Clinical Outcome Assessments (COA) Qualification Program
DDT COA #000015: PROMIS® Fatigue Short Form 10a
Qualification Plan

Section 1: Proposed COA Qualification

1.1.1 Overview of Fatigue in Rheumatoid Arthritis

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 (Helmick et al., 2008; Hootman & Helmick, 2006; Lawrence et al., 2008; Sanderson & Kirwan, 2009; Sanderson, Morris, Calnan, Richards, & Hewlett, 2010a). In addition to symptoms such as pain, impaired physical function, stiffness, sleep disturbance, and emotional distress (Bartlett et al., 2012; Berthelot et al., 2012; Bingham et al., 2011; Felson et al., 1995; Gossec et al., 2009; Hewlett, Cockshott, et al., 2005; Kirwan et al., 2007; Sanderson, Morris, Calnan, Richards, & Hewlett, 2010b; Singh et al., 2012), patient and clinician input reveal that fatigue has been identified as a common, persistent, disabling, and high-priority symptom in RA (Bartlett, Orbai, et al., 2015; Critical Path Institute PRO Consortium RA Working Group, 2012; Gossec et al., 2011; Hewlett et al., 2012; Orbai et al., 2015; Orbai, Smith, Bartlett, De Leon, & Bingham, 2014; Sanderson et al., 2011). Like pain, fatigue is a very important aspect of RA (Bartlett et al., 2012; Berthelot et al., 2012; Hewlett, Cockshott, et al., 2005; Hewlett et al., 2012; Kirwan et al., 2007; Ter Wee et al., 2016; van Tuyl et al., 2016), and a high priority for RA patients seeking treatment (Sanderson et al., 2010a, 2010b). 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 (Hewlett, Cockshott, et al., 2005; Hewlett, Hehir, & Kirwan, 2007; Kirwan & Hewlett, 2007). 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) (Hewlett, Chalder, et al., 2011; Hewlett, Dures, & Almeida, 2011; Pollard, Choy, Gonzalez, Khoshaba, & Scott, 2006; Zonna-Nacach et al., 2000). Findings also indicate the heterogeneous experience of fatigue in RA, and its persistence even when traditional indicators suggest well-controlled RA activity (Hewlett et al., 2012).

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 (Critical Path Institute PRO Consortium RA Working Group, 2012; Felson et al., 1995; Hewlett, Cockshott, et al., 2005; Kirwan et al., 2007; Minnock, Kirwan, & Bresnihan, 2009; Nicklin, Cramp, Kirwan, Urban, & Hewlett, 2010). Moreover, in response to

the increasingly feasible goal of achieving low disease activity and/or remission through intervention (Hewlett, Carr, et al., 2005), 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 (Arnett et al., 1987; Kirwan et al., 2009; Smolen et al., 2010; van Tuyl et al., 2011). 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 (Felson et al., 1995). Input from patients and healthcare professionals emphasized the importance of the patient's perspectives when studying RA remission to ensure optimization of targeted therapy (Sanderson et al., 2011) and the inclusion of domains of importance in clinical trials, especially fatigue. Patient-reported outcome (PRO) measures serve as the best method for assessing symptoms like fatigue, as its severity and impact are best known by the patient. During the development of the Qualification Plan, patient representatives with RA also emphasized that assessments of patient-reported outcomes are important to include in medical research. Taken together, this highlights the need to include fatigue PRO measures when evaluating RA treatments.

In clinical trials and clinical practice, fatigue has not been assessed as often as physical function, pain, and disease activity in RA clinical trials, and 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) (Felson et al., 1995), 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 (Bingham & Bartlett, 2016), and the benefits of monitoring subtle symptom changes in the context of low disease activity/remission (Alten et al., 2011; Bingham et al., 2011; Bingham, Alten, & de Wit, 2012), 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" (Critical Path Institute PRO Consortium RA Working Group, 2012). 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 (Bingham & Bartlett, 2016; Critical Path Institute PRO Consortium RA Working Group, 2012; Keller, Mangrum, & Yang, 2014; Strand, Simon, Bingham, & Hewlett, 2014). 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 (Bingham & Bartlett, 2016; Critical Path Institute PRO Consortium RA Working Group, 2012; Keller et al., 2014; Strand et al., 2014). Additionally, there was insufficient evidence that frequently used fatigue PRO measures captured the full range of the fatigue experience in RA (Strand et al., 2014).

These reports concluded that the *Patient-Reported Outcomes 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 evidence of its reliability and validity in RA populations. **The ten items that comprise the *PROMIS Fatigue Short Form 10a (Kaiser, Shaunfield, Clayman, Ruderman, & Cella, 2016)* are proposed here as the clinical outcome assessment (COA) tool to be qualified for use for 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.**

The development of the ***PROMIS Fatigue Short Form 10a*** has involved multiple phases across several measurement systems. To provide an introduction to the history of this measure's development, the following sub-sections provide a chronological background on the phases of development that have led to the ***PROMIS Fatigue Short Form 10a*** as a candidate for a COA for use in RA clinical trials, beginning with the development of the Functional Assessment of Chronic Illness Therapy (FACIT) Fatigue Scale. Subsequent sections of the Qualification Plan will focus specifically on details related to the ***PROMIS Fatigue Short Form 10a.***

1.1.2 The Functional Assessment of Chronic Illness Therapy (FACIT) Fatigue

The 95-item PROMIS Fatigue Item Bank includes, verbatim, the 13 items that comprise the FACIT Fatigue. Therefore, it is relevant to trace the development of the ***PROMIS Fatigue Short Form 10a*** back to the development and validation of the FACIT Fatigue, especially as it relates to RA. Originally developed in 1994-1995 to assess fatigue secondary to cancer-related anemia, the FACIT Fatigue scale has subsequently demonstrated validity in a wide variety of chronic illness populations, including RA (see Section 1.1.2.1). The FACIT Fatigue scale development and validation methodology took place in four phases: item generation, item-review and selection, scale construction, and psychometric evaluation. These steps are detailed in sequence below.

Candidate items for the FACIT Fatigue were generated using comprehensive, manual-driven semi-structured interviews with 15 anemic oncology patients and 5 medical providers. Questions focused on personal definitions of quality of life, symptoms associated with anemia, and the nature of the impact of cancer and disease/treatment-related anemia on varied

dimensions (physical status, emotional well-being, family issues, sexuality/intimacy, social The patients also rated the relevance of 32 symptoms and concerns associated with cancer, using a 0 to 10 scale (0 = “not at all” and 10 = “as much as I could imagine”). This information was used to help generate new items. These issues were tabulated, and the frequency of responses noted. An exhaustive and inclusive list of items made up the first round of the item pool. A sample of medical experts and patients then provided candidate items.

The purpose of the next phase of the FACIT Fatigue development was to reduce the item list to a manageable and appropriately representative subset of items. Item selection was accomplished using a two-step, team-based systematic item review process led by the principal investigator (D. Cella), and included two clinical psychologists, two oncologists, one oncology nurse, and a research assistant. During item review, every expert stated that of all the problems associated with anemia, fatigue and its effects were paramount. Patient interview data confirmed this inasmuch as many frustrations related to having sufficient energy to conduct one’s daily affairs were acknowledged. The item review team therefore committed to a decision to ensure sufficient emphasis on fatigue and its effects to enable a distinct FACIT Fatigue Scale to emerge from the FACT-Anemia Subscale. This process resulted in retaining 40 candidate items. These 40 items were then brought back to the original panel of five experts and to two new (independent) experts for comment and assistance in developing a “testable instrument of approximately 20 questions.” Expert review and comments were reconciled back against the original patient ratings of importance of general areas, and the original interview data, to derive the final set of items for testing. The final set of 21 items represented anemia symptoms which included but were not limited to fatigue.

