USER MANUAL AND SCORING INSTRUCTIONS

PROMIS FATIGUE

Contents

INTRODUCTION

This **User Manual and Scoring Instructions** includes information on measure development, administration, scoring, and interpretation for the measures listed below. For information on distinguishing one measure from another, look to the Measure Differences summary, which can be found with additional information at www.HealthMeasures.net.

The **User Manual and Scoring Instructions** describes fatigue measures developed for universal use. Information specific to the Patient-Reported Outcomes Measurement Information System Fatigue Measure for Multiple Sclerosis (MS) (PROMIS Short Form v1.0—Fatigue-Multiple Sclerosis 8a), i.e., **PROMIS FatigueMS—8a** is included in Appendix 4. Information specific to the PROMIS Short Form v1.0 – Fatigue 10a (**FACIT-Fatigue-10/PROMIS Fatigue Short Form 10a**) is included in Appendix 5.

PROMIS FATIGUE MEASURES

ADULT	PEDIATRIC	PARENT PROXY
PROMIS Item Bank v1.0 – Fatigue PROMIS Item Bank v1.0 – Fatigue (recommended) PROMIS Item Bank v1.0 – Fatigue (screen-to-CAT) PROMIS Short Form v1.0 – Fatigue 4a PROMIS Short Form v1.0 – Fatigue 6a PROMIS Short Form v1.0 – Fatigue 7a PROMIS Short Form v1.0 – Fatigue 7b Daily PROMIS Short Form v1.0 – Fatigue 8a PROMIS Short Form v1.0 – Fatigue 10a (FACIT-Fatigue-10) PROMIS Short Form v1.0 – Fatigue 13a (FACIT-Fatigue) PROMIS Short Form v1.0 – Fatigue-OA-Knee 8a PROMIS – Ca Bank v1.0 Fatigue PROMIS Short Form v1.0 – Fatigue – Multiple Sclerosis 8a	PROMIS Pediatric Bank GenPop v3.0 – Fatigue PROMIS Pediatric Bank GenPop v3.0 – Fatigue (screen-to-CAT) PROMIS Pediatric Short Form GenPop v3.0 – Fatigue 10a PROMIS Pediatric Bank v2.0 – Fatigue* PROMIS Pediatric Bank v2.0 – Fatigue (recommended)* PROMIS Pediatric Bank v2.0 – Fatigue (screen-to-CAT)* PROMIS Pediatric Short Form v2.0 – Fatigue 10a* PROMIS Pediatric Bank v1.0 – Fatigue* PROMIS Pediatric Short Form v1.0 – Fatigue 10a*	PROMIS Parent Proxy Bank GenPop v3.0 – Fatigue PROMIS Parent Proxy Bank GenPop v3.0 – Fatigue (screen-to-CAT) PROMIS Parent Proxy Short Form GenPop v3.0 – Fatigue 10a PROMIS Parent Proxy Bank v2.0 – Fatigue* PROMIS Parent Proxy Bank v2.0 – Fatigue (recommended)* PROMIS Parent Proxy Bank v2.0 – Fatigue (screen-to-CAT)* PROMIS Parent Proxy Short Form v2.0 – Fatigue 10a* PROMIS Parent Proxy Bank v1.0 – Fatigue* PROMIS Parent Proxy Short Form v1.0 – Fatigue 10a*

*Retired measure
Ca=cancer, OA=osteoarthritis

DESCRIPTION OF FATIGUE AS MEASURED BY PROMIS

The PROMIS Fatigue item banks 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 short forms are universal rather than disease-specific. All assess fatigue over the past seven days.

Fatigue instruments are available for adults (ages 18+), pediatric self-report (ages 8-17) and for parents serving as proxy reporters for their child (youth ages 5-17).

OBTAINING MEASURES

All English and some Spanish PROMIS measures are available to download in respondent-ready PDF format at Search & View Measures on HealthMeasures.net. These measures may be used in **paper administration** without permission or fees as long as the text is not changed.

PROMIS measures are also available in many **digital administration platforms** in patient-reported outcome measure libraries (e.g., REDCap shared library, Epic PROMIS CAT application). These platforms use the Assessment Center API or have a license agreement for distribution of PROMIS measures.

Non-commercial users[1] can program a PROMIS measure into a digital administration platform for research[2] without permission or fees as long as the text is not changed. Programming PROMIS measures into digital administration platforms for other purposes (e.g., registries, routine outcome assessment, quality measurement) and any use by commercial users[3] requires a permission letter and screenshot review (HealthMeasures Electronic Administration Permission or HEAP). The screenshot review ensures the digital implementation retains validity. Contact Help@HealthMeasures.net for permission and to learn more about HEAP.

PROMIS measures are also available in **additional languages**. Information on currently available translations can be found at HealthMeasures.net under Available Translations. Contact translations@HealthMeasures.net to request a translated measure or to learn more.

INTRODUCTION TO ASSESSMENT OPTIONS

There are two administration options for assessing Fatigue: short forms and computer adaptive tests (CATs). When administering a short form, instruct participants to answer all of the items (i.e., questions or statements) presented. With a CAT, participant responses guide the system's choice of subsequent items from the full item bank (165 items in total in adult bank). Although items differ across respondents taking CAT, scores are comparable across participants.

Some administrators may prefer to ask the same question of all respondents or of the same respondent over time, to enable a more direct comparability across people or time. In these cases, or when paper administration is preferred, a short form would be more desirable than CAT. This guide provides information on all Fatigue short form and CAT instruments.

CAT

Stopping rules applied to the CAT assessment determine the number of items required to receive a score for Fatigue. The response to the first item will guide the system's choice of the next item for the participant. The participant's response to the second item will dictate the selection of the following question, and so on. As additional items are administered, the potential for error is reduced and confidence in the respondent's score increases. CAT will continue until either the standard error drops below a specified level (e.g., on the T-score metric 3.0), or the participant has answered the maximum number of questions (e.g., 12), whichever occurs first. For some CATs, specifically "recommended" and "screen-to-CAT" there are additional stopping rules. These include stopping when the standard error isn't improving much or if a respondent is asymptomatic. For details on the exact stopping rules for Fatigue CATs, see below.

[1] A non-commercial user is any institution, organization, or other incorporated entity which is defined as a Tax-exempt Organization under the regulations of the United States' Internal Revenue Service; for users outside the United States, all not-for-profit entities including national health services and their affiliate providers and academic institutions.

[2] Research includes randomized controlled trials, cohort studies, case control studies, and cross-sectional studies.

[3] A commercial user is any company, organization, or other incorporated limited liability entity which does not fit the definition of a Tax-exempt Organization under the regulations of the United States' Internal Revenue Service or which otherwise operates under the rubric of financial profit.

CAT versus Short Form

Whether one uses a short form or CAT, the score metric is Item Response Theory (IRT), a family of statistical models that link individual questions to a presumed underlying trait or concept of fatigue represented by all items in the item bank. When choosing between CAT and a short form, it is useful to consider the demands of computer-based assessment, and the psychological, physical, and cognitive burden placed on respondents as a result of the number of questions asked.

Figure 1 illustrates the correlations (strength of relationship) of the full PROMIS Fatigue Item Bank v1.0 with CAT and with v1.0 short forms of varying length. The correlation of CAT scores with the full bank score is greater than a short form of any length. A longer CAT or longer short form offers greater correlation, as well as greater precision. When evaluating precision, not all questions are equally informative. The flexibility of CAT to choose more informative questions offers more precision.

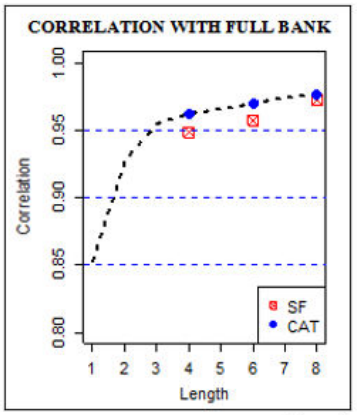


Figure 1

RECOMMENDED SHORT FORMS

PROMIS includes multiple short form options. Here are recommendations for which short form to use.

Comparing Large Groups

These short forms are appropriate for large scale data collection and comparing groups (subgroups larger than 50-75 people).

- **Adult**: PROMIS Short Form v1.0 – Fatigue 7a
- **Pediatric**: PROMIS Pediatric Short Form GenPop v3.0 – Fatigue 10a
- **Parent Proxy**: PROMIS Parent Proxy Short Form GenPop v3.0 – Fatigue 10a

Individual Evaluations and Comparing Small Groups

More detailed or precise assessment can be achieved by administering a slightly longer short form or additional short forms. These short forms are recommended when looking at an individual or making comparisons between smaller groups (subgroups less than 50-75 people).

- **Adult**: PROMIS Short Form v1.0 – Fatigue 13a (FACIT-Fatigue)
- **Pediatric**: PROMIS Pediatric Short Form GenPop v3.0 – Fatigue 10a
- **Parent Proxy**: PROMIS Parent Proxy Short Form GenPop v3.0 – Fatigue 10a

DIFFERENCES BETWEEN MEASURES

Detailed information about the differences between short forms, version numbers, calibration samples, pediatric versus parent proxy, and more is housed in the Measure Differences summary.

[Remainder of page intentionally left blank]

MEASURE DEVELOPMENT

PROMIS uses rigorous methodology for developing its measures, also known as item banks, and testing their validity. (Similar approaches were used for developing PROMIS FatigueMS—8a [Appendix 4] and FACIT-Fatigue-10/PROMIS Fatigue Short Form 10a [Appendix 5]).

- Comprehensive literature searches (e.g., Klem et al., 2009) of existing measures, yielding hundreds of potential items, are performed to ensure content validity (i.e., the assurance that each measure represents all facets of a domain).
- To ensure comprehensive coverage of the conceptual area, focus groups are conducted with relevant participants and thematic analyses are performed of the topics discussed. Following the focus groups, an initial item-review process is completed involving elimination of items that are redundant, confusing, or poorly written (DeWalt et al., 2007).
- Cognitive interviews are also performed so that each candidate item is reviewed by multiple individuals with diverse characteristics (i.e., both genders, members of minority groups, participants with modest reading levels) for feedback on the language and clarity of items and the relevance of the content (DeWalt et al., 2007; Irwin et al., 2009).
- Responses to candidate items are collected from relevant participants, usually via computer administration to both community and clinical samples, during the psychometric testing stage of item bank development (Cella et al., 2010). Data from large samples confirm the factor structure of the domains and allow for analyses at the item and bank level. This approach is described by Reeve et al. (2007) with an update from Hansen et al. (2014). Techniques from both classical test theory (CTT) and item response theory (IRT) were used.
- Validity studies are conducted to determine the degree to which an instrument measures what it is intended to measure. This includes:
    - Cross-sectional validity (e.g., Cella et al., 2010)
    - Responsiveness to change (e.g., Forrest et al., 2019; Pilkonis et al., 2014; Wohlfahrt et al., 2019)
    - Clinical validity (see 2016 Journal of Clinical Epidemiology series with a summary by Cook et al.)
- Translations result from a process of forward and back-translation, multiple expert reviews, harmonization across languages, and cognitive debriefing with a sample of native speakers of the target language (linguistic validation). A universal approach to translation ensures that, whenever possible, one language version is created for multiple countries instead of country-specific versions of the same language (Eremenco et al., 2005; Eremenco et al., 2017).

[Remainder of page intentionally left blank]

Primary Citations

Adult	Fatigue Item Bank v1.0	Lai, J.S., Cella, D., Choi, S.W., Junghaenel, D.U., Christodoulou, C., Gershon, R., & Stone, 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 and Rehabilitation, 92(10 Supplement), S20-S27.
	v1.0 Daily	Schneider S, Choi SW, Junghaenel DU, Schwartz JE, Stone AA. Psychometric characteristics of daily diaries for the Patient-Reported Outcomes Measurement Information System (PROMIS®): a preliminary investigation. Quality of Life Research 2013; 22.
Pediatric	Fatigue Item Bank v1.0	Lai, J.-S., Stucky, B., Thissen, D., Varni, J., DeWitt, E., Irwin, D., et al. (2013). Development and Psychometric Properties of the PROMIS® Pediatric Fatigue Item Banks. Quality of Life Research, 22(9), 2417-2427.
	Fatigue Item Bank v2.0	Quinn H, Thissen D, Liu Y, Magnus B, Lai JS, Amtmann D, Varni JW, Gross HE, DeWalt DA. (2014) Using item response theory to enrich and expand the PROMIS® pediatric self-report banks. Health and Quality of Life Outcomes. 2014 Oct 25;12: 160.
	Fatigue Item Bank v3.0	Lai, J. S., Tang, X., Schalet, B. D., Kallen, M. A., Kaat, A. J., & Cella, D. (2023). The PROMIS Pediatric Item Banks Norming Project. *Journal of Patient-Reported Outcomes*, 7(1): O84.
Parent Proxy	Fatigue Item Bank v1.0 – v2.0	Irwin, D., Gross, H., Stucky, B., Thissen, D., Morgan DeWitt, E., Lai, J.S., et al. (2012). Development of Six PROMIS Pediatrics Proxy-Report Item Banks. Health and Quality of Life Outcomes, 10(1), 22.
	Fatigue Item Bank v3.0	Lai, J. S., Tang, X., Schalet, B. D., Kallen, M. A., Kaat, A. J., & Cella, D. (2023). The PROMIS Pediatric Item Banks Norming Project. *Journal of Patient-Reported Outcomes*, 7(1): O84.

