



Published in final edited form as:

Qual Life Res. 2010 November ; 19(9): 1311–1321. doi:10.1007/s11136-010-9694-5.

Patient-Reported Outcomes Measurement Information System (PROMIS) Domain Names and Definitions Revisions: Further Evaluation of Content Validity in IRT-derived Item Banks

William T. Riley,

National Heart Lung and Blood Institute

Nan Rothrock,

Northwestern University

Bonnie Bruce,

Stanford University

Christopher Christodolou,

Stony Brook University

Karon Cook,

University of Washington

Elizabeth A. Hahn, and

Northwestern University

David Cella

Northwestern University

Abstract

Purpose—Content validity of patient-reported outcomes (PROs) is evaluated primarily during item development, but subsequent psychometric analyses, particularly for item-response theory (IRT)-derived scales, often result in considerable item pruning and potential loss of content. After selecting items for the PROMIS banks based on psychometric and content considerations, we invited external content expert reviews of the degree to which the initial domain names and definitions represented the calibrated item bank content.

Methods—A minimum of four content experts reviewed each item bank and recommended a domain name and definition based on item content. Domain names and definitions then were revealed to the experts who rated how well these names and definitions fit the bank content and provided recommendations for definition revisions.

Results—These reviews indicated that the PROMIS domain names and definitions remained generally representative of bank content following item pruning, but modifications to two domain names and minor to moderate revisions of all domain definitions were needed to optimize fit with the item bank content.

Conclusions—This reevaluation of domain names and definitions following psychometric item pruning, although not previously documented in the literature, appears to be an important procedure for refining conceptual frameworks and further supporting content validity.

Keywords

content validity; conceptual framework; domain definition; item-response theory

Patient-reported outcome (PRO) measurement development has benefited from recent efforts to outline best practices for establishing content validity. These best practices for determining the extent to which an instrument sufficiently represents all facets of the relevant constructs have emphasized the importance of developing a conceptual model that clearly defines the constructs of interest [1,2] and utilizing patient input in the development of item content and the conceptual model [1,3,4]. Best practices also have been offered on the use of qualitative research methodology to obtain patient input for the evaluation of content validity [5].

The Patient-Reported Outcomes Measurement Information System (PROMIS) developed item banks consistent with current content validity guidance, including conceptual model development and inclusion of the patient perspective. The PROMIS domain framework and definitions resulted from extensive literature review, archival data analyses, and a modified Delphi process with content experts [6,7]. Patient feedback from numerous focus groups refined these conceptual definitions, generated content for new items, and documented saturation of the construct [8,9]. Cognitive interviews solicited feedback on item clarity [10]. Item pools were tested in a large sample of general population respondents, augmented by clinical samples, and item response theory (IRT) methodology was used to select and calibrate items for the PROMIS item banks [11-14].

Item pruning, the elimination of items based on psychometric considerations, is inherent in the development of IRT-based item banks. Items may be removed for a variety of psychometric concerns including local dependence (correlated residuals), differential item functioning (DIF), inadequate unidimensionality, lack of monotonicity, and poor IRT model fit [11,15]. For example, 56 items from the PROMIS depression item pool were tested, but only half (28 items) were retained [13]. Although item pruning is consistent with IRT test development, it is also a potential threat to the content validity of the resulting item bank since item content generated from patient and expert feedback may be lost as items with poor psychometric properties are removed. During the PROMIS item bank selection process, domain content experts worked with psychometricians to minimize loss of representative item content due to psychometric concerns. A few items with less than optimal psychometric properties were retained because they uniquely covered a relevant facet of the domain, but a substantial number of items performed too poorly to be retained and were excluded from the calibrated item banks.

To address the effects of content loss from eliminated items on content validity, the ISPOR PRO Task Force has recommended that patient interviews or focus groups be used to determine the importance of omitted versus retained items [4]. However, given the comprehensive initial item pool development process in PROMIS [7-8], new items generated to replace omitted content deemed important by this approach likely would have similar psychometric limitations as the omitted items. Assuming that new items with potentially better psychometric properties could be generated, the testing and calibration of new items with the existing items in a large and diverse sample of respondents would take considerable time and resources. In the interim, the initial domain names and definitions may not accurately represent the currently retained items, thus potentially misleading users as to the content represented by the current item banks. Therefore, we decided to first revise the domain names and definitions to better represent the retained item banks.

Consistent with the ISPOR Task Force recommendation, PRO guidance regarding content validity has focused predominately on the initial item development phase and on developing items that cover the facets or attributes of the conceptual definition [3-5]. The iterative development of PROs, however, includes modifying not only the item content consistent with the concept, but also the concept consistent with the psychometric findings. The Food and Drug Administration (FDA) PRO development model indicates that psychometric findings may lead to revision of the concept [3], and the Mayo/FDA PRO Consensus Meeting Group elaborated on this process of concept refinement, indicating that empiric evidence from psychometric analyses should be used to modify the conceptual framework [2]. This complementary and iterative process of conceptual refinement includes not only generating items that cover all of the purported facets of the concept, but also using the psychometric data to revise the concept consistent with the retained items. Therefore, we report in this study a procedure for revising domain definitions consistent with the retained item content by asking content experts to review the item banks and make recommendations about revising the PROMIS domain names and definitions based on the items retained in these banks.