The initial validation sample consisted of 49 patients ranging in age from 19 to 83 years (median age, 56 years) and represented a broad spectrum of cancer diagnoses (Yellen, Cella, Webster, Blendowski, & Kaplan, 1997). In accordance with expert and patient emphasis provided during previous phases of development, a 13-item FACIT Fatigue Scale was produced and published from this validation testing. Higher scores on the FACIT Fatigue Scale represent lower fatigue. When tested on the sample of 49 anemic oncology patients (range of hemoglobin was 7-15.9 g/dL), the 13-item FACIT Fatigue Scale demonstrated strong psychometric properties. The FACIT Fatigue Scale total score was stable over two distinct (test-retest) time points (Spearman correlation coefficient = 0.90) and internally consistent (Cronbach’s alpha coefficient initial/retest range = 0.93-0.95). The FACIT Fatigue Scale total score also showed significant relationships with performance status and hemoglobin level. Convergent and discriminant validity testing revealed a significant positive relationship with other known measures of fatigue, a significant negative relationship with vigor, and a predicted lack of relationship with social desirability. The FACIT Fatigue Scale differentiated patients by hemoglobin level ($p<0.05$) and patient-rated performance status ($p<0.0001$). This initial evaluation suggested that the FACIT Fatigue Scale would be a reliable, valid, and useful measure of fatigue in cancer patients and perhaps in other groups (Yellen et al., 1997). Subsequent research has confirmed this, and further elaborated upon the meaningfulness of differences and changes over time on the FACIT Fatigue Scale.

Some have suggested that the 13 items fall into two definable components of fatigue: fatigue

experience and fatigue impact (i.e., the impact of fatigue upon activities)(Cella, Lai, & Stone, assessed in two general population (total N=1,878) and two cancer (total N=3,140) samples (Cella et al., 2011). Results suggested that the fatigue experience and the impact of fatigue upon function are reported along a single dimensional continuum, but that experience is more likely than impact on function to be endorsed at lower levels of fatigue. Across samples, experience scores were systematically higher (more fatigue) than impact scores, by a magnitude of 0.21 to 0.46 SD units. Fatigue as an outcome or trial endpoint can be expressed as a single number, and the experience of the symptom is more likely to be endorsed at mild levels of fatigue, presumably before the symptom exerts a measurable or meaningful impact on function.

1.1.2.1 FACIT Fatigue in RA

Over time, the application of the FACIT Fatigue expanded to non-cancer populations, including RA. In 2005, Cella and colleagues (Cella et al., 2005) evaluated the psychometric performance of the FACIT Fatigue using data from the *Safety Trial of Adalimumab in Rheumatoid Arthritis* (hereafter, STAR trial). The STAR trial included a sample of patients with RA enrolled in a 2-arm, double-blind randomized 24-week clinical trial of adalimumab (1999-2000); safety and efficacy results from the trial were published in 2003 (Furst et al., 2003). Study subjects had a median age of 56 years (range 21–86) and were largely female (79%) and non-Hispanic white (88%). One set of confirmatory analyses included data from a second sample of 271 patients with RA enrolled in a separate 4-arm (equal sample sizes) 24 week, randomized, double blind, placebo controlled clinical trial. The FACIT Fatigue was tested along with measures previously validated in RA: the Multidimensional Assessment of Fatigue (MAF) and the Medical Outcomes Study Short-Form 36 (SF-36) Vitality. The clinical endpoint used was the ACR (American College of Rheumatology) 20/50/70, defined by proportion of reduction in symptoms related to joint swelling and pain. The FACIT Fatigue showed good internal consistency (Cronbach's alpha coefficient range across timepoints = 0.86 to 0.87). Spearman correlation coefficients showed a very high degree of association with SF-36 Vitality (r = 0.73 to 0.84) and the MAF (r = -0.84 to -0.88). Coefficients of association with MAF are negative because high scores on that scale reflect worse fatigue. Changes in FACIT Fatigue scale scores over 24 weeks successfully discriminated between groups defined by levels of the clinical endpoint, the ACR20, ACR50, and ACR70. That is, patients with greater clinical improvement in RA also showed larger increases in their FACIT Fatigue scale scores, indicating decreased levels of fatigue. The minimally important difference (MID) is the smallest difference in a scale score that is thought to have clinical significance. The MID values for the FACIT Fatigue scale were confirmed in a second sample of patients with RA that produced nearly identical values. A 3- to 4-point change in the FACIT Fatigue scale is considered clinically significant. The FACIT Fatigue performed similarly in many respects to the other measures, however, the 13-item FACIT Fatigue showed a broader coverage of the fatigue continuum in RA patients than the longer 16-item MAF.

1.1.3 Development of the PROMIS Fatigue Item Bank

Adapting the tripartite framework of physical, mental, and social health from the World Health Organization (WHO; 2007), PROMIS researchers developed multiple item banks for each domain (Cella et al., 2010), including one for fatigue (Christodoulou, Junghaenel, DeWalt, Rothrock, & Stone, 2008; Junghaenel, Christodoulou, Lai, & Stone, 2011; Lai et al., 2011). The WHO's International Classification of Functioning, Disability and Health included the

minimization of fatigue among its stated aims (World Health Organization, 2001), 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 (Christodoulou et al., 2008; Junghaenel et al., 2011; Lai et al., 2011). Given that the 10 items contained within the ***PROMIS Fatigue Short Form 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 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 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 (Cella et al., 2010). 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 (Cella et al., 2005; DeWalt, Rothrock, Yount, Stone, & PROMIS Cooperative Group, 2007). 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 (Kaiser et al., 2016). In total, the PROMIS Fatigue Item Bank includes 38 fatigue experience and 57 fatigue impact (i.e., interference) items (Cella et al., 2016). Although RA patients were among the patient groups included during the development of the PROMIS Fatigue Item Bank (Cella et al., 2010), the items in the bank were not developed specifically for use in RA.

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 (a) could be used to assess a wide range of fatigue across many disease states, (b) capitalized on the benefits of inclusion in the PROMIS measurement framework (particularly standardization and uniformity across measures), and (c) 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 measure development (Cella et al., 2010). These procedures (detailed below) included concept elicitation and definition, the identification of a large pool of candidate items, extensive cognitive interviewing (Christodoulou et al., 2008), 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 (Lai et al., 2011).

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 (Cella et al., 2010; Lai, Cella, Yanez, & Stone, 2014). 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: FDA (through two meetings held at FDA in 2006-2008), OMERACT, the National Institute of Arthritis, Musculoskeletal, and Skin Diseases (NIAMS; through its leadership role in the PROMIS effort), the National Cancer Institute (NCI; through science officer participation at steering committee meetings and Cancer Patient-Reported Outcomes Measurement Information System project meetings at NCI in 2006-2008), the World Health Organization (WHO) 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.

1.1.3.1 PROMIS Fatigue Item Bank Development

1.1.3.1.1 Construct Definition and Item Identification

The team of experts responsible for developing the PROMIS Fatigue Item Bank began by conducting a literature review (Lai et al., 2011) 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 (Christodoulou et al., 2008). 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” (Christodoulou et al., 2008). 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.

1.1.3.1.2 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 (Christodoulou et al., 2008). 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 University of North Carolina (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 (Willis & Lessler, 1999). 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 (Christodoulou et al., 2008).

1.1.3.1.3 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 (Lai et al., 2011).