BEST PRACTICES IN ADMINISTRATION

- PROMIS self-report measures are intended to be completed by the respondent without help from anyone else.
- If respondents are unable to answer on their own, have someone else ("proxy") report on their behalf. Respondents requiring a proxy may include: young children, people in the early stages of dementia who may not recognize the extent of their impairment, people with cognitive or communication deficits, and people with severe disease burden. PROMIS includes measures for parents to report for young children (ages 1-5) and older children (ages 5-17). Note, however, that proxy respondents are not acceptable for measures used to support endpoints in clinical trials.
- Use the same proxy across multiple assessments. Different proxies may have different perspectives.

- Keep respondents' privacy in mind, but have staff readily available to help with any technology issues that may arise.
- It is acceptable for staff to define a term (e.g., "nausea"), but not to define a concept where the respondent's subjective interpretation is the goal of the question (e.g., "quality of life").
- Respondents should be instructed to answer all items to the best of their ability. For a respondent who indicates the item asks about an activity they don't do, instruct them to consider what that activity would be like, and imagine or predict how it would be for them. For example, for the item "Are you able to use a hammer to pound a nail?" imagine the upper body strength, coordination, and dexterity needed to pound a nail with a hammer and evaluate to what degree they have that level of function. Respondents should always have the opportunity to skip an item if needed as well.
- Use the same method (e.g., computer, telephone, or paper) and mode (e.g., self vs. interviewer) of administration. However, this is not always possible, and PROMIS measures have produced similar scores when the method of administration varied.
- In clinical settings, give respondents the optimal time needed to capture the most relevant perspective and complete data (e.g., before/after clinician visit or in between visits). This may depend on the study aims and/or clinic workflow.

GUIDANCE ON THE APPEARANCE OF ITEMS

PROMIS measures have requirements for how they appear to the respondent. This helps support administration that is consistent with how the measures were developed, tested, and validated. The PDF versions of PROMIS measures from HealthMeasures.net's Search & View Measures meet these requirements. If you are integrating PROMIS measures into a digital administration platform, please contact Help@HealthMeasures.net to learn about how to have your implementation approved.

HANDLING MULTIPLE RESPONSES ON A PAPER FORM

Guidelines on how to deal with multiple responses have been established. Resolution depends on the responses noted by the respondent.

- If two or more responses are marked by the respondent and they are next to one another, then a data entry specialist will be responsible for randomly selecting one of them to be entered and will write down on the form which answer was selected. Note: To randomly select one of two responses, the data entry specialist will flip a coin (heads - higher number will be entered; tails – lower number will be entered). To randomly select one of three (or more) responses, a table of random numbers should be used with a statistician's assistance.
- If two or more responses are marked and they are NOT all next to one another, the response will be considered missing.

MEASURE NAMES IN ADMINISTRATION PLATFORMS

Measure names may be abbreviated or use altered terms in some administration platforms.

- Measures named "Item Banks" are administered as Computer Adaptive Tests (CATs). In any administration platform, a measure named Item Bank is the same as a CAT.
- SF = short form

- Ped = pediatric
- NRS = numeric rating scale (e.g., 0 to 10)
- Scale = a set of items that are administered in their entirety. They are not a subset of items from an item bank.
- Profile = a fixed set of short forms or CATs that are administered together (e.g., PROMIS-29). They produce scores for multiple domains (e.g., physical function, pain interference, anxiety).
- Pool = a collection of items that are not calibrated using item response theory. Item pools only generate raw response scores for each item and no summary score.
- Recommended = CAT uses different stopping rules than standard administration. See the Measure Differences summary for details.
- Screen-to-CAT = CAT uses different stopping rules than standard administration. See the Measure Differences summary for details.

SCORING INSTRUCTIONS

Short Forms

Response Pattern Scoring

PROMIS instruments are scored using item-level calibrations. This means that the most accurate way to score a PROMIS short form is to use the HealthMeasures Scoring Service (https://www.assessmentcenter.net/ac_scoringservice) or an administration platform that automatically calculates scores (e.g., REDCap auto-score). This method of scoring uses responses to each item for each participant. We refer to this as "response pattern scoring." **Because response pattern scoring is more accurate than the use of raw summed score/scale score look up tables included in this manual, it is preferred.** Response pattern scoring is especially useful when there is missing data (i.e., a respondent skipped an item), different groups of participants responded to different items, or you have created a new short form using a subset of items from a PROMIS item bank.

Instructions for the HealthMeasures Scoring Service

1. Download the Input Template from the HealthMeasures Scoring Service home page.
2. Add your data (all respondents' individual answers to questions) for one PROMIS measure to the template (multiple timepoints are okay).
3. Determine how much data is missing. In order to produce a precise score, a respondent must answer at least 4 items in measures that include 4 or more items. If a precise score is required, remove respondents who answered less than 4 items from the input template and do not calculate a T-score.
4. Upload your spreadsheet to the HealthMeasures Scoring Service. If this is your first time, register as a new user.
5. Access your email to receive a spreadsheet with calculated scores.

Platforms that use response pattern scoring can produce a T-score for a respondent as long as 1 item has been answered. However, in order to produce a **precise** score using response pattern scoring, a respondent must answer at least 4 items in a short form that includes 4 or more items. If fewer than 4 items have been answered, the resulting T-score may not be sufficiently precise. Depending on how the score will be used, you may want to discard the score or proceed with caution. For short forms with 1-3 items, all items should be answered before using response pattern scoring.

Instructions for Using Raw Sum Score to T-score Conversion Tables

To use the scoring tables in this manual, calculate a summed score. Each question usually has five response options ranging in value from one to five. To find the total raw score for a short form with all questions answered, sum the values of the response to each question. For example, for the adult 8-item form, the lowest possible raw score is 8; the highest possible raw score is 40 (see all short form scoring tables in Appendix 1). **All questions must be answered in order to produce a valid score using the scoring tables**. If a participant has skipped a question, use the HealthMeasures Scoring Service (https://www.assessmentcenter.net/ac_scoringservice) to generate a final score.

Locate the applicable score conversion table in Appendix 1 and use this table to translate the total raw score into a T-score for each participant. The T-score rescales the raw score into a standardized score with a mean of 50 and a standard deviation (SD) of 10. Therefore, a person with a T-score of 40 is one SD below the mean.

For the adult PROMIS Fatigue 7a Short Form, a raw score of 10 converts to a T-score of 39.6 with a standard error (SE) of 4 (see scoring table for the 7a short form in Appendix 1). Thus, the 95% confidence interval around the observed score ranges from 47.4 to 31.8.7 (T-score ± (1.96*SE) or 39.6 ± (1.96*4).

CAT

A minimum number of items (e.g., 4) must be answered in order to receive a score for the Fatigue CATs. The response to the first item will guide the system's choice of the next item for the participant. The participant's response to the second item will dictate the selection of the following question, and so on. As additional items are administered, the potential for error is reduced and confidence in the respondent's score increases. The CAT will continue until either the standard error drops below a specified level (e.g., on the T-score metric 3.0), or the participant has answered the maximum number of questions (e.g., 12), whichever occurs first. For some CATs, specifically "recommended" and "screen-to-CAT" there are additional stopping rules. These include stopping when the standard error isn't improving much or if a respondent is asymptomatic. For details on the exact stopping rules for Fatigue CATs, view the Measure Differences summary.

DIRECTION OF SCORES

For most PROMIS instruments, a score of 50 is the average for the general population with a standard deviation of 10 because calibration testing was performed on a large sample of the general population. You can read more about the calibration and centering samples in the Reference Population section below. The T-score is provided with an error term (Standard Error or SE). The Standard Error is a statistical measure of variance and represents the "margin of error" for the T-score.

Important: *A higher PROMIS T-score represents more of the concept being measured.* For negatively-worded concepts like Fatigue, a T-score of 60 is one SD worse than average. By comparison, a Fatigue T-score of 40 is one SD better than average.

REPORTING SCORES

T-scores and associated Standard Errors should be reported as integers. Additional decimal places may be retained in interim computations, but the reported value should be rounded to an integer to maintain the appropriate level of significant figures.

A Publication Checklist should be used when writing about PROMIS measures in publications and presentations. Checklist categories include measure details, administration, scoring, and reporting. The checklist is available on HealthMeasures.net at Publication Checklist as well as in Hanmer et al., 2020.

REFERENCE POPULATIONS

A PROMIS score of 50 is the mean score for a specific, relevant group of people (e.g., the general population, kids with a painful condition). That group is the **reference population**.

Centering Sample and Calibration Sample

Multiple steps are used to create T-scores.

1. Items are completed by a large sample of respondents. This is the **Calibration Sample**. The purpose of a calibration sample is to estimate item parameters (item characteristics such as difficulty and discrimination) using an item response theory (IRT) model.
2. Measure developers then decide which score on the IRT metric is going to be the middle score—a score of 50. This is done by collecting scores from a large sample that represents the reference population and then calculating the mean for that sample. That sample is the **Centering Sample**. That score becomes the middle score (e.g., 50 for T-scores).
3. A linear transformation spaces all other scores along the continuum so they have the correct values relative to the middle score (the mean of the Centering Sample).

Sometimes a single sample was used as both the Calibration Sample AND the Centering Sample. Other times one sample was used as the Calibration Sample and another was used as the Centering Sample. Sometimes, a subset of the Calibration Sample served as the Centering Sample.

For some PROMIS measures the reference population (and the centering sample) was a clinical population. This is the case for PROMIS Smoking measures, for which the reference populations was daily smokers; the Centering Sample was a sample of daily smokers. For many PROMIS measures, the reference population was the general population.

The Reference Population tables show the Calibration and the Centering samples for PROMIS. Most users will be particularly interested in the last column (Centering Sample). **If you want to know the reference population (what sample has a mean of 50), go to the Centering Sample column.**

Item Bank/Scale	Calibration Sample	Centering Sample
PROMIS Adult		
Fatigue	General population	General population
PROMIS-Cancer - Fatigue	General population	General population
PROMIS Pediatric		
Fatigue v1.0, v2.0	General population subset + Clinical sample	General population subset + Clinical sample
Fatigue 3.0	General population	General population
PROMIS Parent Proxy		
Fatigue v1.0, v2.0	General population subset + Clinical sample	General population subset + Clinical sample
Fatigue 3.0	General population	General population

STATISTICAL AND PSYCHOMETRIC TERMS

PROMIS measures produce T-scores and standard errors.

T-score

A T-score is a standardized score, like z-scores and IQ scores. All standardized scores have a "middle" score; it is zero for z-scores, 100 for IQ scores, and 50 for T-scores. This middle score is the mean of a large group of people. This group is known as the reference population. T-scores have a standard deviation (SD) of 10. Consequently, if a respondent produces a score of 60, that is one SD higher than the mean.

Precision and Standard Error (SE)

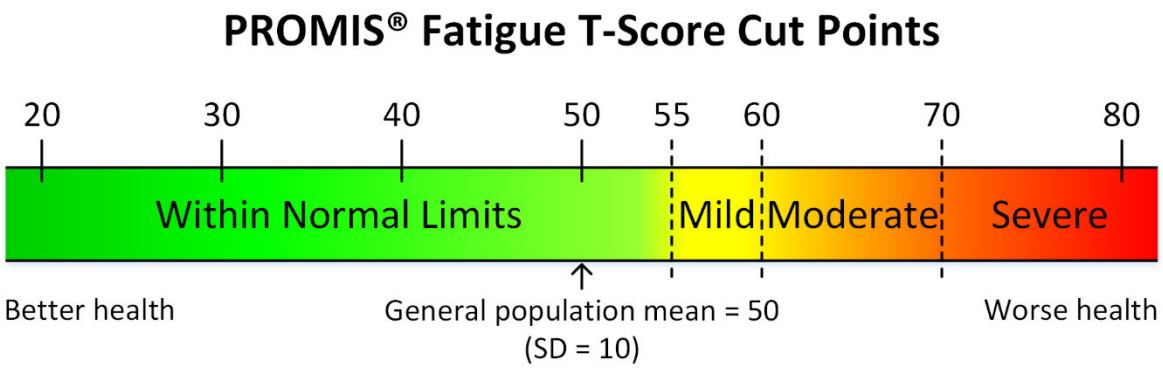
A PROMIS score includes a T-score and a standard error (SE). The standard error is a measure of precision or the variability for a given T-score across hypothetical repeated measurements. It represents the "margin of error" for a T-score. The standard error can be used to estimate precision by calculating confidence intervals around a T-score. A 95% confidence interval is commonly used. A 95% confidence interval means there is a 95% probability that the true T-score is within this range. The formula for a 95% confidence interval is (T-score $\pm$ (1.96*SE). For example, if T=52 and SE=2, the lower boundary of the confidence interval is (52 – (1.96*2)) = 48 and the upper boundary is (52 + (1.96*2)) = 56. Similarly, a 68% confidence interval can be calculated by (T-score $\pm$ SE). If T=52 and SE=2, the lower boundary of the confidence interval is (52 – 2) = 49 and the upper boundary is (52 + 2) = 54. T-scores should be rounded to whole numbers due to the number of significant figures in their computation.

