support. *Rheumatology* 1998; 37(10), 1042-1046.

[52]Huyser, BA, Parker JC, Thoreson R, Smarr KL, Johnson JC, Hoffman R. Predictors of subjective fatigue among individuals with rheumatoid arthritis. *Arthritis & Rheumatology* 1998; 41(12), 2230-2237.

[53]Dekkers, JC, Geenen, R, Godaert, GL, Doornen, LJ, Bijlsma, JW. Diurnal courses of cortisol, pain, fatigue, negative mood, and stiffness in patients with recently diagnosed rheumatoid arthritis. *International Journal of Behavioral Medicine* 2000; 7(4), 353-371.

[54]Davis MC, Zautra AJ, Younger J, Motivala SJ, Attrep J, Irwin MR. Chronic stress and regulation of cellular markers of inflammation in rheumatoid arthritis: implications for fatigue. *Brain, Behavior & Immunity* 2008; 22(1):24-32.

[55]Druce KL, Jones GT, Macfarlane GJ, Basu N. Patients receiving anti-TNF therapies experience clinically important improvements in RA-related fatigue: results from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis. *Rheumatology (Oxford)* 2015; 54(6):964-971.

[56]Bingham CO, Weinblatt M, Han C et al. The effect of intravenous golimumab on health-related quality of life in rheumatoid arthritis: 24-week results of the phase III GO-FURTHER trial. *Journal of Rheumatology* 2014; 41(6):1067-1076.

[57]Strand V, Burmester GR, Ogale S, Devenport J, John A, Emery P. Improvements in health-related quality of life after treatment with tocilzumab in patients with rheumatoid arthritis refractory to tumour necrosis factor inhibitors: results from the 24-week randomized controlled RADIATE study. *Rheumatology (Oxford)* 2012; 51(10):1860-1869.

[58]Bartlett SJ, Bykerk VP, Cooksey R et al. Feasibility and Domain Validation of Rheumatoid Arthritis (RA) Flare Core Domain Set: Report of the OMERACT 2014 RA Flare Group Plenary. *Journal of Rheumatology* 2015; 42(11):2185-2189.

[59]Minnock P, Bresnihan B. Pain outcomes and fatigue levels reported by women with established rheumatoid arthritis. *Arthritis & Rheumatolgy* 2004; 50 Suppl S471.

[60]Van Tuyl L, Hewlett S, Sadlonova M et al. The patient perspective on remission in rheumatoid arthritis: 'You've got limits, but you're back to being you again.' *Annals of the Rheumatic Diseases* 2015; 74(6):1004-1010.

[61]Carr A, Hewlett S, Hughes R et al. Rheumatology outcomes: the patient's perspective. *Journal of Rheumatology* 2003; 30(4):880-883.

[62]Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. *Arthritis & Rheumatology* 1995; 38:727-735.

[63]Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis & Rheumatology* 1995; 38:44-48.

[64]Lai JS, Cella D, Choi SW, Junghaenel DU, Christodolou C, Gershon R et al. How item banks and their application can influence measurement practice in rehabilitation medicine: a PROMIS fatigue item bank example. *Archives of Physical Medicine & Rehabilitation* 2011; 92:S20e7.

[65]Christodoulou C, Junghaenel DU, DeWalt DA, Rothrock N, Stone AA. Cognitive interviewing in the evaluation of fatigue items: results from the patient-reported outcomes measurement information system (PROMIS). *Quality of Life Research* 2008; 17:1239e46.

[66]Junghaenel DU, Christodoulou C, Lai JS, Stone AA. Demographic correlates of fatigue in the US general population: results from the patient-reported outcomes measurement information system (PROMIS) initiative. *Journal of Psychosomatic Research* 2011; 71(3):117e23.

[67]World Health Organization. International Classification of Functioning, Disability and Health (ICF). Geneva, Switzerland: *World Health Organization* 2001.

[68]Furst DE, Schiff MH, Fleischmann RM, Strand V, Birbara CA, Compagnone D, Fischkoff SA, Chartash EK. Adalimumab, a fully human anti-tumor necrosis factor-alpha monoclonal antibody, and concomitant standard antirheumatic therapy for the treatment of rheumatoid arthritis: results of STAR (Safety Trial of Adalimumab in Rheumatoid Arthritis). *Journal of Rheumatology* 2003; 30(12):2563-2571.