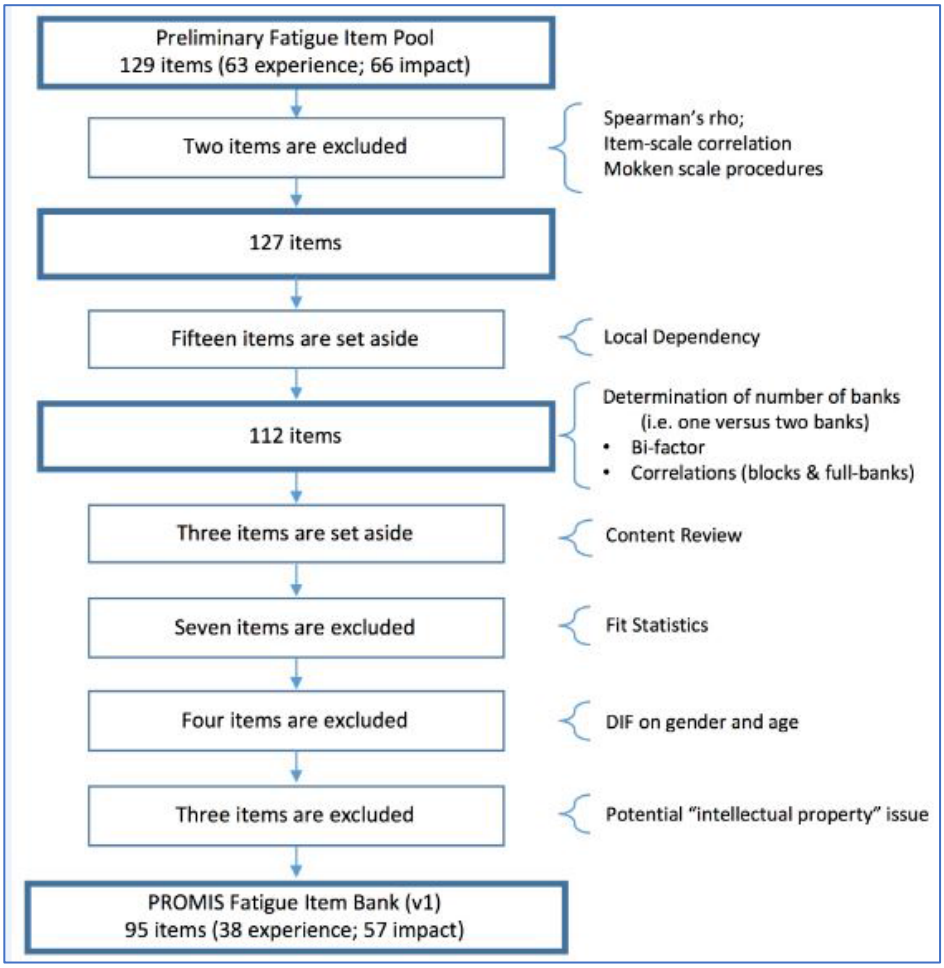
1.1.3.1.4 Large-Scale Field Testing

The remaining candidate items were field tested by 21,133 participants (Cella et al., 2010). The majority of the sample (93%; n = 19,601) was recruited by Polimetrix, a polling firm based in Palo Alto, California; the remaining sample 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 demographic proportions from the Year 2000 US Census. 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), chronic obstructive pulmonary disease (COPD) (n = 1,214), psychiatric disorders (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.

[Remainder of page intentionally left blank]

Data from the fieldtesting sample were used for a range of quantitative analyses (Cella et al., 2010; Lai et al., 2011) (see Figure 1). 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 (Lai et al., 2011). In the first step, two items from the "impact" item set were removed due to low itemscale 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). Bifactor 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 (Lai et al., 2011).

Figure 1: Steps to develop the PROMIS Fatigue Item Bank



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 (Lai et al., 2011). Relevant to this context, none of the 13 FACIT Fatigue items were removed in this process.

These items were then calibrated and placed on the same metric using a graded response IRT model (Lai et al., 2011). The resulting calibrations are used to score the general PROMIS Fatigue Item Bank and all PROMIS Fatigue short forms, including the ***PROMIS Fatigue Short Form***

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 (Edelen & Reeve, 2007; Reeve & Fayers, 2005). 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 (Reeve et al., 2007; Samejima, 1969), the items in PROMIS banks are highly inter-correlated, yielding a single dimension that explains the large majority of variance in person-level scores. Higher scores on the PROMIS fatigue items reflect greater fatigue.

1.1.3.2 Relationship between FACIT Fatigue and PROMIS Fatigue Item Bank

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 (Lai et al., 2011). All 13 items were among the 95 items retained in the final item bank after following the PROMIS instrument development and validation standards (see Attachment 1). The primary goal 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., ability to participate in social roles and satisfaction with that ability), and mental health (e.g., depression, anxiety). A thorough description of the methods used to support this goal is provided in Attachment 1.

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 Fatigue Short Form 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.

1.1.3.3 Description of the PROMIS Fatigue 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; not disease-specific. Yet, items can be selected preferentially, based upon their relevance to a given disease or condition. 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 (Cella et al., 2010; Junghaenel et al., 2011). This means that all PROMIS Fatigue measures – including the ***PROMIS Fatigue Short Form 10a*** – have a mean score of 50 and a standard deviation of 10. Thus, a T-score of 60 is one standard deviation (SD) higher (more fatigue) than the "average person" in the US. The theoretical range of scores for the ***PROMIS Fatigue Short***

Form 10a extends infinitely in both directions, but in practice the range is from 30 to 85 T units (Lai et al., 2014). This covers 5.5 standard deviation units, quite a broad range, in the general population. Additional information about scoring (including tables for the conversion of "raw" scores to the T-score metric) is provided in Attachment 2.

1.1.4 Administration of PROMIS Fatigue

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 among respondents taking a CAT, scores are comparable across respondents.

Fixed-length short forms have also been developed for the PROMIS Fatigue domain. The ***PROMIS Fatigue Short Form 10a*** (see Appendix A) is one of several available short forms made from the PROMIS Fatigue Item Bank (all of the items in the PROMIS Fatigue Item Bank are shown in Appendix B).

Several modes of data collection are possible using the ***PROMIS Fatigue Short Form 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 regard to the evidence regarding differences across modes, Bjorner and colleagues (Bjorner et al., 2014) 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 COPD, depression, or RA. The four methods included (a) paper-and-pencil, (b) automated data collection by telephone using interactive voice response, (c) computer connected to the internet, and (d) personal digital assistant (PDA). In difference score analyses, no significant method differences were found, and all confidence intervals (CIs) 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 to 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.

1.1.5 Previous PROMIS Fatigue Research in RA

1.1.5.1 PROMIS Fatigue CAT administration in John Hopkins Arthritis Center

A growing body of research has utilized items from the PROMIS Fatigue Item Bank to assess fatigue in individuals with RA. 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. (Bartlett, Orbai, et al., 2015). In addition to other PROMIS measures and clinical instruments, the PROMIS Fatigue CAT was administered. The CAT was programmed to administer 4 to 8 items until a standard error of 0.3 or less for respondent's scores was detected. Cronbach's alpha coefficient, a measure of internal consistency reliability,

was 0.99. Thirty-four participants completed a test-retest assessment approximately two days later. Among these 34 participants, the Pearson correlation coefficient for reliability was 0.88.

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) (Bingham & Bartlett, 2016). Further evidence of construct validity is indicated in Table 1; PROMIS Fatigue scores were also moderately correlated with Clinical Disease Activity Index (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 2 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 3.

Table 1: Correlations of PROMIS Fatigue T-scores with related clinician- or patient reported measures in an RA cohort (Bingham & Bartlett, 2016)

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

Table 2: PROMIS Fatigue T-scores among groups with differing levels of disease activity based on the Clinical Disease Activity Index (CDAI) (Bingham & Bartlett, 2016)

	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 3: PROMIS Fatigue T-scores and "flare" activity in RA patients (Bingham & Bartlett, 2016)

	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.

In the Johns Hopkins Arthritis Center cohort (Bartlett, Orbai, et al., 2015), 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 4 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 4: 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 5).

Table 5: Change in PROMIS Fatigue CAT scores by Clinical Disease Activity Index (CDAI) change 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 (Cella et al., 2016). 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 6. 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 6: Change in PROMIS Fatigue scores by general health change rating

		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

Note: SRM = standardized response mean = mean change / SD of change.

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 meaningful within-patient change will be evaluated in future RA clinical trials.

1.1.5.1 History of the PROMIS Fatigue in Rheumatoid Arthritis Drug Development Tool (DDT) COA Qualification Process

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 Fatigue Short Form 10a.*** At the time that the Initial Briefing Package was submitted, comprehensive evidence of content validity for the *PROMIS Fatigue Short Form 7a* was not available to the RA Working Group, while published evidence of the content validity for the ***PROMIS Fatigue Short Form 10a*** was available. In addition, the use of the FACIT Fatigue in a number of RA clinical trials would provide data to support the longitudinal properties of the ***PROMIS Fatigue Short Form 10a***, making it the better candidate measure to propose for qualification. The ***PROMIS Fatigue Short Form 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 past seven days and includes a 5-point verbal rating scale ranging from "Not at all" to "Very much." The ***PROMIS Fatigue Short Form 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 (Strand et al., 2014). There is considerable published evidence from these studies to support the content validity, reliability, convergent validity, concurrent validity, predictive validity, responsiveness to ACR clinical classification, and minimally important differences (MID) of the FACIT Fatigue in RA patients (Bechman et al., 2018; Campbell et al., 2012; Cella et al., 2005; Cohen et al., 2006; Emery et al., 2017; Janoudi et al., 2017; Keystone et al., 2017; Mayoux-Benhamou et al., 2008; Mease et al., 2008; Strand et al., 2016; Strand, Rentz, et al., 2012). Thus, the FACIT Fatigue provides a strong foundation to support the ***PROMIS Fatigue Short Form 10a*** given the former's excellent measurement properties. The ***PROMIS Fatigue Short Form 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).