Reliability

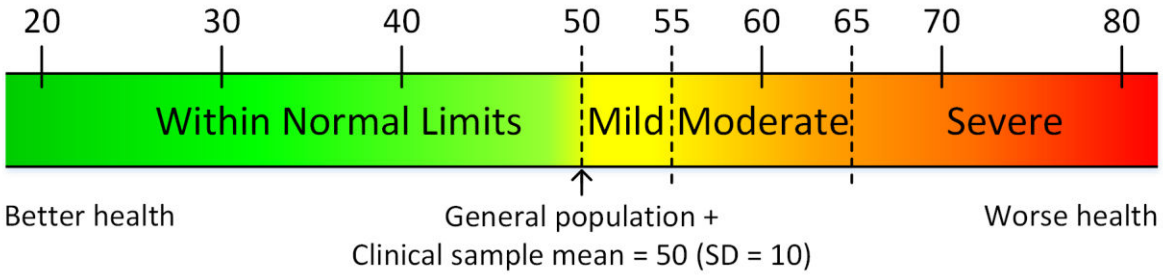
Reliability is the degree to which a measure is free of error. For IRT-based scaled scores (including PROMIS), reliability = $1 - SE^2$ using SEs on the theta metric (i.e., the T-score SE divided by 10). IRT-based reliability varies across the measurement continuum. That is, a measue may have much higher reliability for some score ranges (e.g., near the mean) compared to other score ranges (e.g., extreme ends of the continuum such as +/- 2 SDs from the mean).

SCORE CUT POINTS

Below, each figure shows score ranges and descriptive labels for PROMIS Fatigue item banks. The information in the figure applies to short forms and computer adaptive tests from that item bank. Be sure to select the correct figure for the measure you are using (e.g., respondent, version).



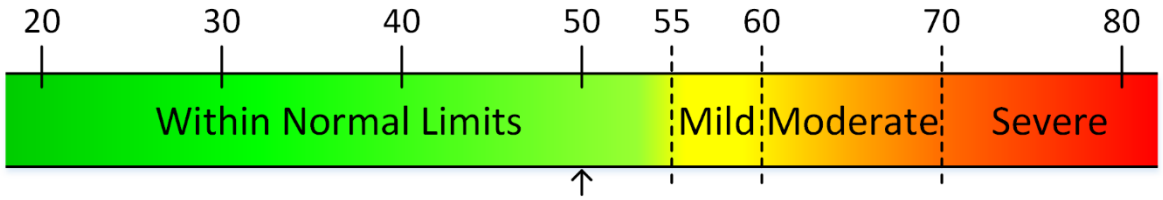
PROMIS® Pediatric and Parent Proxy Fatigue v2 T-Score Cut Points



Better health | General population + Clinical sample mean = 50 (SD = 10) | Worse health

Feb 2025

PROMIS® Pediatric and Parent Proxy Fatigue v3 T-Score Cut Points



Better health | General population mean = 50 (SD = 10) | Worse health

Feb 2025

[Remainder of page intentionally left blank]

REFERENCE VALUES FOR SPECIFIC PATIENT POPULATIONS

Few studies using representative samples have calculated reference values for individuals with specific health conditions or from countries other than the US. Using the general population mean of 50 is the best starting point for comparisons across all populations.

Oncology

Jensen et al. (2017) published reference values for people with cancer by age, stage, and cancer type (breast, prostate, uterine, cervical, colorectal, lung, and non-Hodgkin's lymphoma.) Level of fatigue varied by cancer type; patients with lung cancer reported the most fatigue. Patients with more advanced cancers (e.g., stage IV) reported more fatigue than patients with earlier stage cancers (e.g., stage I).

Country-specific Reference Values

Many studies have evaluated the measurement properties of translations of PROMIS measures in different countries. Some of these include reference values specific to a country. For example, there are reference values for Dutch translations of PROMIS anxiety, depression, fatigue, sleep, and global health (Elsman et al., 2021; Elsman, 2022; Terwee, Elsman et al., 2022; Terwee, van Litsenburg et al., 2022). Kaman et al. (2021) published reference values for German translations of PROMIS pediatric anger. To learn more about PROMIS reference values within a specific country, contact your PROMIS National Center representative.

COMPARING SCORES ACROSS VERSIONS

Some PROMIS domains have multiple versions of instruments (i.e., v1.0, v1.1, v2.0, v3.0). Generally, **it is recommended that you use the most recent version available which can be identified as the instrument with the highest version number.** In most cases, an instrument that has a decimal increase (v1.0 to v1.1) retains the same item-level parameters as well as instrument reliability and validity. In cases where a version number increases by a whole number (e.g., v1.0 to v2.0), the changes to the instrument are more substantial.

- For Fatigue Pediatric and Parent Proxy measures, the calibrations between v1.0 and v2.0 are identical. Consequently, T-scores from v1.0 and v2.0 are comparable.
- Scores from Pediatric and Parent Proxy GenPop v3.0 are NOT comparable with v1.0, v1.1, or v2.0 measure scores.

Transforming Scores between Versions:

To transform T-scores from Pediatric/Parent Proxy v1.0/v2.0 Fatigue measures to the Pediatric/Parent Proxy GenPop v3.0 metric, rescore v1.0/v2.0 measures using the HealthMeasures Scoring Service or Assessment Center API. This is the preferred strategy as it produces precise scores.

1. Download the v3.0 measure from Search & View Measures on HealthMeasures.net.
2. Revise your dataset as needed to match the v3.0 item IDs and response scores:

    - **PROMIS Parent Proxy Bank GenPop v3.0 – Fatigue**, **PROMIS Parent Proxy Bank GenPop v3.0 – Fatigue (screen-to-CAT)**, and **PROMIS Parent Proxy Short Form GenPop v3.0 – Fatigue 10a**:
        - The Item ID **Pf4fatigue4r** was changed to **Pf4fatigue4r2** to account for the revised Response Scores from 1-2-3-4-5 to 1-2-3-4-4.
        - The Item ID **Pf4fatigue12r** was changed to **Pf4fatigue12r2** to account for the revised Response Scores from 1-2-3-4-5 to 1-2-3-4-4.
    - **PROMIS Parent Proxy Bank GenPop v3.0 – Fatigue** and **PROMIS Parent Proxy Bank GenPop v3.0 – Fatigue (screen-to-CAT)**:
        - The Item ID **Pf4fatigue5r** was changed to **Pf4fatigue5r2** to account for the revised Response Scores from 1-2-3-4-4 to 1-2-3-4-5.
        - The Item ID **Pf1fatigue3r** was changed to **Pf1fatigue3r2** to account for the revised Response Scores from 1-2-3-4-4 to 1-2-3-4-5.

[Continued on next page]

3. Follow the instructions for using the HealthMeasures Scoring Service. Contact help@HealthMeasures.net if using the Assessment Center API for rescoring.

The crosswalk tables in Appendix 3 may also be used as a reference to guide understanding of how scores have shifted from the older v1.0 and v2.0 measures to the v3.0 measures. If these tables are used to convert older T-scores to GenPop v3.0, please note that the resulting T-scores are not as precise as those generated by using HealthMeasures Scoring Service or Assessment Center API.

[Remainder of page intentionally left blank]

REFERENCES

Cella, D., Choi, S. W., Condon, D. M., Schalet, B., Hays, R. D., Rothrock, N. E., Yount, S., Cook, K. F., Gershon, R. C., Amtmann, D., DeWalt, D. A., Pilkonis, P. A., Stone, A. A., Weinfurt, K., & Reeve, B. B. (2019). PROMIS® Adult Health Profiles: Efficient Short-Form Measures of Seven Health Domains. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, *22*(5), 537–544. https://doi.org/10.1016/j.jval.2019.02.004

Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., Amtmann, D., Bode, R., Buysse, D., Choi, S., Cook, K., Devellis, R., DeWalt, D., Fries, J. F., Gershon, R., Hahn, E. A., Lai, J. S., Pilkonis, P., Revicki, D., Rose, M., … 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. https://doi.org/10.1016/j.jclinepi.2010.04.011

Cook, K. F., Jensen, S. E., Schalet, B. D., Beaumont, J. L., Amtmann, D., Czajkowski, S., Dewalt, D. A., Fries, J. F., Pilkonis, P. A., Reeve, B. B., Stone, A. A., Weinfurt, K. P., & Cella, D. (2016). PROMIS measures of pain, fatigue, negative affect, physical function, and social function demonstrated clinical validity across a range of chronic conditions. *Journal of Clinical Epidemiology*, *73*, 89–102. https://doi.org/10.1016/j.jclinepi.2015.08.038

DeWalt DA, Rothrock N, Yount S, Stone AA; PROMIS Cooperative Group. Evaluation of item candidates: the PROMIS qualitative item review. *Medical Care*. 2007 May;45(5 Suppl 1):S12-21. doi: 10.1097/01.mlr.0000254567.79743.e2. PMID: 17443114; PMCID: PMC2810630.

Elsman, E. B. M., Flens, G., de Beurs, E., Roorda, L. D., & Terwee, C. B. (2022). Towards standardization of measuring anxiety and depression: Differential item functioning for language and Dutch reference values of PROMIS item banks. *PLoS ONE*, 17(8), e0273287. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9398458/

Elsman, E. B. M., Roorda, L. D., Crins, M. H. P., Boers, M., & Terwee, C. B. (2021). Dutch reference values for the Patient-Reported Outcomes Measurement Information System Scale v1.2 - Global Health (PROMIS-GH). *Journal of Patient Reported Outcomes*, 5(1), 38. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8116369/

Eremenco, S. L., Cella, D., & Arnold, B. J. (2005). A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Evaluation & the Health Professions*, *28*(2), 212–232. https://doi.org/10.1177/0163278705275342

Eremenco, S., Pease, S., Mann, S., Berry, P., & PRO Consortium's Process Subcommittee (2017). Patient-Reported Outcome (PRO) Consortium translation process: consensus development of updated best practices. *Journal of Patient-Reported Outcomes*, *2*(1), 12. https://doi.org/10.1186/s41687-018-0037-6

Forrest, C. B., Zorc, J. J., Moon, J., Pratiwadi, R., Becker, B. D., Maltenfort, M. G., & Guevara, J. P. (2019). Evaluation of the PROMIS pediatric global health scale (PGH-7) in children with asthma. *The Journal of Asthma: Official Journal of the Association for the Care of Asthma*, *56*(5), 534–542. https://doi.org/10.1080/02770903.2018.1471701

Hanmer, J., Jensen, R.E., & Rothrock, N. (2020). A reporting checklist for HealthMeasures' patient-reported outcomes: ASCQ-Me, Neuro-QoL, NIH Toolbox, and PROMIS. *Journal of Patient Reported Outcomes*, 4, 21. https://doi.org/10.1186/s41687-020-0176-4

Hansen, M., Cai, L., Stucky, B. D., Tucker, J. S., Shadel, W. G., & Edelen, M. O. (2014). Methodology for developing and evaluating the PROMIS smoking item banks. *Nicotine & Tobacco Research: Official Journal of the Society for Research on Nicotine and Tobacco*, *16 Suppl 3*(Suppl 3), S175–S189. https://doi.org/10.1093/ntr/ntt123

Irwin, D.E., Varni, J.W., Yeatts, K., & DeWalt, D.A. (2009). Cognitive Interviewing Methodology in the Development of a Pediatric Item Bank: A Patient Reported Outcomes Measurement Information System (PROMIS) Study. *Health and Quality of Life Outcomes*, 7(3).

Jensen RE, Potosky AL, Moinpour CM, et al. United States Population-Based Estimates of Patient-Reported Outcomes Measurement Information System Symptom and Functional Status Reference Values for Individuals With Cancer. *J Clin Oncol*, 2017;35(17):1913-1920. doi: 10.1200/JCO.2016.71.4410

Kaman, A., Otto, C., Devine, J., Erhart, M., Döpfner, M., Banaschewski, T., Görtz-Dorten, A., Hanisch, C., Kölch, M., Roessner, V., & Ravens-Sieberer, U. (2021). Assessing anger and irritability in children: psychometric evaluation and normative data for the German version of the PROMIS® Parent Proxy Anger Scale. *Quality of Life Research*. https://pubmed.ncbi.nlm.nih.gov/34586583/

Klem M, Saghafi E, Abromitis R, Stover A, Dew MA, Pilkonis P. (2009). Building PROMIS item banks: librarians as co-investigators. Quality of Life Research. 18(7):881-8. doi: 10.1007/s11136-009-9498-7. PMID: 19548118; PMCID: PMC2744389.

Pilkonis, P. A., Yu, L., Dodds, N. E., Johnston, K. L., Maihoefer, C. C., & Lawrence, S. M. (2014). Validation of the depression item bank from the Patient-Reported Outcomes Measurement Information System (PROMIS) in a three-month observational study. *Journal of Psychiatric Research*, *56*, 112–119. https://doi.org/10.1016/j.jpsychires.2014.05.010

Reeve, B. B. , Hays, R. D. , Bjorner, J. B. , Cook, K. F. , Crane, P. K. , Teresi, J. A. , Thissen, D. , Revicki, D. A. , Weiss, D. J. , Hambleton, R. K. , Liu, H. , Gershon, R. , Reise, S. P. , Lai, J. & Cella, D. (2007). Psychometric Evaluation and Calibration of Health-Related Quality of Life Item Banks. *Medical Care, 45* (5), S22-S31. doi: 10.1097/01.mlr.0000250483.85507.04.