[69]Lai, J-S, Cella D, Yanez B, Stone A. Linking fatigue measures on a common reporting metric. *Journal of Pain and Symptom Management* 2014; 48(4): 639-648.

[70]Bjorner JB, Rose M, Gandek B, Stone AA, Junghaenel DU, Ware, JE. Method of administration of PROMIS scales did not significantly impact score level, reliability, or validity. *Journal of Clinical Epidemiology* 2014; 67(1), 108-113.

[71]Paul E, Fleischmann R, Filipowicz-Sosnowska A, Schechtman J, Szczepanski L, Kavanaugh A, Racewicz AJ et al. The efficacy and safety of rituximab in patients with active rheumatoid arthritis despite methotrexate treatment: results of a phase IIB randomized, double-blind, placebo-controlled, dose-ranging trial. *Arthritis & Rheumatology* 2006; 54(5): 1390-1400.

[72]Genovese MC, McKay JD, Nasonov EL, Mysler EF, da Silva NA, Alecock E, Woodworth T, Gomez-Reino JJ. Interleukin-6 receptor inhibition with tocilizumab reduces disease activity in rheumatoid arthritis with inadequate response to disease-modifying antirheumatic drugs: The tocilizumab in combination with traditional disease-modifying antirheumatic drug therapy study. *Arthritis & Rheumatism* 2008; 58(10): 2968-2980.

[73]Fleischmann R, Kremer J, Cush J, Schulze-Koops H, Connell CA, Bradley JD, Gruben D, Wallenstein GV, Zwillich SH, Kanik KS. Placebo-controlled trial of tofacitinib monotherapy in rheumatoid arthritis. *New England Journal of Medicine* 2012; 367(6): 495-507.

[74]Weinblatt ME, Kavanaugh A, Genovese MC, Musser TK, Grossbard EB, Magilavy DB. An oral spleen tyrosine kinase (Syk) inhibitor for rheumatoid arthritis. *New England Journal of Medicine* 2010; 363(14): 1303-1312.

[75]Cheng G, Saleh MN, Marcher C, Vasey S, Mayer B, Aivado M, Arning M, Stone AL, Bussel JB. Eltrombopag for management of chronic immune thrombocytopenia (RAISE): a 6-month, randomised, phase 3 study. *The Lancet* 2011; 377(9763): 393-402.

[76]Emery P, Deodhar A, Rigby WF, Isaacs JD, Combe B, Racewicz AJ, Latinis K, et al. Efficacy and safety of different doses and retreatment of rituximab: a randomised, placebo-controlled trial in patients who are biological naive with active rheumatoid arthritis and an inadequate response to methotrexate (Study Evaluating Rituximab's Efficacy in MTX iNadequate rEsponders (SERENE). *Annals of the Rheumatic Diseases* 2010.

[77]Willis GB, Lessler JT. Question appraisal system: QAS-99. *National Cancer Institute* 1999.
[78]Reeve B, Fayers PM. Applying item response theory modeling for evaluating questionnaire item and scale properties. In: Fayers PM, Hays R, eds. Assessing quality of life in clinical trials. New York, NY: Oxford University Press 2005; 55-73.

[79]Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research* 2007; 16 Suppl 1:5-18.

[80]Reeve, BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the patient-reported outcomes measurement information system (PROMIS). *Medical Care* 2007; 45(5), S22-S31.

[81]Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, No. 17. Richmond, VA: Psychometric Society; 1969: http://www.psychometrika.org/journal/online/MN17.pdf

[82]Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced cancer patients. *Journal of Clinical Epidemiology* 2011; 64(5):507-516.

[83]Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value in Health* 2005; 8(2), 94-104.

[84]Wild D, Eremenco S, Mear I, Martin M, Houchin C, Gawlicki M, Hareendran A, Wiklund I, Chong LY, von Maltzahn R, Cohen L, Molsen E. 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 in Health* 2009; 12(4), 430-440.

[85]Eremenco SL, Cella D, & Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Evaluation and the Health Professions* 2005; 28(2), 212–232.

Updated October 17, 2018