1.1.5.2 Development of the PROMIS Fatigue Short Form 10a

Drawing from the history of research with the FACIT Fatigue and PROMIS Fatigue Item Bank in RA, Kaiser and colleagues (Kaiser et al., 2016) conducted an evaluation of the content validity of the original items in the 13-item FACIT Fatigue (all items are also included in the PROMIS Fatigue Item Bank) in a sample of 17 participants with moderately to highly active RA and found strong support for both the coverage and relevance of 10 items. These analyses were based on semi-structured interviews with each participant 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, participants 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 FACIT Fatigue**,** Kaiser and colleagues (Kaiser et al., 2016) concluded that while all 13 items related to aspects of fatigue that were relevant and important in the experience of fatigue among RA patients, 10 items exhibited the strongest support for content validity. Kaiser et al. (2016) did not identify any additional fatigue-related concepts that needed to be added to the ***PROMIS Fatigue Short Form 10a,*** further confirming sufficient coverage for people with RA. This resulted in the ***PROMIS Fatigue Short Form 10a***.

1.1 Concept of Interest for meaningful treatment benefit

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" (Cella et al., 2007). 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 (Keller et al., 2014). Sub-dimensions of fatigue experience in the RA literature include intensity, frequency, duration, variations in fatigue (e.g., unpredictable, irregular), and differentiation by cause (Keller et al., 2014). Sub-dimensions of fatigue impact include impact/consequences, sleep/rest, requirements/problems, physical ability, cognition, emotions, and coping (Keller et al., 2014). 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.

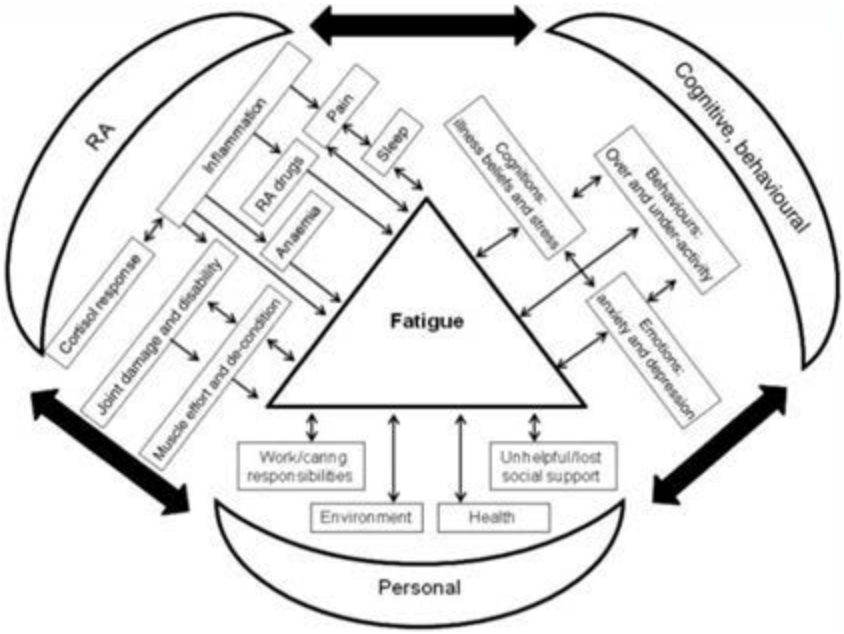
Patients with chronic disease frequently identify fatigue as one of the key factors affecting their quality of life (Swain, 2000) and the experience of fatigue can confer decrements to multiple domains of quality of life, including physical, emotional, social, and cognitive well-being (Bingham & Bartlett, 2016). 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 (Piper, 1989).

Fatigue represents a prominent and common symptom experienced by patients with RA, affecting an estimated 88 to 93% of patients (Piper, 1989). The etiology of fatigue in RA is multifactorial, with inflammatory processes, pain, anemia, sleep quality, and psychosocial factors all playing potential roles (Dekkers, Geenen, Godaert, van Doornen, & Bijlsma, 2000; Hewlett, Chalder, et al., 2011; Huyser et al., 1998; Riemsma et al., 1998; Wolfe, Hawley, & Wilson, 1996). Hewlett and colleagues (Hewlett, Chalder, et al., 2011) have developed a multidimensional conceptual model of fatigue in RA (see Figure 2) 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.

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) (Davis et al., 2008), as well as between fatigue and C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), and disease activity score (DAS28) (Davis et al., 2008). Further support for the relationship between inflammatory processes and RA fatigue is evidenced by findings demonstrating improvements in fatigue associated with disease-modifying therapies (Bingham et al., 2014; Druce, Jones, Macfarlane, & Basu, 2015; Keller et al., 2014; Strand, Burmester, et al., 2012).

Figure 2: Conceptual model of fatigue in RA



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 CDAI, with significant differences between groups with high disease activity, low/moderate disease activity, and in remission (Bartlett, Bykerk, et al., 2015). This is consistent with significantly higher fatigue scores among RA patients experiencing an inflammatory flare as compared to those not in flare (Bingham & Bartlett, 2016). 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 (Bingham & Bartlett, 2016). These findings are aligned with previous reports that, in spite of achieving clinical RA remission, many patients do not experience remission of their fatigue (Druce et al., 2015).

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 (Minnock & Bresnihan, 2004) 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 (Sanderson et al., 2010a, 2010b). 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 (van Tuyl et al., 2015). 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" (van Tuyl et al., 2015). 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, in magnitude of degree, 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 (Carr et al., 2003; Hewlett, Chalder, et al., 2011; Hewlett, Cockshott, et al., 2005) by describing the latter as overwhelming, difficult to resolve, and undertreated in clinical settings (Hewlett, Cockshott, et al., 2005).

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 (Bingham & Bartlett, 2016). 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 (Hewlett et al., 2012) 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 (Hewlett et al., 2012).

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 (Carr et al., 2003; Hewlett, Chalder, et al., 2011). 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 (Bingham & Bartlett, 2016).

1.3 Context of use

1.3.1 Identify the targeted study population, including a definition of the disease and selection criteria for clinical trials.

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 7 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.

1.3.2 Identify the targeted study design.

The targeted study design (i.e., trial design for the qualification plan) 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 (Felson et al., 1995), the Disease Activity Score for 28 joints [DAS28] (Prevoo et al., 1995)). 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 (Felson et al., 1995)) 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 endpoint measures in the trial.

Table 7: 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.

1.3.3 Identify the targeted study objectives and endpoint positioning

(i.e., planned set of primary and secondary endpoints with hierarchy)

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 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 to the degree known

1.4.1 Reporter

The ***PROMIS Fatigue Short Form 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.

1.4.2 Item content or description of the measure

The ***PROMIS Fatigue Short Form 10a*** is a 10-item fixed-length short form of the PROMIS Fatigue Item Bank assessing fatigue severity in terms of fatigue experience and impact. It has a recall period of past seven days and includes a 5-point verbal rating scale ranging from "Not at all" to "Very much." The content of the items reflects the following concepts: sleep during the day, usual activities (2 items), having energy, feeling tired, feeling fatigued, trouble starting things because of tiredness, trouble finishing things because of tiredness, frustrated by tiredness interfering with activities, and limiting social activity because of tiredness. Scores on the ***PROMIS Fatigue Short Form 10a*** are reported as a T-score metric (mean = 50 and SD = 10), with higher scores reflecting greater fatigue. For additional information about development of the ***PROMIS Fatigue Short Form 10a***, please see Section 1.1.5.2.

1.4.3 Mode of administration and data collection method

All of the items in the ***PROMIS Fatigue Short Form 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 Fatigue Short Form 10a***, including paper-and pencil; electronic data collection via tablet computer, smartphone, or similar device; and data capture via telephone using interactive voice response. The specific data collection method will be dependent upon future clinical trial methodology and researchers' preference; however, as referenced in Section 1.1.4, previous research has found no differential effect by method of data collection on items from the PROMIS Fatigue Item Bank (Bjorner et al., 2014).