Terwee, C.B., Elsman, E.B.M., & Roorda, L.D. (2022). Towards standardization of fatigue measurement: Psychometric properties and reference values of the PROMIS Fatigue item bank in the Dutch general population. *Research Methods in Medicine & Health Sciences*, 3(3), 86-98. doi:10.1177/26320843221089628

Terwee, C. B., van Litsenburg, R. R. L., Elsman, E. B. M., & Roorda, L. D. (2022). Psychometric properties and reference values of the Patient-Reported Outcomes Measurement Information System (PROMIS) sleep item banks in the Dutch general population. *Journal of Sleep Research*, e13753. https://doi.org/10.1111/jsr.13753

Wohlfahrt, A., Bingham, C. O., 3rd, Marder, W., Phillips, K., Bolster, M. B., Moreland, L. W., Zhang, Z., Neogi, T., & Lee, Y. C. (2019). Responsiveness of Patient-Reported Outcomes Measurement Information System Measures in Rheumatoid Arthritis Patients Starting or Switching a Disease-Modifying Antirheumatic Drug. *Arthritis Care & Research*, *71*(4), 521–529. https://doi.org/10.1002/acr.23617

[Remainder of page intentionally left blank]

APPENDIX 1—SCORING TABLES

Fatigue 4a - Adult v1.0		
Short Form Conversion Table		
Raw Score	T-score	SE*
4	33.7	4.9
5	39.7	3.1
6	43.1	2.7
7	46.0	2.6
8	48.6	2.5
9	51.0	2.5
10	53.1	2.4
11	55.1	2.4
12	57.0	2.3
13	58.8	2.3
14	60.7	2.3
15	62.7	2.4
16	64.6	2.4
17	66.7	2.4
18	69.0	2.5
19	71.6	2.7
20	75.8	3.9

*SE = Standard Error on T-score metric

Fatigue 6a - Adult v1.0		
Short Form Conversion Table		
Raw Score	T-score	SE*
6	33.4	4.9
7	39.1	2.9
8	42.0	2.4
9	44.2	2.2
10	46.1	2.1
11	47.8	2.1
12	49.4	2.1
13	50.9	2.0
14	52.4	2.0
15	53.7	2.0
16	55.1	2.0
17	56.3	1.9
18	57.5	1.9
19	58.8	1.9
20	60.0	1.9
21	61.2	1.9
22	62.4	1.9
23	63.7	2.0
24	65.0	2.0
25	66.4	2.0
26	67.8	2.0
27	69.3	2.0
28	71.0	2.1
29	73.0	2.5
30	76.8	3.8

*SE = Standard Error on T-score metric

Fatigue 7a - Adult v1.0		
Short Form Conversion Table		
Raw Score	T-score	SE*
7	29.4	5.3
8	33.4	4.8
9	36.9	4.3
10	39.6	4.0
11	41.9	3.8
12	43.9	3.5
13	45.8	3.3
14	47.6	3.2
15	49.2	3.1
16	50.8	3.0
17	52.2	3.0
18	53.7	3.0
19	55.1	3.0
20	56.4	2.9
21	57.8	2.9
22	59.2	2.9
23	60.6	2.9
24	62.0	2.9
25	63.4	2.9
26	64.8	2.9
27	66.3	2.9
28	67.8	2.9
29	69.4	2.9
30	71.1	3.0
31	72.9	3.0
32	74.8	3.1
33	77.1	3.3
34	79.8	3.6
35	83.2	4.1

*SE = Standard Error on T-score metric

PROMIS Adult v1.0 – Fatigue 7b Daily		
Short Form Conversion Table		
Raw Summed Score	**T-score**	**SE***
7	31.0	4.9
8	36.3	3.5
9	39.4	3.0
10	41.6	2.7
11	43.5	2.5
12	45.2	2.4
13	46.7	2.4
14	48.1	2.3
15	49.5	2.3
16	50.9	2.3
17	52.2	2.3
18	53.6	2.4
19	54.9	2.4
20	56.2	2.4
21	57.6	2.4
22	58.9	2.4
23	60.3	2.4
24	61.7	2.4
25	63.0	2.4
26	64.4	2.4
27	65.7	2.4
28	67.2	2.4
29	68.6	2.4
30	70.2	2.4
31	71.8	2.5
32	73.6	2.6
33	75.6	2.9
34	78.1	3.2
35	81.4	3.6

*SE = Standard Error on T-score metric

Fatigue 8a - Adult v1.0 *Short Form Conversion Table*					
Raw Score	**T-score**	**SE***	**Raw Score**	**T-score**	**SE***
8	33.1	4.8	25	58.5	1.7
9	38.5	2.7	26	59.4	1.7
10	41	2.2	27	60.4	1.7
11	42.8	2	28	61.3	1.7
12	44.3	1.9	29	62.3	1.7
13	45.6	1.8	30	63.3	1.7
14	46.9	1.8	31	64.3	1.7
15	48.1	1.8	32	65.3	1.7
16	49.2	1.8	33	66.4	1.7
17	50.4	1.8	34	67.5	1.7
18	51.5	1.7	35	68.6	1.7
19	52.5	1.7	36	69.8	1.8
20	53.6	1.7	37	71	1.8
21	54.6	1.7	38	72.4	2
22	55.6	1.7	39	74.2	2.4
23	56.6	1.7	40	77.8	3.7
24	57.5	1.7			
*SE = Standard Error on T-score metric					

PROMIS Adult v1.0 – Fatigue 10a (FACIT-Fatigue-10)					
Short Form Conversion Table					
Raw Summed Score	**T-Score**	**SE***	**Raw Summed Score**	**T-Score**	**SE***
10	30.4	4.8	31	59.7	2
11	35.1	3.5	32	60.5	2
12	38.2	3.1	33	61.2	2
13	40.7	2.9	34	62	2
14	42.7	2.7	35	62.8	2
15	44.3	2.6	36	63.5	2
16	45.8	2.5	37	64.3	2
17	47.2	2.3	38	65.1	2.1
18	48.4	2.2	39	65.9	2.1
19	49.5	2.2	40	66.7	2.1
20	50.5	2.1	41	67.6	2.1
21	51.5	2.1	42	68.5	2.2
22	52.4	2.1	43	69.5	2.3
23	53.3	2	44	70.6	2.4
24	54.2	2	45	71.7	2.5
25	55	2	46	72.9	2.7
26	55.9	2	47	74.3	2.9
27	56.7	2	48	75.9	3.1
28	57.4	2	49	77.9	3.4
29	58.2	2	50	80.7	3.8
30	59	2			

SE* = Standard Error on T-Score metric

Fatigue 13a (FACIT-Fatigue) - Adult v1.0 *Short Form Conversion Table*					
Raw Score	T-score	SE*	Raw Score	T-score	SE*
13	30.3	4.7	40	60.8	1.8
14	35.0	3.5	41	61.4	1.8
15	38.0	3.0	42	62.0	1.8
16	40.3	2.8	43	62.6	1.8
17	42.1	2.6	44	63.2	1.8
18	43.7	2.5	45	63.8	1.8
19	45.0	2.3	46	64.4	1.8
20	46.3	2.2	47	65.0	1.8
21	47.3	2.1	48	65.6	1.8
22	48.3	2.0	49	66.2	1.9
23	49.3	2.0	50	66.9	1.9
24	50.1	1.9	51	67.5	1.9
25	51.0	1.9	52	68.2	1.9
26	51.7	1.9	53	68.9	2.0
27	52.5	1.9	54	69.6	2.0
28	53.2	1.9	55	70.4	2.0
29	53.9	1.8	56	71.2	2.1
30	54.6	1.8	57	72.0	2.2
31	55.3	1.8	58	72.9	2.3
32	55.9	1.8	59	73.9	2.4
33	56.6	1.8	60	75.0	2.5
34	57.2	1.8	61	76.2	2.7
35	57.8	1.8	62	77.5	2.9
36	58.4	1.8	63	79.1	3.1
37	59.0	1.8	64	81.2	3.3
38	59.6	1.8	65	83.5	3.4
39	60.2	1.8			

*SE = Standard Error on T-score metric

PROMIS Adult Short Form v1.0 – Fatigue-OA-Knee 8a					
Short Form Conversion Table					
Raw Summed Score	**T-Score**	**SE***	**Raw Summed Score**	**T-Score**	**SE***
8	32.8	4.8	25	59.3	2.0
9	38.1	3.0	26	60.3	2.0
10	40.7	2.6	27	61.4	2.0
11	42.7	2.4	28	62.4	2.0
12	44.4	2.3	29	63.4	2.0
13	45.8	2.2	30	64.4	2.0
14	47.2	2.2	31	65.4	2.0
15	48.4	2.1	32	66.5	2.0
16	49.6	2.1	33	67.6	2.1
17	50.8	2.0	34	68.8	2.1
18	51.9	2.0	35	70.0	2.1
19	53.0	2.0	36	71.4	2.3
20	54.1	2.0	37	72.9	2.4
21	55.1	2.0	38	74.7	2.7
22	56.2	2.0	39	77.0	3.1
23	57.2	2.0	40	80.2	3.7
24	58.2	2.0			

*SE = Standard Error on T-score metric

PROMIS Adult v1.0 – Fatigue – Multiple Sclerosis 8a		
Short Form Conversion Table		
Raw Summed Score	**T-Score**	**SE***
8	34.1	5.0
9	39.3	3.1
10	41.8	2.6
11	43.6	2.3
12	45.1	2.1
13	46.4	2.1
14	47.6	2.0
15	48.7	2.0
16	49.8	2.0
17	50.9	2.0
18	52.0	2.0
19	53.0	2.0
20	54.1	2.0
21	55.1	2.0
22	56.2	2.0
23	57.3	2.0
24	58.3	2.0
25	59.3	2.0
26	60.4	2.0
27	61.5	2.0
28	62.5	2.0
29	63.6	2.0
30	64.7	2.0
31	65.8	2.0
32	66.9	2.0
33	68.1	2.0
34	69.3	2.0
35	70.5	2.1
36	71.9	2.1
37	73.4	2.3
38	75.1	2.5
39	77.5	3.0
40	80.9	3.9

*SE = Standard Error on T-Score metric

PROMIS Pediatric GenPop v3.0 – Fatigue 10a		
Short Form Conversion Table		
Raw Summed Score	**T-score**	**SE***
10	41.1	6.1
11	48.3	3.1
12	50.5	2.6
13	52.0	2.2
14	53.2	1.9
15	54.2	1.8
16	55.1	1.7
17	55.9	1.6
18	56.6	1.6
19	57.3	1.6
20	58.0	1.6
21	58.7	1.6
22	59.3	1.6
23	60.0	1.6
24	60.7	1.6
25	61.3	1.6
26	62.0	1.6
27	62.7	1.6
28	63.3	1.6
29	64.0	1.6
30	64.6	1.6
31	65.2	1.6
32	65.9	1.6
33	66.5	1.6
34	67.1	1.6
35	67.7	1.6
36	68.3	1.5
37	68.9	1.5
38	69.5	1.5
39	70.1	1.5
40	70.7	1.5
41	71.4	1.5
42	72.0	1.6
43	72.7	1.6
44	73.4	1.7
45	74.2	1.7
46	75.1	1.8
47	76.1	2.0
48	77.3	2.2
49	78.9	2.5
50	81.7	3.1

*SE = Standard Error on T-score metric

PROMIS Parent Proxy GenPop v3.0 – Fatigue 10a		
Short Form Conversion Table		
Raw Summed Score	**T-score**	**SE***
10	41.9	6.3
11	49.5	3.1
12	51.9	2.4
13	53.5	2.0
14	54.7	1.8
15	55.7	1.7
16	56.6	1.6
17	57.5	1.6
18	58.3	1.6
19	59.1	1.6
20	59.9	1.6
21	60.7	1.6
22	61.4	1.6
23	62.2	1.6
24	62.9	1.6
25	63.6	1.6
26	64.4	1.6
27	65.1	1.6
28	65.8	1.6
29	66.5	1.6
30	67.2	1.6
31	67.9	1.6
32	68.5	1.6
33	69.2	1.6
34	69.8	1.6
35	70.5	1.6
36	71.1	1.5
37	71.8	1.5
38	72.5	1.6
39	73.1	1.6
40	73.8	1.6
41	74.5	1.6
42	75.3	1.7
43	76.2	1.8
44	77.2	1.9
45	78.3	2.0
46	79.7	2.2
47	81.6	2.5
48	84.1	2.8
*Items **Pf4fatigue4r** and **Pf4fatigue12r** have collapsed categories 4 and 5*		

*SE = Standard Error on T-score metric

APPENDIX 2—SCORING TABLES FOR RETIRED MEASURES

Fatigue 10a - Pediatric v1.0					
Short Form Conversion Table					
Raw Score	**T-Score**	**SE***	**Raw Score**	**T-Score**	**SE***
0	30.3	5.5	21	60.6	3.3
1	34.3	4.7	22	61.6	3.3
2	36.9	4.4	23	62.6	3.3
3	39	4.1	24	63.6	3.3
4	40.9	3.9	25	64.6	3.3
5	42.5	3.8	26	65.6	3.3
6	44	3.7	27	66.7	3.3
7	45.4	3.6	28	67.7	3.3
8	46.7	3.5	29	68.7	3.3
9	47.9	3.5	30	69.8	3.3
10	49.1	3.4	31	70.9	3.3
11	50.2	3.4	32	72	3.4
12	51.3	3.4	33	73.2	3.4
13	52.4	3.4	34	74.4	3.4
14	53.5	3.4	35	75.7	3.5
15	54.5	3.4	36	77	3.6
16	55.6	3.4	37	78.5	3.6
17	56.6	3.4	38	80.2	3.7
18	57.6	3.4	39	82	3.7
19	58.6	3.3	40	84	3.5
20	59.6	3.3			
*SE = Standard Error on T-score					

Fatigue 10a - Pediatric v2.0					
Short Form Conversion Table					
Raw Score	**T-Score**	**SE***	**Raw Score**	**T-Score**	**SE***
10	30.3	5.5	31	60.6	3.3
11	34.3	4.7	32	61.6	3.3
12	36.9	4.4	33	62.6	3.3
13	39	4.1	34	63.6	3.3
14	40.9	3.9	35	64.6	3.3
15	42.5	3.8	36	65.6	3.3
16	44	3.7	37	66.7	3.3
17	45.4	3.6	38	67.7	3.3
18	46.7	3.5	39	68.7	3.3
19	47.9	3.5	40	69.8	3.3
20	49.1	3.4	41	70.9	3.3
21	50.2	3.4	42	72	3.4
22	51.3	3.4	43	73.2	3.4
23	52.4	3.4	44	74.4	3.4
24	53.5	3.4	45	75.7	3.5
25	54.5	3.4	46	77	3.6
26	55.6	3.4	47	78.5	3.6
27	56.6	3.4	48	80.2	3.7
28	57.6	3.4	49	82	3.7
29	58.6	3.3	50	84	3.5
30	59.6	3.3			
*SE= Standard Error on T-Score metric					

Fatigue 10a - Parent Proxy v1.0 *Short Form Conversion Table*					
Raw Score	**T-Score**	**SE***	**Raw Score**	**T-Score**	**SE***
0	34	5	21	63	2
1	39	4	22	64	2
2	42	3	23	65	2
3	44	3	24	66	2
4	45	3	25	67	2
5	47	3	26	68	2
6	48	2	27	69	2
7	49	2	28	70	2
8	50	2	29	71	2
9	51	2	30	72	2
10	52	2	31	72	2
11	53	2	32	73	2
12	54	2	33	74	2
13	55	2	34	75	2
14	56	2	35	76	2
15	57	2	36	77	2
16	58	2	37	79	3
17	59	2	38	80	3
18	60	2	39	82	3
19	61	2	40	85	4
20	62	2			
*SE = Standard Error on T-score					

Fatigue 10a - Parent Proxy v2.0 *Short Form Conversion Table*					
Raw Score	T-Score	SE*	Raw Score	T-Score	SE*
10	34	5	31	63	2
11	39	4	32	64	2
12	42	3	33	65	2
13	44	3	34	66	2
14	45	3	35	67	2
15	47	3	36	68	2
16	48	2	37	69	2
17	49	2	38	70	2
18	50	2	39	71	2
19	51	2	40	72	2
20	52	2	41	72	2
21	53	2	42	73	2
22	54	2	43	74	2
23	55	2	44	75	2
24	56	2	45	76	2
25	57	2	46	77	2
26	58	2	47	79	3
27	59	2	48	80	3
28	60	2	49	82	3
29	61	2	50	85	4
30	62	2			

*Standard Error on T-score metric

All scoring tables are based on default Parent Proxy calibrations.

APPENDIX 3—CROSSWALK BETWEEN PEDIATRIC AND PARENT PROXY V2.0 T-SCORES AND PEDIATRIC AND PARENT PROXY GENPOP V3.0 T-SCORES

Table 1. PROMIS Pediatric Fatigue: Crosswalk between v2.0 T-scores and GenPop v3.0 T-scores

PROMIS Pediatric Fatigue T-Score	
v2.0	**v3.0 GenPop**
24	37.0
25	38.5
26	41.6
27	42.6
28	44.1
29	45.2
30	46.1
31	47.3
32	47.9
33	48.4
34	49.3
35	49.9
36	50.7
37	51.4
38	51.9
39	52.4
40	52.9
41	53.6
42	54.0
43	54.4
44	54.7
45	55.2
46	55.9
47	56.5
48	57.0
49	57.6
50	58.0
51	58.7
52	59.3
53	59.7
54	60.3
55	60.9
56	61.7
57	62.4
58	63.1
59	63.6
60	64.3
61	64.8

PROMIS Pediatric Fatigue T-Score	
v2.0	**v3.0 GenPop**
62	65.3
63	65.9
64	66.9
65	67.6
66	68.0
67	68.5
68	69.2
69	70.0
70	70.5
71	70.7
72	71.5
73	73.0
74	73.0
75	73.0
76	73.0
77	73.8
78	74.5
79	75.2

Table 2. PROMIS Parent Proxy Fatigue: Crosswalk between v2.0 T-scores and GenPop v3.0 T-scores

PROMIS Parent Proxy Fatigue T-Score	
v2.0	**v3.0 GenPop**
29	37.0
30	38.5
31	39.0
32	41.8
33	42.9
34	44.6
35	45.6
36	46.4
37	47.8
38	48.8
39	49.5
40	50.4
41	51.2
42	52.1
43	52.9
44	53.4
45	54.1
46	54.7
47	55.6
48	56.3
49	56.9
50	57.5
51	58.4
52	59.1
53	59.6
54	60.3
55	61.4
56	62.3
57	62.9
58	63.4
59	64.2
60	64.9
61	65.6
62	66.5
63	67.3
64	68.0
65	68.8
66	70.0

PROMIS Parent Proxy Fatigue T-Score	
v2.0	**v3.0 GenPop**
67	70.0
68	71.0
69	72.0
70	72.5
71	73.0
72	73.5
73	73.5
74	74.0
75	76.0
76	76.0
77	76.0
78	78.0
79	80.0

APPENDIX 4— SUPPLEMENTAL INFORMATION FOR THE *PROMIS SHORT FORM v1.0—FATIGUE-MULTIPLE SCLEROSIS 8A* (*PROMIS FatigueMS—8a*)

OVERVIEW OF FATIGUE IN MULTIPLE SCLEROSIS

Multiple sclerosis (MS) is a chronic, disabling disease that affects the central nervous system and causes a range of neurological symptoms and impacts on activities of daily living [1-3] As of 2017, approximately 900,000 individuals in the United States (U.S.) are living with an MS diagnosis.[4] Currently, there is no cure for MS; treatment is a lifelong process that includes medication, rehabilitation, and emotional support.[3]

The majority of people with MS experience severe, debilitating fatigue as a result of their condition.[5,6] Fatigue is frequently identified as the most prevalent symptom of MS[7-20] and fatigue may be the first presenting symptom,[21] manifesting even before a diagnosis of MS is made.[22] People with all stages of MS, from mild to severe, can experience significant fatigue.[21,23] The symptom can be acute or chronic.[5,6,24] However, fatigue is a daily experience for many people living with MS, lasting six or more hours, and often worsening later in the day.[25,26]

Independent of disability,[27,28] fatigue reduces health-related quality of life in people with MS.[21,23,28-31] For example, fatigue in MS can affect one's ability to work and to retain employment.[13,16,22,27,32,33] According to the U.S. Social Security Administration, fatigue is a leading cause of disability in persons with MS.[23] Fatigue in MS also limits physical exertion; physical limitations can lead to deconditioning and exercise intolerance if patients curtail physical activity to minimize fatigue.[23,34,35] The impact of fatigue can extend to a decline in overall health in the person with MS, given that it represents a barrier to health-promoting behaviors and activities in this population.[36] Fatigue can exacerbate other MS symptoms,[21,37] including negatively affecting cognitive function.[38]

OVERVIEW OF THE *PROMIS SHORT FORM V1.0—FATIGUE-MULTIPLE SCLEROSIS 8A* (*PROMIS FATIGUEMS—8A*)

Description

The *PROMIS FatigueMS—8a* is an 8-item short form derived from the PROMIS Fatigue Item Bank to assess the experience and impact of fatigue in people with MS.[39,40] It has a recall period of the past 7 days. The *PROMIS FatigueMS—8a* consists of seven items with frequency-based response options ("Never," "Rarely," "Sometimes," "Often," "Always"), and one item with response options of "Not at all," "A little bit," "Somewhat," "Quite a bit," or "Very much." Consistent with scoring guidelines for PROMIS measures, this short form is scored on a T-score metric, which is centered on the U.S. general population (mean = 50; SD = 10).

Reporter

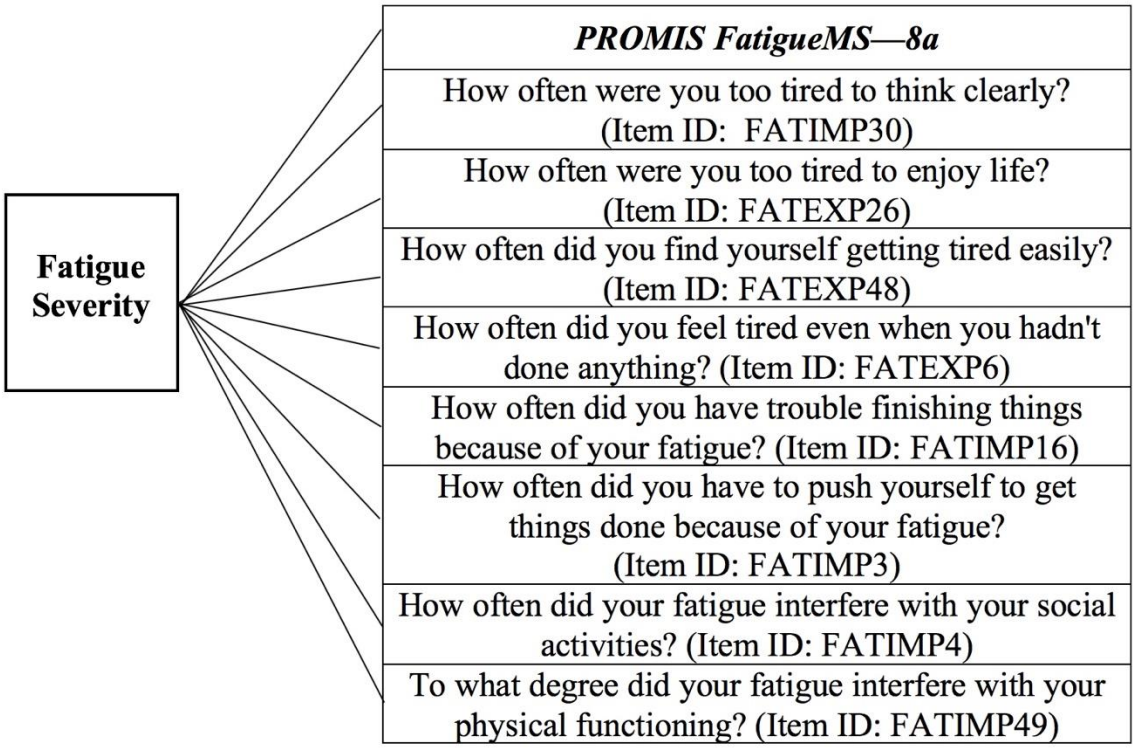
The *PROMIS FatigueMS—8a* is a self-report measure. Respondents choose the response option that most accurately describes their experience of fatigue and its impact on their daily living.

Conceptual Framework

The conceptual framework for the assessment of fatigue in MS based on the *PROMIS FatigueMS—8a* is shown in Figure 1. The *PROMIS FatigueMS—8a* is derived from the PROMIS Fatigue Item Bank, where all items met the unidimensionality criteria and were calibrated on the same measurement continuum, i.e., fatigue. IRT allows for the modeling of the relationship of the individual items to the construct (i.e., fatigue) being measured.[41] The *PROMIS FatigueMS—8a* is hypothesized as a unidimensional construct of the severity of fatigue consisting of 8 items that generate a total score. The measure's items address the following content: fatigue impact on thinking clearly, influence of fatigue on life enjoyment, easily tired, feeling tired without doing anything, trouble finishing

things because of fatigue, need to push oneself, fatigue interference with social activities, and fatigue inference with physical functioning.

Figure 1. The *PROMIS FatigueMS—8a* Conceptual Framework



DEVELOPMENT AND EVALUATION OF THE *PROMIS FATIGUEMS—8A*

Development

Prior to development of the *PROMIS FatigueMS—8a*, three published PROMIS Fatigue short forms were available;[42] however, none were developed specifically for use among individuals with MS. Therefore, Cook and colleagues sought to develop a content valid measure of fatigue for individuals with MS.[39,40] To reduce participant burden, investigators (two clinicians and four experts in measure development) reduced the PROMIS Fatigue item pool from 95 items to 44 candidate items using two criteria. First, 13 items previously published as a measure of fatigue in cancer which was considered a legacy measure during the PROMIS Fatigue Item Bank testing were removed because their wording (e.g., "I feel…" "I have…" "I need…") was inconsistent with the rest of PROMIS Fatigue Item Bank items (e.g., "How often…" "To what degree…").[43] Second, items with duplicate content were reviewed and those with frequency-based response options were retained. The decision to prioritize frequency-based response options was based on prior cognitive interview findings indicating that individuals with disabilities prefer frequency-based response options (e.g., "Never," "Rarely," "Sometimes," "Often," "Always") to intensity response options (e.g., "Not at all," "A little bit," "Somewhat," "Quite a bit," "Very much").[44] Additionally, when compared to other fatigue response options, fatigue frequency response options have reduced floor and ceiling effects, and their response threshold locations are more equidistant.[45] Through the item review process, 38 items with response options "Not at all," "A little bit," "Somewhat," "Quite a bit," and "Very much" were dropped and items with duplicate content that included response options "Never," "Rarely," "Sometimes," "Often," and "Always" were retained. The remaining six items had response options of "Not at all," "A little bit," "Somewhat," "Quite a bit," or "Very much" because frequency options were not appropriate for these items. Of these six items, 1 was retained for the final measure. The 44 remaining candidate items were selected to represent the range of content in the PROMIS Fatigue Item Bank.[42]

Next, the 44 candidate items were reviewed by clinicians (N=37) and ranked based on relevance to MS, the results of which were used to further reduce the candidate items to 20. Forty-six people with MS reviewed the 20 candidate items and ranked the top 10 on relevance to their experience with MS. Final item selection for the *PROMIS FatigueMS—8a* was informed by four criteria: (1) rankings by individuals with MS and experts; (2) high discrimination parameters based on IRT calibration; (3) cumulative representatives of the full PROMIS Fatigue Item Bank content; and (4) content validity assessment.[39]

Evidence of Content Validity

Content validity of the *PROMIS FatigueMS—8a* was established through expert input, literature review, and qualitative input from individuals with MS. The content validity evidence is summarized in Table 1. Taken together, this evidence provides strong support for the coverage, relevance, and content validity of *PROMIS FatigueMS—8a*.