1.5 Description of the involvement of external expertise, including scientific communities or other international regulatory agencies

As detailed in Section 1.1.3, the original PROMIS Fatigue Item Bank was developed by the PROMIS Fatigue working group, chaired by Dr. Stone from Stony Brook University at the time of development (Cella et al., 2010; Lai et al., 2014). The research was reviewed and approved by the PROMIS Steering Committee, composed of investigators from NIH as well as Duke University, Northwestern University, Stanford University, Stony Brook University, University of North Carolina – Chapel Hill, University of Pittsburgh, and University of Washington. In addition, the following external experts and agencies provided input and expertise for the development and validation of the PROMIS Fatigue Item Bank: FDA, OMERACT, WHO, and many consultants with considerable expertise in the measurement of fatigue and health.

With regard to the development of this Qualification Plan, 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: Brandon Becker, PhD (GlaxoSmithKline), Pam Berry, MSc (formerly of GlaxoSmithKline); Kate Burslem, MSc (Boehringer Ingelheim); Carol Gaich, PharmD, RPh (Eli Lilly and Company); Tristan Gloede, PhD (Boehringer Ingelheim); Kristina Harris, PhD (UCB Pharma); Christian Henke (Merck KGaA), Paul Kamudoni, PhD (Merck KGaA), April Naegeli,

DrPH, MPH (Co-Chair of the Working Group; Eli Lilly and Company); Enkeleida Nikai, MSc, MB (formerly Co-Chair of the Working Group; formerly of Eli Lilly and Company); and Josephine Park (GlaxoSmithKline).

- Other participants include: Susan J. Bartlett, PhD (McGill University); Clifton O. Bingham III, MD (Johns Hopkins University); David Cella, PhD (PROMIS, Northwestern University); Robert Chapman (Northwestern University); George J. Greene, PhD (Northwestern University); Kathryn Jackson, MS (Northwestern University); Sally Jensen, PhD (Northwestern University); San Keller, PhD (PROMIS, American Institutes for Research [AIR]); Amye Leong, MBA (Patient Representative); John Devin Peipert, PhD (Northwestern University); Lee S. Simon, MD (OMERACT); Vibeke Strand, MD (OMERACT); and James Witter, MD, PhD, FACR (PROMIS, NIH/NIAMS).

Additionally, patient representatives reviewed and provided detailed feedback on the format and content of the Qualification Plan, including input on the descriptions of how patients were engaged in RA research throughout the proposal, interpretation of research findings, and the barriers and facilitators to implementing the ***PROMIS Fatigue Short Form 10a*** in medical research settings.

Section 2: Executive Summary

The PRO Consortium's RA Working Group submitted a Letter of Intent for the qualification of the *PROMIS Fatigue Short Form 7a* for use in clinical trials with patients with RA. Subsequent to this submission and interactions with the FDA, the RA Working Group conducted a literature and measure review and determined that the ***PROMIS Fatigue Short Form 10a*** could be proposed for qualification given the evidence available for content validity and the widespread use of the parent measure, FACIT Fatigue, in clinical trials from which data could be reanalyzed to support the ***PROMIS Fatigue Short Form 10a***. The RA Working Group, in collaboration with Northwestern University, developed an Initial Briefing Package (IBP) for the qualification of the ***PROMIS Fatigue Short Form 10a*** for use in assessing treatment benefit in RA clinical trials. Following the submission of the IBP and subsequent feedback from the FDA, the RA Working Group is submitting this Qualification Plan for the ***PROMIS Fatigue Short Form 10a*** for use in clinical trials in patients with RA as the next step in the FDA's Drug Development Tool (DDT) process. This Qualification Plan follows the FDA's Clinical Outcome Assessment (COA) template posted publicly under Section 507 of the FD&C Act and summarizes the analyses to be conducted to generate evidence supporting the use of the ***PROMIS Fatigue Short Form 10a*** to address the need for a reliable, valid, and precise PRO measure of fatigue to support a secondary efficacy endpoint in RA clinical trials.

There is a growing interest in the broader inclusion of measures of fatigue in RA clinical trials, largely in response to the growing evidence highlighting the centrality of fatigue among RA symptoms. However, existing fatigue measures are limited by inadequate qualitative work with RA patients, insufficient evidence that the measures capture the full range of the fatigue experience in RA, variable quality in terms of psychometric properties, and lack of validation in RA samples. Given that PRO measures serve as the best method for assessing symptoms like fatigue that are only known to the patient, precise and valid measurement of fatigue is required to

fully evaluate the effects of RA interventions in clinical trials. The 10-item ***PROMIS Fatigue Short Form 10a*** is proposed as a COA in order to determine whether RA interventions can provide overall benefit to patients above and beyond the existing indicators for disease progression, symptom maintenance, and clinical remission. Specifically, ***PROMIS Fatigue Short Form 10a*** is proposed to be qualified for use in assessing fatigue severity via patient reported fatigue experience and impact as a secondary efficacy endpoint measure in RA treatment trials.

The ***PROMIS Fatigue Short Form 10a*** was developed based on qualitative work with RA patients. Kaiser and her colleagues (Kaiser et al., 2016) evaluated the content validity of the 13-item FACIT Fatigue in a sample of participants (n = 17) with moderately to highly active RA. Semi-structured interviews with participants inquired about the importance, experience, and impact of fatigue in daily living (i.e., concept elicitation). After completing the 13 items, participants completed cognitive interviews to assess 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. Analyses revealed that the fatigue items were relevant and important in the experience of fatigue among RA patients. However, Kaiser and colleagues (Kaiser et al., 2016) recommended the removal of 3 of the 13 fatigue items as they were not reflected in the responses from study participants, and highlighted that the 10 items exhibiting the strongest content validity should be retained to form the ***PROMIS Fatigue Short Form 10a***. In addition, qualitative analyses did not identify additional fatigue-related concepts that needed to be added to the ***PROMIS Fatigue Short Form 10a,*** further confirming sufficient coverage for participants with RA. Taken together, these qualitative results provide strong support for the coverage, relevance, and validity of 10 of the 13 FACIT Fatigue items in RA (Kaiser et al., 2016) leading to the development of the ***PROMIS Fatigue Short Form 10a***.

This Qualification Plan proposes analytic plans that will generate rigorous evidence in support of the use of the ***PROMIS Fatigue Short Form 10a*** as a PRO measure of fatigue severity in RA clinical trials. The primary data set that will be used in the analyses is from the Safety Trial of Adalimumab in Rheumatoid Arthritis (STAR trial) (Furst et al., 2003). These data consist of the 10 items included in the ***PROMIS Fatigue Short Form 10a*** (HI7, AN2, AN3, AN4, AN5, AN7, AN8, AN14, AN15, AN16). The items were collected at baseline, 12 weeks, and 24 weeks. At the item-level, the proposed descriptive statistics include frequency distribution of both item response and overall scores, floor and ceiling effects, and percentage of missing responses. To describe the distribution of items in the ***PROMIS Fatigue 10a***, we will calculate the frequency and percentage of each response option selected for each item, at baseline and 24-week administration time points. A threshold of 35% of the sample selecting either the highest or lowest response will be used to indicate a floor or ceiling effect for that item. Missing data at the item level will be evaluated by calculating the frequency and percentage of missing data per item at each time point. In addition, we will summarize the ***PROMIS Fatigue Short Form 10a*** Tscores from the STAR trial at baseline and 24 weeks by calculating the mean and standard deviation of the scores and graphing the density of the T-scores, comparing the distributions across time points. With regard to inter-item relationships and dimensionality analysis, we will use the STAR trial data to describe the relationship between individual instrument items by calculating inter-item Spearman's rank correlations between each pair of items. Unidimensionality for the ***PROMIS Fatigue Short Form 10a*** will be examined by calculating

McDonald's omega hierarchical and explained common variance (ECV) from an exploratory bifactor analysis.

The analysis plan includes reliability and validity assessments for the ***PROMIS Fatigue Short Form 10a***. To evaluate internal consistency, we will calculate the Cronbach's alpha coefficient using the STAR trial data. Since the STAR trial data set does not allow for test-retest analysis, we will rely on data acquired for the development and initial testing of the Functional Assessment of Cancer Therapy – Anemia (FACT-An) quality of life instrument (Yellen et al., 1997). For this analysis, we will calculate summary scores and assess the intraclass correlation coefficients (ICCs) between time points among this sample. To examine the association of the ***PROMIS Fatigue Short Form 10a*** and similar measures (convergent and discriminant validity), we will calculate the Pearson correlation between the ***PROMIS Fatigue Short Form 10a*** and the SF-36 Vitality and MAF at both baseline and 24-week assessments using STAR trial data. To examine if ***PROMIS Fatigue Short Form 10a*** scores differ between subgroups of subjects, we will segment the STAR trial population into groups based on their Patient Global Assessment of Disease Activity scores and compare mean ***PROMIS Fatigue Short Form 10a*** scores across these three groups using a one-way analysis of variance (ANOVA).

The analysis plan also describes how the ***PROMIS Fatigue Short Form 10a*** can be incorporated into clinical trials for longitudinal evaluation and describes the measure's ability to detect change, as well as the evaluation of individual patient change. In addition, details are provided on the administration and scoring of the measure, as well as PROMIS standards for translations, including information on the items in the ***PROMIS Fatigue Short Form 10a*** have been translated into 59 languages.

References:

Alten, R., Pohl, C., Choy, E. H., Christensen, R., Furst, D. E., Hewlett, S. E., . . . OMERACT RA Flare Definition Working Group. (2011). Developing a construct to evaluate flares in rheumatoid arthritis: a conceptual report of the OMERACT RA Flare Definition Working Group. *Journal of Rheumatology, 38*(8), 1745-1750. doi:10.3899/jrheum.110400

Arnett, F. C., Edworthy, S. M., Bloch, D. A., Mcshane, D. J., Fries, J. F., Cooper, N. S., . . . Hunder, G. G. (1987). The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis & Rheumatology, 31*, 315-324.

Bartlett, S. J., Bykerk, V. P., Cooksey, R., Choy, E. H., Alten, R., Christensen, R., . . . Bingham, C. O., 3rd. (2015). Feasibility and Domain Validation of Rheumatoid Arthritis (RA) Flare Core Domain Set: Report of the OMERACT 2014 RA Flare Group Plenary. *Journal of Rheumatology, 42*(11), 2185-2189. doi:10.3899/jrheum.141169

Bartlett, S. J., Hewlett, S., Bingham, C. O., 3rd, Woodworth, T. G., Alten, R., Pohl, C., . . . OMERACT RA Flare Working Group. (2012). Identifying core domains to assess flare in rheumatoid arthritis: an OMERACT international patient and provider combined Delphi consensus. *Ann Rheum Dis, 71*(11), 1855-1860. doi:10.1136/annrheumdis-2011-201201

Bartlett, S. J., Orbai, A. M., Duncan, T., DeLeon, E., Ruffing, V., Clegg-Smith, K., & Bingham, C. O., 3rd. (2015). Reliability and Validity of Selected PROMIS Measures in People with Rheumatoid Arthritis. *PLoS One, 10*(9), e0138543. doi:10.1371/journal.pone.0138543

Bechman, K., Sin, F. E., Ibrahim, F., Norton, S., Scott, D., Cope, A., & Galloway, J. (2018). Psychological and Functional States Predict Disease Flare Following Tnf Inhibitor Tapering in Patients with Rheumatoid Arthritis: A Post-Hoc Analysis of Data from the Optimisingtnf Tapering in Ra (Opttira) Cohort. *Annals of the Rheumatic Diseases, 77*, 279-280. doi:10.1136/annrheumdis-2018-eular.3077

Berthelot, J. M., De Bandt, M., Morel, J., Benatig, F., Constantin, A., Gaudin, P., . . . STPR Group of French Society of Rheumatology. (2012). A tool to identify recent or present rheumatoid arthritis flare from both patient and physician perspectives: the 'FLARE' instrument. *Annals of the Rheumatic Diseases, 71*(7), 1110-1116. doi:10.1136/ard.2011.150656

Bingham, C. O., 3rd, Alten, R., Bartlett, S. J., Bykerk, V. P., Brooks, P. M., Choy, E., . . . OMERACT RA Flare Definition Working Group. (2011). Identifying preliminary domains to detect and measure rheumatoid arthritis flares: report of the OMERACT 10 RA Flare Workshop. *Journal of Rheumatology, 38*(8), 1751-1758. doi:10.3899/jrheum.110401

Bingham, C. O., 3rd, Alten, R., & de Wit, M. P. (2012). The importance of patient participation in measuring rheumatoid arthritis flares. *Annals of the Rheumatic Diseases, 71*(7), 1107-1109. doi:10.1136/annrheumdis-2011-200870

Bingham, C. O., 3rd, & Bartlett, S. J. (2016). *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*. Retrieved from

Bingham, C. O., 3rd, Weinblatt, M., Han, C., Gathany, T. A., Kim, L., Lo, K. H., . . . Westhovens, R. (2014). 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, 41*(6), 1067-1076. doi:10.3899/jrheum.130864

Bjorner, J. B., Rose, M., Gandek, B., Stone, A. A., Junghaenel, D. U., & Ware, J. E. (2014). Method of administration of PROMIS scales did not significantly impact score level, reliability, or validity. *Journal of Clinical Epidemiology, 67*(1), 108-113. doi:10.1016/j.jclinepi.2013.07.016

Campbell, R. C. J., Batley, M., Hammond, A., Ibrahim, F., Kingsley, G., & Scott, D. L. (2012). The impact of disease activity, pain, disability and treatments on fatigue in established rheumatoid arthritis. *Clinical Rheumatology, 31*(4), 717-722. doi:10.1007/s10067-011-1887-y

Carr, A., Hewlett, S., Hughes, R., Mitchell, H., Ryan, S., Carr, M., & Kirwan, J. (2003). Rheumatology outcomes: the patient's perspective. *Journal of Rheumatology, 30*(4), 880-883.

Cella, D., Lai, J., & Stone, A. A. (2011). Self-reported fatigue: One dimension or more? Lessons from the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) questionnaire. *Supportive Care in Cancer, 19*(9), 1441-1450. doi:10.1007/s00520-010-0971-1

Cella, D., Lai, J. S., Jensen, S. E., Christodoulou, C., Junghaenel, D. U., Reeve, B. B., & Stone, A. A. (2016). PROMIS Fatigue Item Bank had Clinical Validity across Diverse Chronic Conditions. *Journal of Clinical Epidemiology, 73*, 128-134. doi:10.1016/j.jclinepi.2015.08.037

Cella, D., Riley, W., Stone, A. A., Rothrock, N., Reeve, B., Yount, S., . . . PROMIS Cooperative Group. (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. doi:10.1016/j.jclinepi.2010.04.011

Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., . . . PROMIS Cooperative Group. (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Medical Care, 45*(5 Suppl 1), S3-S11. doi:10.1097/01.mlr.0000258615.42478.55

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. *Journal of Rheumatology, 32*(5), 811-819.

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 patientreported outcomes measurement information system (PROMIS). *Quality of Life Research, 17*(10), 1239-1246. doi:10.1007/s11136-008-9402-x

Cohen, S. B., Emery, P., Greenwald, M. W., Dougados, M., Furie, R. A., Genovese, M. C., . . . Grp, R. T. (2006). Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy - Results of a multicenter, randomized, double-blind, placebo-controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks. *Arthritis and rheumatism, 54*(9), 2793-2806. doi:10.1002/art.22025

Critical Path Institute PRO Consortium RA Working Group. (2012). *Toward Consensus Development: Qualifying Endpoint Measures for Rheumatoid Arthritis Clinical Trials*. Retrieved from https://c-path.org/towards-consensus-development-qualifying-endpointmeasures-for-rheumatoid-arthritis-clinical-trails/

Davis, M. C., Zautra, A. J., Younger, J., Motivala, S. J., Attrep, J., & Irwin, M. R. (2008). Chronic stress and regulation of cellular markers of inflammation in rheumatoid arthritis: implications for fatigue. *Brain, Behavior & Immunity, 22*(1), 24-32. doi:10.1016/j.bbi.2007.06.013