Table 1. Summary of Content Validity Support for the *PROMIS FatigueMS—8a*

Activity	Source of Evidence	Citation
Review and reduction of the PROMIS Fatigue Item Bank	**Expert input** from 2 clinicians with expertise in the treatment of MS and 4 experts in measure development	Cook KF, Bamer AM, Roddey TS, Kraft GH, Kim J, Amtmann D. A PROMIS fatigue short form for use by individuals who have multiple sclerosis. Quality of Life Research. 2012;21(6):1021-1030.
Expert review of candidate items for the *PROMIS FatigueMS—8a*	**Clinician input** from 37 clinicians with expertise in the treatment of MS	Cook KF, Bamer AM, Roddey TS, Kraft GH, Kim J, Amtmann D. A PROMIS fatigue short form for use by individuals who have multiple sclerosis. Quality of Life Research. 2012;21(6):1021-1030.
Qualitative and quantitative data evaluating the candidate items for the *PROMIS FatigueMS—8a*, including item ranking and open-ended input on content validity from people with MS	Input from 46 **people with MS**,	Cook KF, Bamer AM, Roddey TS, Kraft GH, Kim J, Amtmann D. Multiple sclerosis and fatigue: understanding the patient's needs. Physical Medicine & Rehabilitation Clinics. 2013;24(4):653-661.
Summary of qualitative research in patients with MS (Literature review conducted by RTI Health Solutions)	**Literature review**	N/A
MS targeted literature and instrument review (Literature review conducted by RTI Health Solutions)	**Literature review**	Martin SA, Olayinka-Amao OC, Henke C, Kamudoni P, Gnanasakthy A. Conceptual Model and Instrument Review in Multiple Sclerosis. Value in Health. 2017;20:A399-A811.
Supplementary review of measures of fatigue in MS (Literature review conducted by Northwestern University)	**Literature review**	N/A

Interviews with experts about the MS fatigue experience (Conducted by RTI Health Solutions)	**Clinician input** from 2 clinicians with expertise in the treatment of MS	N/A
Concept elicitation interviews with people with MS (Conducted by RTI Health Solutions)	Input from 14 **people with MS**	N/A
Cognitive interviews with people with MS (Conducted by RTI Health Solutions)	Input from 29 **people with MS**	N/A

Measurement Properties and Psychometric Evaluation

Psychometric properties of the *PROMIS FatigueMS—8a* were evaluated in 2022 using data from two projects: 1) a survey of individuals with MS in the United Kingdom (UK-MS Study),[46] and 2) the *PROMIS FatigueMS—8a* Development Study (Development Study).[39] Cross-sectional analyses were conducted using baseline data from both studies. Longitudinal analyses, including ability to detect change analyses and estimation of within-person meaningful change thresholds, were conducted using the UK-MS Study data collected at baseline, week 24, and week 52. Descriptive statistics included frequency distributions of both item response and overall scores, floor and ceiling effects, and percentage of missing responses. The frequency and percentage of each response option selected for each item at baseline on both studies were used to describe the distribution of items in the *PROMIS FatigueMS—8a*. A threshold of 35% (approximately one-third) of the sample selecting either the highest or lowest response was used to indicate a floor or ceiling effect for that item. No items demonstrated floor or ceiling effects across both studies. Missing data at the item-level were not observed because the platform did not allow for items to be skipped. Missing data at the form-level were noticed as the sample size decreased over time in the UK-MS Study. Inter-item consistency was supported with Cronbach's Alpha ≥ 0.90 in both studies. Strong test-retest reliability was found on UK-MS Study data with the intraclass correlation coefficients (ICCs) = 0.90. Test-retest reliability was not available for the Development Study. Regarding IRT-based reliabilities, the *PROMIS FatigueMS—8a* has excellent reliability (≥ 0.70) across a wide range of fatigue scores reported by participants from both studies.

Convergent validity analyses yielded strong correlations between the *PROMIS FatigueMS—8a* and other clinical outcomes measures in both studies. The known-groups validity analysis examined differences in *PROMIS FatigueMS—8a* scores between subgroups of participants with known status using different outcome measures as anchors, including the patient-reported Expanded Disability Status Scale, the PROMIS Global Health Scale (general health, physical health, fatigue), the Multiple Sclerosis Impact Scale (MSIS-29), the Fatigue Severity Scale (FSS). The *PROMIS FatigueMS—8a* scores showed expected and statistically significant ($p < 0.001$) differences across patient subgroups for all anchor measures applied in both studies.

Results from a longitudinal evaluation of the measurement properties of the *PROMIS FatigueMS—8a* using UK-MS study data indicated that the measure is sensitive to change from baseline to week 24 and from baseline to week 52 using clinical anchors. Results showed that *PROMIS FatigueMS—8a* T-scores significantly worsened on all worsening groups both from baseline to week 24 and from baseline to week 54. Similarly, *PROMIS FatigueMS—8a* T-scores improved on most improving groups, both from baseline to week 24 and from baseline to week 54. Most measures of effect size were greater than 0.2 on both improving and worsening groups from baseline to week 24 as well as from baseline to week 52.

Taken together, these analyses demonstrate that the *PROMIS FatigueMS—8a* can be incorporated into clinical trials for longitudinal evaluation given the measure's ability to detect change over time and its strong psychometric properties.

Scoring and Interpretation Procedures

Please see the section on scoring of short forms in this PROMIS Fatigue User Manual and Scoring Instructions.

Handling of Missing Data

- **Form-level missing data.** Form-level missing data refers to a respondent missing an entire PRO assessment for a given time point. In general, form-level data may be missing due to a respondent's early withdrawal from the study, the inability to evaluate an endpoint at a particular time point, or non-compliance. Imputation is not recommended for *PROMIS FatigueMS—8a* form-level missing data. If a participant does not complete the *PROMIS FatigueMS—8a*, such as due to attrition in longitudinal studies or due to forgetting to complete an individual assessment, his or her *PROMIS FatigueMS—8a* should not be computed for that time point.
- **Item-level missing data.** Administrators may decide to allow respondents to skip individual items within the *PROMIS FatigueMS—8a*. If respondents are allowed to skip items, missing data at the item-level may be present. Based on missing data simulation analyses, a 4-item rule will be employed for missing data at the item-level with scores only calculated if data for 4 or more items in each assessment are available, with the available item values summed and divided by the number of items available. If 5 or more items are missing, the assessment itself will be considered missing and will be handled under the form-level missing data procedures.

[Remainder of page intentionally left blank]

REFERENCES FOR APPENDIX 4

1. Loma I, Heyman R. Multiple sclerosis: pathogenesis and treatment. *Current Neuropharmacology.* 2011;9(3):409-416.
2. Motl RW, McAuley E, Wynn D, Suh Y, Weikert M, Dlugonski D. Symptoms and physical activity among adults with relapsing-remitting multiple sclerosis. *The Journal of Nervous and Mental Disease.* 2010;198(3):213-219.
3. National Multiple Sclerosis Society. What is MS? http://www.nationalmssociety.org/What-is-MS. Published 2019. Accessed July 15, 2019.
4. Wallin MT, Culpepper WJ, Campbell JD, et al. The prevalence of MS in the United States: a population-based estimate using health claims data. *Neurology.* 2019;92(10):e1029-e1040.
5. Egner A, Phillips VL, Vora R, Wiggers E. Depression, fatigue, and health-related quality of life among people with advanced multiple sclerosis: results from an exploratory telerehabilitation study. *NeuroRehabilitation.* 2003;18(2):125-133.
6. Mathiowetz V. Test-retest reliability and convergent validity of the Fatigue Impact Scale for persons with multiple sclerosis. *The American Journal of Occupational Therapy.* 2003;57(4):389-395.
7. Blaney BE, Lowe-Strong A. The impact of fatigue on communication in multiple sclerosis. The insider's perspective. *Disability and Rehabilitation.* 2009;31(3):170-180.
8. Courts NF, Buchanan EM, Werstlein PO. Focus groups: the lived experience of participants with multiple sclerosis. *Journal of Neuroscience Nursing* 2004;36(1):42-47.
9. DiLorenzo TA, Becker-Feigeles J, Halper J, Picone MA. A qualitative investigation of adaptation in older individuals with multiple sclerosis. *Disability and Rehabilitation.* 2008;30(15):1088-1097.
10. Douglas C, Windsor C, Wollin J. Understanding chronic pain complicating disability: finding meaning through focus group methodology. *Journal of Neuroscience Nursing.* 2008;40(3):158-168.
11. Isaksson AK, Ahlstrom G. From symptom to diagnosis: illness experiences of multiple sclerosis patients. *Journal of Neuroscience Nursing.* 2006;38(4):229-237.
12. Isaksson AK, Gunnarsson LG, Ahlstrom G. The presence and meaning of chronic sorrow in patients with multiple sclerosis. *Journal of Clinical Nursing.* 2007;16(11C):315-324.
13. Knaster ES, Yorkston KM, Johnson K, McMullen KA, Ehde DM. Perspectives on self-management in multiple sclerosis: a focus group study. *International Journal of MS Care.* 2011;13(3):146-152.
14. Moriya R, Suzuki S. A qualitative study relating to the experiences of people with MS: differences by disease severity. *British Journal of Neuroscience Nursing.* 2011;7(4):593-600.
15. Newland PK, Flick LH, Thomas FP, Shannon WD. Identifying symptom co-occurrence in persons with multiple sclerosis. *Clinical Nursing Research.* 2014;23(5):529-543.
16. Newland PK, Thomas FP, Riley M, Flick LH, Fearing A. The use of focus groups to characterize symptoms in persons with multiple sclerosis. *Journal of Neuroscience Nursing.* 2012;44(6):351-357.
17. Nicolson P, Anderson P. The psychosocial impact of spasticity-related problems for people with multiple sclerosis: a focus group study. *Journal of Health Psychology.* 2001;6(5):551-567.
18. Somerset M, Sharp D, Campbell R. Multiple sclerosis and quality of life: a qualitative investigation. *Journal of Health Services Research & Policy.* 2002;7(3):151-159.
19. Stuifbergen AK, Rogers S. The experience of fatigue and strategies of self-care among persons with multiple sclerosis. *Applied Nursing Research.* 1997;10(1):2-10.
20. Vickers MH. Life and work with multiple sclerosis (MS): The role of unseen experiential phenomena on unreliable bodies and uncertain lives. *Illness, Crisis & Loss.* 2009;17(1):7-21.

21. Fisk JD, Pontefract A, Ritvo PG, Archibald CJ, Murray TJ. The impact of fatigue on patients with multiple sclerosis. *Le Journal Canadien des Sciences Neurologiques.* 1994;21(1):9-14.
22. Iriarte J, Subira ML, Castro P. Modalities of fatigue in multiple sclerosis: correlation with clinical and biological factors. *Multiple Sclerosis Journal.* 2000;6(2):124-130.
23. Bakshi R. Fatigue associated with multiple sclerosis: diagnosis, impact and management. *Multiple Sclerosis Journal.* 2003;9(3):219-227.
24. Flensner G, Lindencrona C. The cooling-suit: case studies of its influence on fatigue among eight individuals with multiple sclerosis. *Journal of advanced nursing.* 2002;37(6):541-550.
25. Schwartz CE, Coulthard-Morris L, Zeng Q. Psychosocial correlates of fatigue in multiple sclerosis. *Archives of Physical Medicine and Rehabilitation.* 1996;77(2):165-170.
26. Vercoulen JH, Hommes OR, Swanink CM, et al. The measurement of fatigue in patients with multiple sclerosis. A multidimensional comparison with patients with chronic fatigue syndrome and healthy subjects. *Archives of neurology.* 1996;53(7):642-649.
27. Forbes A, While A, Mathes L, Griffiths P. Health problems and health-related quality of life in people with multiple sclerosis. *Clinical rehabilitation.* 2006;20(1):67-78.
28. Merkelbach S, Sittinger H, Koenig J. Is there a differential impact of fatigue and physical disability on quality of life in multiple sclerosis? *The Journal of Nervous and Mental Disease.* 2002;190(6):388-393.
29. Amato MP, Ponziani G, Rossi F, Liedl CL, Stefanile C, Rossi L. Quality of life in multiple sclerosis: the impact of depression, fatigue and disability. *Multiple sclerosis (Houndmills, Basingstoke, England).* 2001;7(5):340-344.
30. Krupp LB, Alvarez LA, LaRocca NG, Scheinberg LC. Fatigue in multiple sclerosis. *Archives of Neurology.* 1988;45(4):435-437.
31. Olsson M, Lexell J, Soderberg S. The meaning of fatigue for women with multiple sclerosis. *Journal of advanced nursing.* 2005;49(1):7-15.
32. O'Connor AB, Schwid SR, Herrmann DN, Markman JD, Dworkin RH. Pain associated with multiple sclerosis: systematic review and proposed classification. *Pain.* 2008;137(1):96-111.
33. Pompeii LA, Moon SD, McCrory DC. Measures of physical and cognitive function and work status among individuals with multiple sclerosis: a review of the literature. *Journal of occupational rehabilitation.* 2005;15(1):69-84.
34. Crayton H, Heyman RA, Rossman HS. A multimodal approach to managing the symptoms of multiple sclerosis. *Neurology.* 2004;63(11 Suppl 5):S12-18.
35. Rampello A, Franceschini M, Piepoli M, et al. Effect of aerobic training on walking capacity and maximal exercise tolerance in patients with multiple sclerosis: a randomized crossover controlled study. *Physical therapy.* 2007;87(5):545-555.
36. Becker H, Stuifbergen A. What makes it so hard? Barriers to health promotion experienced by people with multiple sclerosis and polio. *Family & community health.* 2004;27(1):75-85.
37. Hubsky EP, Sears JH. Fatigue in multiple sclerosis: guidelines for nursing care. *Rehabilitation Nursing.* 1992;17(4):176-180.
38. Bryant D, Chiaravalloti ND, DeLuca J. Objective Measurement of Cognitive Fatigue in Multiple Sclerosis. *Rehabilitation Psychology.* 2004;49(2):114.
39. Cook KF, Bamer AM, Roddey TS, Kraft GH, Kim J, Amtmann D. A PROMIS fatigue short form for use by individuals who have multiple sclerosis. *Quality of Life Research.* 2012;21(6):1021-1030.
40. Cook KF, Bamer AM, Roddey TS, Kraft GH, Kim J, Amtmann D. Multiple sclerosis and fatigue: understanding the patient's needs. *Physical Medicine & Rehabilitation Clinics.* 2013;24(4):653-661.
41. Nguyen TH, Han HR, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. *Patient.* 2014;7(1):23-35.