Dekkers, J. C., Geenen, R., Godaert, G. L. R., van Doornen, L. J. P., & Bijlsma, J. W. J. (2000). Diurnal courses of cortisol, pain, fatigue, negative mood, and stiffness in patients with recently diagnosed rheumatoid arthritis. *International Journal of Behavioral Medicine, 7*(4), 353-371. doi:Doi 10.1207/S15327558ijbm0704_06

DeWalt, D. A., Rothrock, N., Yount, S., Stone, A. A., & PROMIS Cooperative Group. (2007). Evaluation of item candidates: the PROMIS qualitative item review. *Medical Care, 45*(5 Suppl 1), S12-21. doi:10.1097/01.mlr.0000254567.79743.e2

Druce, K. L., Jones, G. T., Macfarlane, G. J., & Basu, N. (2015). 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), 54*(6), 964-971. doi:10.1093/rheumatology/keu390

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research, 16 Suppl 1*(S1), 5-18. doi:10.1007/s11136-007-9198-0

Emery, P., Blanco, R., Cocco, J. M., Chen, Y. C., Gaich, C. L., DeLozier, A. M., . . . Dougados, M. (2017). Patient-reported outcomes from a phase III study of baricitinib in patients with conventional synthetic DMARD-refractory rheumatoid arthritis. *Rmd Open, 3*(1). doi:UNSP e000410 10.1136/rmdopen-2016-000410

Felson, D. T., Anderson, J. J., Boers, M., Bombardier, C., Chernoff, M., Fried, B., . . . Wolfe, F. (1995). 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, 36*(6), 729–740. doi:10.1002/art.1780360601

Furst, D. E., Schiff, M. H., Fleischmann, R. M., Strand, V., Birbara, C. A., Compagnone, D., . . . Chartash, E. K. (2003). 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, 30*(12), 2563-2571.

Gossec, L., Dougados, M., Rincheval, N., Balanescu, A., Boumpas, D. T., Canadelo, S., . . . Kvien, T. K. (2009). Elaboration of the preliminary Rheumatoid Arthritis Impact of Disease (RAID) score: a EULAR initiative. *Annals of the Rheumatic Diseases, 68*(11), 1680-1685. doi:10.1136/ard.2008.100271

Gossec, L., Paternotte, S., Aanerud, G. J., Balanescu, A., Boumpas, D. T., Carmona, L., . . .

Kvien, T. K. (2011). Finalisation and validation of the rheumatoid arthritis impact of disease score, a patient-derived composite measure of impact of rheumatoid arthritis: a EULAR initiative. *Annals of the Rheumatic Diseases, 70*(6), 935-942. doi:10.1136/ard.2010.142901

Helmick, C. G., Felson, D. T., Lawrence, R. C., Gabriel, S., Hirsch, R., Kwoh, C. K., . . . National Arthritis Data Workgroup. (2008). Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. *Arthritis & Rheumatism, 58*(1), 15-25. doi:10.1002/art.23177

Hewlett, S., Carr, M., Ryan, S., Kirwan, J., Richards, P., Carr, A., & Hughes, R. (2005). Outcomes generated by patients with rheumatoid arthritis: how important are they? *Musculoskeletal Care, 3*, 131-142. doi:10.1002/msc.3

Hewlett, S., Chalder, T., Choy, E., Cramp, F., Davis, B., Dures, E., . . . Kirwan, J. (2011). Fatigue in rheumatoid arthritis: time for a conceptual model. *Rheumatology (Oxford), 50*(6), 1004-1006. doi:10.1093/rheumatology/keq282

Hewlett, S., Cockshott, Z., Byron, M., Kitchen, K., Tipler, S., Pope, D., & Hehir, M. (2005). Patients' perceptions of fatigue in rheumatoid arthritis: overwhelming, uncontrollable, ignored. *Arthritis & Rheumatism, 53*(5), 697-702. doi:10.1002/art.21450

Hewlett, S., Dures, E., & Almeida, C. (2011). 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 (Hoboken), 63 Suppl 11*, S263-286. doi:10.1002/acr.20579

Hewlett, S., Hehir, M., & Kirwan, J. R. (2007). Measuring fatigue in rheumatoid arthritis: a systematic review of scales in use. *Arthritis & Rheumatism, 57*(3), 429-439. doi:10.1002/art.22611

Hewlett, S., Sanderson, T., May, J., Alten, R., Bingham, C. O., 3rd, Cross, M., . . . Bartlett, S. J. (2012). '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), 51*(1), 69-76. doi:10.1093/rheumatology/keq455

Hootman, J. M., & Helmick, C. G. (2006). Projections of US prevalence of arthritis and associated activity limitations. *Arthritis & Rheumatism, 54*(1), 226-229. doi:10.1002/art.21562

Huyser, B. A., Parker, J. C., Thoreson, R., Smarr, K. L., Johnson, J. C., & Hoffman, R. (1998).

Predictors of subjective fatigue among individuals with rheumatoid arthritis. *Arthritis and rheumatism, 41*(12), 2230-2237. doi:Doi 10.1002/1529-0131(199812)41:12<2230::Aid-Art19>3.0.Co;2-D

Janoudi, N., Omran, N., Hussain, W., Al-osaimi, H., Baamer, M., Siddiqui, M. I., . . . Almoallim, H. (2017). Adalimumab impact on fatigue measured by FACIT-F: a study of rheumatoid arthritis patients in Saudi Arabia. *Open Access Rheumatology-Research and Reviews, 9*, 181-184. doi:10.2147/Oarrr.S142411

Junghaenel, D. U., Christodoulou, C., Lai, J. S., & Stone, A. A. (2011). Demographic correlates of fatigue in the US general population: results from the patient-reported outcomes measurement information system (PROMIS) initiative. *Journal of Psychosomatic Research, 71*(3), 117-123. doi:10.1016/j.jpsychores.2011.04.007

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

Keller, S., Mangrum, R., & Yang, M. (2014). *The measurement of fatigue in rheumatoid arthritis: review and identification of gaps in evidence*. Retrieved from American Institutes for Research (AIR) http://www.nihpromis.com/Documents/FatigueMeasureRA2014_ES_508.pdf?AspxAutoDetectCookieSupport=1

Keystone, E. C., Taylor, P. C., Tanaka, Y., Gaich, C., DeLozier, A. M., Dudek, A., . . . Weinblatt, M. E. (2017). Patient-reported outcomes from a phase 3 study of baricitinib versus placebo or adalimumab in rheumatoid arthritis: secondary analyses from the RABEAM study. *Annals of the Rheumatic Diseases, 76*(11), 1853-1861. doi:10.1136/annrheumdis-2017-211259

Kirwan, J. R., & Hewlett, S. (2007). Patient perspective: reasons and methods for measuring fatigue in rheumatoid arthritis. *Journal of Rheumatology, 34*(5), 1171-1173.

Kirwan, J. R., Newman, S., Tugwell, P. S., et al., Wells, G., Hewlett, S., Idzera, L., . . . Welch, V. (2009). Progress on incorporating the patient perspective in outcome assessment in rheumatology and the emergence of life impact measures at OMERACT 9. *Journal of Rheumatology, 36*, 2071-2076. doi:10.3899/jrheum.090360

Lai, J. S., Cella, D., Choi, S. W., Junghaenel, D. U., Christodolou, C., Gershon, R., & Stone, A. A. (2011). How item banks and their application can influence measurement practice in rehabilitation medicine: a PROMIS fatigue item bank example. *Archives of Physical Medicine & Rehabilitation, 92*(s20e7). doi:10.1016/j.apmr.2010.08.033

Lai, J. S., Cella, D., Yanez, B., & Stone, A. A. (2014). Linking fatigue measures on a common reporting metric. *Journal of Pain and Symptom Management, 48*(4), 639-648. doi:10.1016/j.jpainsymman.2013.12.236

Lawrence, R. C., Felson, D. T., Helmick, C. G., Arnold, L. M., Choi, H., Deyo, R. A., . . . National Arthritis Data Workgroup. (2008). Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. *Arthritis & Rheumatism, 58*(1), 26-35. doi:10.1002/art.23176

Mayoux-Benhamou, A., Quintrec, J. S. G. L., Ravaud, P., Champion, K., Dernis, E., Zerkak, D., . . . Dougados, M. (2008). Influence of patient education on exercise compliance in RA: A prospective 12-month randomized controlled trial. *Journal of Rheumatology, 35*(2), 216-223.