42. Lai J-S, Cella D, Choi S, et al. How item banks and their application can influence measurement practice in rehabilitation medicine: A PROMIS Fatigue Item Bank example. *Archives of Physical Medicine and Rehabilitation.* 2011;92(10):S20-S27.
43. Yellen SB, Cella DF, Webster K, Blendowski C, Kaplan E. Measuring fatigue and other anemia-related symptoms with the Functional Assessment of Cancer Therapy (FACT) measurement system. *Journal of Pain and Symptom Management.* 1997;13(2):63-74.
44. Amtmann D, Cook KF, Johnson KL, Cella D. The PROMIS initiative: involvement of rehabilitation stakeholders in development and examples of applications in rehabilitation research. *Archives of Physical Medicine and Rehabilitation.* 2011;92(10 Suppl):S12-19.
45. Chang CH, Cella D, Clarke S, Heinemann AW, Von Roenn JH, Harvey R. Should symptoms be scaled for intensity, frequency, or both? *Palliative Supportive Care.* 2003;1(1):51-60.
46. Kamudoni P, Johns J, Cook KF, et al. Standardizing fatigue measurement in multiple sclerosis: the validity, responsiveness and score interpretation of the PROMIS SF v1.0 - Fatigue (MS) 8a. *Multiple Sclerosis and Related Disorders.* 2021;54:103117.

[Remainder of page intentionally left blank]

APPENDIX 5— SUPPLEMENTAL INFORMATION FOR THE *PROMIS SHORT FORM V1.0 – FATIGUE 10A (FACIT-FATIGUE-10/PROMIS FATIGUE SHORT FORM 10A*)

This appendix includes information specific to the PROMIS Short Form v1.0 – Fatigue 10a (FACIT-Fatigue-10/*PROMIS Fatigue Short Form 10a*) under consideration for qualification as a drug development tool by FDA as a measure of fatigue severity to derive efficacy endpoints in clinical trials for products intended to treat rheumatoid arthritis (RA).

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 et al., 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 et al., 2010b; Singh et al., 2012), patient and clinician input reveal that fatigue is a common, persistent, disabling, and high-priority symptom in RA (Bartlett 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 et al., 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 persons with RA 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 et al., 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 persons with RA not only consider fatigue to be 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, et al., 2011; Pollard et al., 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 overall 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 et al., 2009; Nicklin et al., 2010). Moreover, in response to the increasingly feasible goal of achieving low disease activity and/or remission through intervention (Hewlett, Carr, et al., 2005), ACR, the European League Against Rheumatism (EULAR), and Outcome Measures in Rheumatology (OMERACT) 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. Hence, patients and clinicians agree on the need to include patient-reported assessment of fatigue when evaluating RA treatments.

In RA clinical trials and clinical practice, fatigue has not been assessed as often as physical function, pain, and disease activity, 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 by persons with RA, providers, and researchers. Given the importance persons with RA 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 et al., 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 with RA 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 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 Critical Path Institute personnel, participants included individuals with RA, representatives from the U.S. Food and Drug Administration (FDA), National Institute for Arthritis and Musculoskeletal and Skin Diseases, ACR, OMERACT, and EULAR. The conclusion of the workshop was that the greatest gap in terms of PRO measurement in RA treatment trials was demonstration of the incremental value of fatigue as a key symptom beyond the ACR criteria for assessing clinical 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 et al., 2014; Strand et al., 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.

Overview of the PROMIS Short Form v1.0 – Fatigue 10a (FACIT-Fatigue-10/*PROMIS Fatigue Short Form 10a*)

Description

The PROMIS Short Form V1.0 – Fatigue 10a is a 10-item fixed-length short form assessing fatigue severity in terms of experience and impact. It has a recall period of past 7 days and includes a 5-point verbal rating scale ranging from "Not at all" to "Very much." The short form reflects the following concepts: feelings of energy/fatigue (2 items) tiredness/sleepiness (2 items), activity limitation (5 items), and frustration with limitations (1 item). The PROMIS Short Form V1.0 – Fatigue 10a also enables the assessment of both fatigue experience and impact within a single brief measure, producing a score that locates the respondent on a unidimensional fatigue T-score metric.

This metric is linked to all the items in the PROMIS Fatigue Item Bank (and other PROMIS Fatigue short forms). Scores on the PROMIS Short Form V1.0 – Fatigue 10a are reported as a T-score metric (mean = 50 and SD = 10), with higher scores reflecting greater fatigue.

Reporter

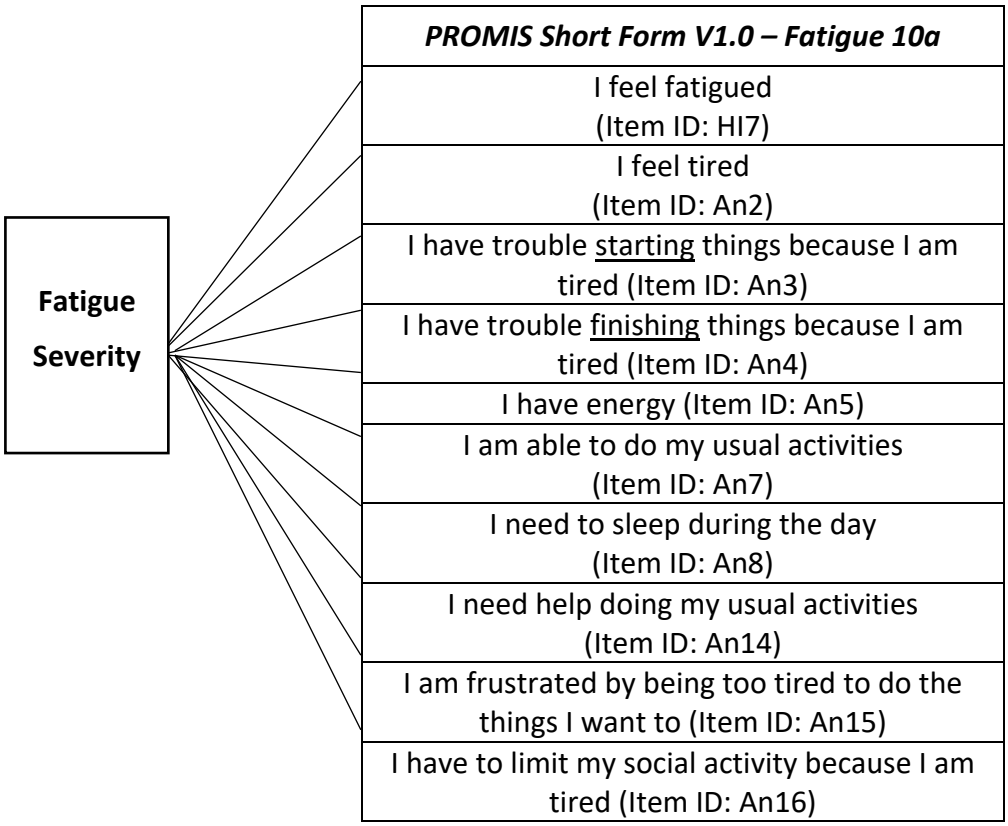
The PROMIS Short Form V1.0 – Fatigue 10a is a self-report measure. Individuals rate themselves on statements ("items") about fatigue. Respondents choose the response option that most accurately describes their experience of fatigue and its impact.

Conceptual framework

The conceptual framework for the assessment of fatigue in RA based on the PROMIS Short Form V1.0 – Fatigue 10a is shown in Figure 1. The PROMIS Short Form V1.0 – Fatigue 10a measures the unidimensional construct of fatigue severity consisting of 10 items that generate a total score. The content consists of the following concepts: sleep during the day (1 item), usual activities (2 items), having energy (1 item), feeling tired (1 item), feeling fatigued (1 item), trouble starting things (1 item), trouble finishing things (1 item), frustrated by fatigue interference with activities (1 item), and limiting social activity due to tiredness (1 item).

[Remainder of page intentionally left blank]

Figure 1: Conceptual framework



Development and Evaluation of the PROMIS Short Form v1.0 – Fatigue 10a (FACIT-Fatigue-10/*PROMIS Fatigue Short Form 10a*)

Development

The PROMIS Short Form V1.0 – Fatigue 10a is one of several available short forms made from the PROMIS Fatigue Item Bank v1.0. This item bank 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 7 days using 5-level verbal rating scales.

The PROMIS Short Form V1.0 – Fatigue 10a is comprised of 10 of the 13 items in the Functional Assessment of Chronic Illness Therapy – Fatigue (FACIT-F) Scale, 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 Scale 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 et al., 2012). Thus, the FACIT Fatigue Scale provides a strong foundation to support the PROMIS Short Form V1.0 – Fatigue 10a given the former's excellent measurement properties.

Evidence of content validity

The PROMIS Short Form V1.0 – Fatigue 10a was developed based on qualitative work with individuals diagnosed with RA. Kaiser and colleagues evaluated the content validity of the 13-item FACIT Fatigue Scale in a sample of participants (N = 17) with moderately- to highly-active RA. Semi-structured, concept elicitation interviews with participants explored the importance, experience, and impact of fatigue in daily living (Kaiser et al., 2016). 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 items were relevant and important in the experience of fatigue among RA participants. However, Kaiser and colleagues 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 Short Form V1.0 – Fatigue 10a. In addition, qualitative analyses did not identify additional fatigue-related concepts that needed to be added to the PROMIS Short Form V1.0 – Fatigue 10a, further confirming sufficient content coverage for participants with RA. These qualitative results provide strong support for the coverage, relevance, and content validity of 10 of the 13 FACIT Fatigue Scale items in RA leading to the derivation of the PROMIS Short Form V1.0 – Fatigue 10a (Kaiser et al., 2016).

Measurement properties and psychometric evaluation

Secondary analyses of two trials were conducted for the psychometric evaluation of the PROMIS Short Form V1.0 – Fatigue 10a: *Safety Trial of Adalimumab in Rheumatoid Arthritis* (STAR) (Furst et al., 2003) and *A Randomized, Double-Blind, Placebo- and Active-Controlled, Phase 3 Study Evaluating the Efficacy and Safety of Baricitinib in Patients with Moderately to Severely Active Rheumatoid Arthritis Who Have Had an Inadequate Response to Methotrexate Therapy* (RA-BEAM) (Taylor et al., 2017).

Descriptive statistics included item and score descriptive statistics, frequency distributions of item response and overall scores, floor and ceiling effects, and percentage of missing responses. To describe the distribution of items in the PROMIS Short Form V1.0 – Fatigue 10a, the frequency and percentage of each response option selected for each item was calculated, at the Baseline and Week 24 time points in the STAR trial data. A threshold of 35% (approximately one-third) of the sample selecting either the highest or lowest response was used to indicate a floor or ceiling effect for that item. Two items (AN14 and AN16) were identified as having a potential ceiling effect; there were no scale-level floor or ceiling effects (> 15% of the sample). Missing data at the item level was evaluated by calculating the frequency and percentage of missing data per item at each timepoint. This evaluation determined that the STAR trial had little missing data. With regard to inter-item relationships and dimensionality analysis, relationships between individual items were assessed by calculating inter-item Spearman's rank correlations between each pair of items in the STAR trial. Two item pairs (HI7 and AN2; AN3 and AN4) were found to have high inter-item correlations, but these items were ultimately retained. Unidimensionality for the PROMIS Short Form V1.0 – Fatigue 10a was supported by McDonald's omega hierarchical and explained common variance (ECV) values from an exploratory bifactor analysis and adequate model fit indices of a confirmatory factor model. Differential item functioning (DIF) analysis between the RA sample in the STAR trial data and a general population sample indicated that no DIF was found for the items in the PROMIS Short Form V1.0 – Fatigue 10a.