Mease, P. J., Revicki, D. A., Szechinski, J., Greenwald, M., Kivitz, A., Barile-Fabris, L., . . . Leirisalo-Repo, M. (2008). Improved health-related quality of life for patients with active rheumatoid arthritis receiving rituximab - Results of the dose-ranging assessment: International clinical evaluation of rituximab in rheumatoid arthritis (DANCER) trial. *Journal of Rheumatology, 35*(1), 20-30.

Minnock, P., & Bresnihan, B. (2004). Pain outcomes and fatigue levels reported by women with established rheumatoid arthritis. *Arthritis & Rheumatolgy, 50*, S471.

Minnock, P., Kirwan, J., & Bresnihan, B. (2009). Fatigue is a reliable, sensitive and unique outcome measure in rheumatoid arthritis. *Rheumatology (Oxford), 48*(12), 1533-1536. doi:10.1093/rheumatology/kep287

Nicklin, J., Cramp, F., Kirwan, J., Urban, M., & Hewlett, S. (2010). Collaboration with patients in the design of patient-reported outcome measures: capturing the experience of fatigue in rheumatoid arthritis. *Arthritis Care & Research (Hoboken), 62*(11), 1552-1558. doi:10.1002/acr.20264

Orbai, A. M., Halls, S., Hewlett, S., Bartlett, S. J., Leong, A. L., Bingham, C. O., 3rd, & OMERACT RA Flare Group Steering Committee. (2015). More than Just Minutes of Stiffness in the Morning: Report from the OMERACT Rheumatoid Arthritis Flare Group Stiffness Breakout Sessions. *Journal of Rheumatology, 42*(11), 2182-2184. doi:10.3899/jrheum.141172

Orbai, A. M., Smith, K. C., Bartlett, S. J., De Leon, E., & Bingham, C. O., 3rd. (2014). "Stiffness has different meanings, I think, to everyone": examining stiffness from the perspective of people living with rheumatoid arthritis. *Arthritis Care & Research (Hoboken), 66*(11), 1662-1672. doi:10.1002/acr.22374

Piper, B. F. (1989). *Fatigue: current bases for practice*. New York: Spring Publishing.

Pollard, L. C., Choy, E. H., Gonzalez, J., Khoshaba, B., & Scott, D. L. (2006). Fatigue in rheumatoid arthritis reflects pain, not disease activity. *Rheumatology (Oxford), 45*(7), 885-889. doi:10.1093/rheumatology/kel021

Reeve, B., & Fayers, P. M. (2005). *Applying item response theory modeling for evaluating questionnaire item and scale properties*. New York, NY: Oxford University Press.

Reeve, B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., . . . PROMIS Cooperative Group. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care, 45*(5 Suppl 1), S22-31. doi:10.1097/01.mlr.0000250483.85507.04

Riemsma, R. P., Rasker, J. J., Taal, E., Griep, E. N., Wouters, J. M., & Wiegman, O. (1998). Fatigue in rheumatoid arthritis: the role of self-efficacy and problematic social support. *British Journal of Rheumatology, 37*(10), 1042-1046.

Samejima, F. (1969). Estimation of Latent Ability Using a Response Pattern of Graded Scores. *Psychometrika, 34*(4p2), 1-&.

Sanderson, T., Hewlett, S. E., Flurey, C., Dures, E., Richards, P., & Kirwan, J. R. (2011). The impact triad (severity, importance, self-management) as a method of enhancing measurement of personal life impact of rheumatic diseases. *Journal of Rheumatology, 38*(2), 191-194. doi:10.3899/jrheum.100700

Sanderson, T., & Kirwan, J. (2009). Patient-reported outcomes for arthritis: time to focus on personal life impact measures? *Arthritis & Rheumatism, 61*(1), 1-3. doi:10.1002/art.24270

Sanderson, T., Morris, M., Calnan, M., Richards, P., & Hewlett, S. (2010a). Patient perspective of measuring treatment efficacy: the rheumatoid arthritis patient priorities for pharmacologic interventions outcomes. *Arthritis Care & Resesearch (Hoboken), 62*(5), 647-656. doi:10.1002/acr.20151

Sanderson, T., Morris, M., Calnan, M., Richards, P., & Hewlett, S. (2010b). What outcomes from pharmacologic treatments are important to people with rheumatoid arthritis? Creating the basis of a patient core set. *Arthritis Care & Research (Hoboken), 62*(5), 640-646. doi:10.1002/acr.20034

Singh, J. A., Furst, D. E., Bharat, A., Curtis, J. R., Kavanaugh, A. F., Kremer, J. M., . . . Saag, K. G. (2012). 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 (Hoboken), 64*(5), 625-639. doi:10.1002/acr.21641

Smolen, J. S., Aletaha, D., Bijlsma, J. W., Breedveld, F. C., Boumpas, D., Burmester, G., . . . T. T. Expert Committee. (2010). Treating rheumatoid arthritis to target: recommendations of an international task force. *Annals of the Rheumatic Diseases, 69*(4), 631-637. doi:10.1136/ard.2009.123919

Strand, V., Burmester, G. R., Ogale, S., Devenport, J., John, A., & Emery, P. (2012). 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), 51*(10), 1860-

1869. doi:10.1093/rheumatology/kes131

Strand, V., Kosinski, M., Chen, C. I., Joseph, G., Rendas-Baum, R., Graham, N. M. H., . . . Genovese, M. C. (2016). Sarilumab plus methotrexate improves patient-reported outcomes in patients with active rheumatoid arthritis and inadequate responses to methotrexate: results of a phase III trial. *Arthritis Research & Therapy, 18*. doi:ARTN 198 10.1186/s13075-016-1096-9

Strand, V., Rentz, A. M., Cifaldi, M. A., Chen, N. J., Roy, S., & Revicki, D. (2012). Healthrelated Quality of Life Outcomes of Adalimumab for Patients with Early Rheumatoid Arthritis: Results from a Randomized Multicenter Study. *Journal of Rheumatology, 39*(1), 63-72. doi:10.3899/jrheum.101161

Strand, V., Simon, L. S., Bingham, C. O., 3rd, & Hewlett, S. (2014). *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*. Retrieved from

Swain, M. (2000). Fatigue in chronic disease. *Clinical Science, 9*, 1-8.

Ter Wee, M. M., van Tuyl, L. H., Blomjous, B. S., Lems, W. F., Boers, M., & Terwee, C. B. (2016). 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, 18*, 22. doi:10.1186/s13075-015-0911-z

van Tuyl, L., Hewlett, S., Sadlonova, M., Davis, B., Flurey, C., Hoogland, W., . . . Boers, M. (2015). 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, 74*(6), 1004-1010. doi:10.1136/annrheumdis-2013-204798

van Tuyl, L., Sadlonova, M., Davis, B., Flurey, C., Goel, N., Hewlett, S. E., . . . Boers, M. (2016). Remission in Rheumatoid Arthritis: Working Toward Incorporation of the Patient Perspective at OMERACT 12. *Journal of Rheumatology, 43*(1), 203-207. doi:10.3899/jrheum.141113

van Tuyl, L., Smolen, J. S., Wells, G. A., Scholte-Voshaar, M., Hoogland, W., & Boers, M. (2011). Patient perspective on remission in rheumatoid arthritis. *Journal of Rheumatology, 38*(8), 1735-1738. doi:10.3899/jrheum.110394

Willis, G. B., & Lessler, J. T. (1999). *Question appraisal system: QAS-99*. Retrieved from

World Health Organization. (2001). *International Classification of Functioning, Disability and Health (ICF)*. Retrieved from Geneva, Switzerland:

Wolfe, F., Hawley, D. J., & Wilson, K. (1996). The prevalence and meaning of fatigue in rheumatic disease. *J Rheumatol, 23*(8), 1407-1417.

Yellen, S. B., Cella, D. F., Webster, K., Blendowski, C., & Kaplan, E. (1997). Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. *Journal of Pain and Symptom Management, 13*(2), 63-74.

Zonna-Nacach, A., Roseman, J. M., McGwin, G. J., Friedman, A. W., Baethge, B. A., Reveille, J. D., & Alarcon, G. S. (2000). Systemic lupus erythematosus in three ethnic groups. VI. Factors associated with fatigue within 5 years of criteria diagnosis. *Lupus, 9*, 101-109