Further analysis of the STAR trial data confirmed the reliability and construct validity of the PROMIS Short Form V1.0 – Fatigue 10a. The Cronbach's alpha coefficient for the PROMIS Short Form V1.0 – Fatigue 10a showed excellent internal consistency reliability ($\alpha = 0.91$). Test-retest reliability analyses were conducted on a sub-sample of STAR trial participants who did not show clinically meaningful change in 2 measures of fatigue symptoms – the Multidimensional Assessment of Fatigue (MAF; a measure developed to evaluate self-reported fatigue in persons with RA) and the SF-36 Vitality subscale (a general measure of energy/fatigue), between 2 timepoints. Additionally, this analysis calculated summary scores and assessed the intraclass correlation coefficients (ICCs) between timepoints. The ICC of 0.91 indicated excellent test-retest reliability for the PROMIS Short Form V1.0 – Fatigue 10a. Convergent validity analyses yielded strong Spearman's rank correlations between the PROMIS Short Form V1.0 – Fatigue 10a and the SF-36 Vitality subscale and MAF at both Baseline and Week 24 assessments. The known-groups validity analysis based on both the STAR and the RA-BEAM trial data examined differences in PROMIS Short

Form V1.0 – Fatigue 10a scores between subgroups of participants with known status. The STAR trial population was segmented into groups based on their Patient Global Assessment of Disease Activity (PGA) scores, Clinical Disease Activity Index (CDAI) scores, RA treatment response (American College of Rheumatology [ACR] response criteria) scores, and fatigue severity (MAF) scores, and compared mean PROMIS Short Form V1.0 – Fatigue 10a scores across these groups using a one-way analysis of variance (ANOVA). Results indicated that most sub-groups had significantly different scores, providing evidence of known-groups validity for the PROMIS Short Form V1.0 – Fatigue 10a.

Longitudinal evaluation of the measurement properties of the PROMIS Short Form V1.0 – Fatigue 10a using RA-BEAM trial data indicated that the measure is sensitive to change. The ability to detect change was tested using 4 clinical anchors available in the RA-BEAM trial data: the SF-36 v2 (Acute Version) VT9e and VT9i items, the SF-36 v2 (Acute Version) Vitality subscale score, and the Severity of Worst Tiredness item. Differences in these anchors were calculated from Baseline to Week 24 follow-up (Baseline to Week 12 for the Severity of Worst Tiredness item). The first 2 anchors were categorized into improved (change score is larger than 0), same (change score is 0), and worsened (change score is smaller than 0). Across all anchors, participants with an improved score on the anchor also had a significantly lower PROMIS Short Form V1.0 – Fatigue 10a score; similarly, participants with a worsened anchor score had a significantly smaller PROMIS Short Form V1.0 – Fatigue 10a change score. Finally, analysis of the RA-BEAM trial using the same anchors yielded estimates of meaningful within-person change. The range of estimates based on 1-level change in the 4 anchors ranged from -7 06 to -11.68 and ranged from -11.28 to -15.88 based on 2-level change in the 4 anchors for improvement in fatigue. Based on these results, the initial recommendations for meaningful within-person change thresholds are ranges of 5 to 8 points for improvement and 2 to 5 points for worsening.

Taken together, these analyses demonstrate that the PROMIS Short Form V1.0 – Fatigue 10a can be incorporated into clinical trials for longitudinal evaluation given the measure's ability to detect change over time, as well as its strong psychometric properties.

[Remainder of page intentionally left blank]

Handling of missing data

- **Form-level missing data.** Form-level missing data refers to a respondent missing an entire PRO assessment for a given time point (e.g., a given evening for the PROMIS Short Form V1.0 – Fatigue 10a). In general, form-level data may be missing due to a respondent's early withdrawal from the study, the inability to evaluate an endpoint at a particular time point, or non-compliance. Imputation is not recommended for PROMIS Short Form V1.0 – Fatigue 10a form-level missing data. If a participant does not complete the PROMIS Short Form V1.0 – Fatigue 10a, such as due to attrition in longitudinal studies due to forgetting to complete an individual assessment, his or her PROMIS Short Form V1.0 – Fatigue 10a score should not be computed for that timepoint.
- **Item-level missing data.** Administrators may decide to allow respondents to skip individual items within the PROMIS Short Form V1.0 – Fatigue 10a. If respondents are allowed to skip items, missing data at the item level may be present. Based on missing data simulation analyses, scores should only be calculated if data for 4 or more items in a given assessment are available using response pattern scoring from HealthMeasures.net. If 7 or more items are missing, the assessment itself will be considered missing and will be handled under the form-level missing data procedures.

[Remainder of page intentionally left blank]

References for Appendix 5

Alten, R., Pohl, C., Choy, E. H., Christensen, R., Furst, D. E., Hewlett, S. E., Leong, A., May, J. E., Sanderson, T. C., Strand, V., Woodworth, T. G., Bingham, C. O., 3rd, & 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. https://doi.org/10.3899/jrheum.110400

Arnett, F. C., Edworthy, S. M., Bloch, D. A., Mcshane, D. J., Fries, J. F., Cooper, N. S., Healey, L. A., Kaplan, S. R., Liang, H. L., Luthra, H. S., Medsger, T. A., Mitchell, D. M., Neustadt, D. H., Pinals, R. S., Schaller, J. G., Sharp, J. T., Wilder, R. L., & 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., Hewlett, S., Bingham, C. O., 3rd, Woodworth, T. G., Alten, R., Pohl, C., Choy, E. H., Sanderson, T., Boonen, A., Bykerk, V., Leong, A. L., Strand, V., Furst, D. E., Christensen, R., & 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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1136/annrheumdis-2018-eular.3077

Berthelot, J. M., De Bandt, M., Morel, J., Benatig, F., Constantin, A., Gaudin, P., Le Loet, X., Maillefert, J. F., Meyer, O., Pham, T., Saraux, A., Solau-Gervais, E., Spitz, E., Wendling, D., Fautrel, B., Guillemin, F., & 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. https://doi.org/10.1136/ard.2011.150656

Bingham, C. O., 3rd, Alten, R., Bartlett, S. J., Bykerk, V. P., Brooks, P. M., Choy, E., Christensen, R., Furst, D. E., Hewlett, S. E., Leong, A., May, J. E., Montie, P., Pohl, C., Sanderson, T. C., Strand, V., Woodworth, T. G., Guillemin, F., & 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. https://doi.org/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. https://doi.org/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*.

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. https://doi.org/10.1007/s10067-011-1887-y

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. https://www.ncbi.nlm.nih.gov/pubmed/15868614

Cohen, S. B., Emery, P., Greenwald, M. W., Dougados, M., Furie, R. A., Genovese, M. C., Keystone, E. C., Loveless, J. E., Burmester, G. R., Cravets, M. W., Hessey, E. W., Shaw, T., Totoritis, 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. https://doi.org/10.1002/art.22025

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

Emery, P., Blanco, R., Cocco, J. M., Chen, Y. C., Gaich, C. L., DeLozier, A. M., de Bono, S., Liu, J. J., Rooney, T., Chang, C. H. C., & 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). https://doi.org/UNSP e000410 10.1136/rmdopen-2016-000410

Felson, D. T., Anderson, J. J., Boers, M., Bombardier, C., Chernoff, M., Fried, B., Furst, D., Goldsmith, C., Kieszak, S., Lightfoot, R., Paulus, H., Tugwell, P., Weinblatt, M., Widmark, R., Williams, J., & 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. https://doi.org/10.1002/art.1780360601

Furst, D. E., Schiff, M. H., Fleischmann, R. M., Strand, V., Birbara, C. A., Compagnone, D., Fischkoff, S. A., & 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. https://www.ncbi.nlm.nih.gov/pubmed/14719195

Gossec, L., Dougados, M., Rincheval, N., Balanescu, A., Boumpas, D. T., Canadelo, S., Carmona, L., Daures, J. P., de Wit, M., Dijkmans, B. A., Englbrecht, M., Gunendi, Z., Heiberg, T., Kirwan, J. R., Mola, E. M., Matucci-Cerinic, M., Otsa, K., Schett, G., Sokka, T., . . . 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. https://doi.org/10.1136/ard.2008.100271

Gossec, L., Paternotte, S., Aanerud, G. J., Balanescu, A., Boumpas, D. T., Carmona, L., de Wit, M., Dijkmans, B. A., Dougados, M., Englbrecht, M., Gogus, F., Heiberg, T., Hernandez, C., Kirwan, J. R., Mola, E. M., Cerinic, M. M., Otsa, K., Schett, G., Scholte-Voshaar, M., . . . 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. https://doi.org/10.1136/ard.2010.142901

Helmick, C. G., Felson, D. T., Lawrence, R. C., Gabriel, S., Hirsch, R., Kwoh, C. K., Liang, M. H., Kremers, H. M., Mayes, M. D., Merkel, P. A., Pillemer, S. R., Reveille, J. D., Stone, J. H., & 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. https://doi.org/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. https://doi.org/10.1002/msc.3

Hewlett, S., Chalder, T., Choy, E., Cramp, F., Davis, B., Dures, E., Nicholls, C., & Kirwan, J. (2011). Fatigue in rheumatoid arthritis: time for a conceptual model. *Rheumatology (Oxford)*, *50*(6), 1004-1006. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1002/art.22611

Hewlett, S., Sanderson, T., May, J., Alten, R., Bingham, C. O., 3rd, Cross, M., March, L., Pohl, C., Woodworth, T., & 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. https://doi.org/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. https://doi.org/10.1002/art.21562

Janoudi, N., Omran, N., Hussain, W., Al-osaimi, H., Baamer, M., Siddiqui, M. I., Fathaldin, O., & 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. https://doi.org/10.2147/Oarrr.S142411

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). https://doi.org/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*. 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., Zamora, J. V., Cobos, J. A. C., Rooney, T., de Bono, S., Arora, V., Linetzky, B., & 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 RA-BEAM study. *Annals of the Rheumatic Diseases*, *76*(11), 1853-1861. https://doi.org/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. https://www.ncbi.nlm.nih.gov/pubmed/17477481

Kirwan, J. R., Minnock, P., Adebajo, A., Bresnihan, B., Choy, E., de Wit, M., Hazes, M., Richards, P., Saag, K., Suarez-Almazor, M., Wells, G., & Hewlett, S. (2007). Patient perspective: fatigue as a recommended patient centered outcome measure in rheumatoid arthritis. *Journal of Rheumatology*, *34*(5), 1174–1177.

Kirwan, J. R., Newman, S., Tugwell, P. S., et al., Wells, G., Hewlett, S., Idzera, L., Laslo, B., March, L., Minnock, P., Montie, P., Nicklin, J., Rader, T., Richards, P., T., S., Suarez-Almazor, M., Tanjong-Ghogomu, E., Ueffing, E., & 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. https://doi.org/10.3899/jrheum.090360

Lawrence, R. C., Felson, D. T., Helmick, C. G., Arnold, L. M., Choi, H., Deyo, R. A., Gabriel, S., Hirsch, R., Hochberg, M. C., Hunder, G. G., Jordan, J. M., Katz, J. N., Kremers, H. M., Wolfe, F., & 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. https://doi.org/10.1002/art.23176

Mayoux-Benhamou, A., Quintrec, J. S. G. L., Ravaud, P., Champion, K., Dernis, E., Zerkak, D., Roy, C., Kahan, A., Revel, M., & 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. <Go to ISI>://WOS:000253202800007

Mease, P. J., Revicki, D. A., Szechinski, J., Greenwald, M., Kivitz, A., Barile-Fabris, L., Kalsi, J., Eames, J., & 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. <Go to ISI>://WOS:000252255700007

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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1002/acr.22374

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. https://doi.org/10.1093/rheumatology/kel021

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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1002/acr.20034

Singh, J. A., Furst, D. E., Bharat, A., Curtis, J. R., Kavanaugh, A. F., Kremer, J. M., Moreland, L. W., O'Dell, J., Winthrop, K. L., Beukelman, T., Bridges, S. L., Jr., Chatham, W. W., Paulus, H. E., Suarez-Almazor, M., Bombardier, C., Dougados, M., Khanna, D., King, C. M., Leong, A. L., . . . 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. https://doi.org/10.1002/acr.21641

Smolen, J. S., Aletaha, D., Bijlsma, J. W., Breedveld, F. C., Boumpas, D., Burmester, G., Combe, B., Cutolo, M., de Wit, M., Dougados, M., Emery, P., Gibofsky, A., Gomez-Reino, J. J., Haraoui, B., Kalden, J., Keystone, E. C., Kvien, T. K., McInnes, I., Martin-Mola, E., . . . 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. https://doi.org/10.1136/ard.2009.123919

Strand, V., Kosinski, M., Chen, C. I., Joseph, G., Rendas-Baum, R., Graham, N. M. H., van Hoogstraten, H., Bayliss, M., Fan, C. P., Huizinga, T., & 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*. https://doi.org/ARTN 198 10.1186/s13075-016-1096-9

Strand, V., Rentz, A. M., Cifaldi, M. A., Chen, N. J., Roy, S., & Revicki, D. (2012). Health-related 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. https://doi.org/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*.

Taylor, P. C., Keystone, E. C., van der Heijde, D., Weinblatt, M. E., Del Carmen Morales, L., Reyes Gonzaga, J., Yakushin, S., Ishii, T., Emoto, K., Beattie, S., Arora, V., Gaich, C., Rooney, T., Schlichting, D., Macias, W. L., de Bono, S., & Tanaka, Y. (2017). Baricitinib versus Placebo or Adalimumab in Rheumatoid Arthritis. *N Engl J Med*, *376*(7), 652-662. https://doi.org/10.1056/NEJMoa1608345

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. https://doi.org/10.1186/s13075-015-0911-z

van Tuyl, L., Sadlonova, M., Davis, B., Flurey, C., Goel, N., Hewlett, S. E., Hill, C. L., Hoogland, W., Kirwan, J. R., van Schaardenburg, D., Scholte-Voshaar, M., Smolen, J. S., Stamm, T., Wells, G. A., & Boers, M. (2016). Remission in Rheumatoid Arthritis: Working Toward Incorporation of the Patient Perspective at OMERACT 12. *Journal of Rheumatology*, *43*(1), 203-207. https://doi.org/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. https://doi.org/10.3899/jrheum.110394

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.

[Remainder of page intentionally left blank]