Method

Participants

PROMIS domain groups (physical function, pain, fatigue, emotional distress, social health, sleep) identified content experts with experience developing and validating instruments in the domain or conducting clinical research in which the domain is a primary outcome. To participate, experts could not be supported by the PROMIS cooperative agreement, but could have served as consultants to the project in a limited capacity. We identified potential content experts from a number of sources, including developers of legacy scales from whom we had asked permission to use their scales for item pool development and/or concurrent validity testing. Approximately eight experts for each PROMIS domain were identified with the goal of at least four per domain completing the review. This expert feedback was considered exempt from Institutional Review Board review.

Expert Feedback Procedures

The identified content experts were contacted by email, described the study purpose, and asked to review attached item banks and provide online feedback. Nonrespondents were recontacted after approximately one month. As needed, additional experts were contacted until a minimum of four experts in each domain provided responses. Expert reviews were performed independently of each another.

To balance burden and utilize expertise across similar areas, content experts reviewed either one large item bank or multiple smaller item banks as follows:

Bank A:	Physical Function (124 items)
Bank B:	Fatigue (95 items)
Banks C & D:	Pain Behavior (39 items), Pain Impact (41 items)
Banks E, F, & G:	Depression (28 items), Anxiety (29 items), Anger (28 items)
Banks H & I:	Satisfaction with Participation in Social Roles (14 items), Satisfaction with Participation in Discretionary Social Activities (12 items)
Banks J & K:	Sleep Disturbance (27 items), Wake Disturbance (16 items)

Experts received and reviewed only the banks assigned to them, but all experts also reviewed and provided feedback on the 10 PROMIS Global Health items [16].

Prior to revealing the existing domain name and definition, experts were asked based on the item bank content alone (i.e. blind feedback) to: a) provide a 1-4 word name for the bank, and (b) describe and define in a few sentences what the item bank measures. The domain names and definitions were then revealed, and experts were asked to respond to the following questions.

1. How well does this name reflect the item content? (Not at all, A little bit, Somewhat, Quite a bit, Very much)
2. How well does this definition reflect the item content being measured? (Not at all, A little bit, Somewhat, Quite a bit, Very much)
3. Is there item content that is not adequately reflected in the definition? (Yes, No)
4. How could the definition be expanded to fully capture the item content of this bank?
5. Is there item content missing that the bank definition suggests is present in the bank? (Yes, No)
6. How could the definition be narrowed to accurately reflect the content of the bank?
7. Please provide any additional feedback you have on how to improve the name or definition of this item bank.

All experts then provided feedback on the PROMIS Global Health items.

Experts completed a short sociodemographic questionnaire including information on English language background and years of experience in the domain area. Each expert received a \$200 honorarium for participation. No further feedback was obtained. PROMIS Domain Group and Steering Committee Review Procedures.

Representatives from each domain group reviewed and summarized the expert feedback for their respective PROMIS domain groups to consider in revising domain names or definitions. Domain groups considered the quantitative and qualitative feedback subjectively (i.e. no predetermined criteria for revision) and revised domain names and definitions. These revised definitions were presented to the PROMIS Steering Committee for discussion and approval. Definitions were further revised by study investigators (WR, NR) for format and content consistency across domains, and were then reviewed and approved by the respective domain chairs and by the PROMIS Steering Committee.

During the expert review process, the social domain group received supplemental funding to further develop and test social domain items. As a result, the social domain deferred any name or definition revisions until further item bank development was completed. Therefore, social domain expert feedback is included only for the global items.

Results

Participants

Thirty-five participants, 23 males and 12 females, provided expert review. One was Hispanic and one was African-American. Twenty-eight were Ph.D.s and seven were M.D.s. Participants had a mean of 27 (SD = 9) years of experience in the domain area. Thirty-four of the 35 indicated English as their first language (see Table 1). Eighty-three percent either had no prior contact with PROMIS (21/35) or had only contributed legacy items or scales to PROMIS (8/35). The remaining 6 experts had served as consultants to PROMIS in some limited capacity.

Physical Function (Bank A)

Blinded to the “physical function” domain name, three of the six expert reviewers provided “physical function”, two provided “physical activity”, and one provided “PROMIS Health Assessment Questionnaire” as the domain name. After unblinding, five indicated that the name reflected the item bank “very much” (1 rated “quite a bit”). Blinded definitions for this bank included, “Capacity to do a large number of physical activities,” “wide range of usual daily physical activities, exercise, and household chores,” and “measures the participant’s current ability to perform particular tasks that involve use of limbs or core and coordinated movements.” Two of the experts indicated that there was item content not adequately reflected in the physical function definition, with one commenting that the bank was missing goal attainment scaling. Although asked how the *definition* could be expanded to fully capture the item bank content, the experts in all domains often responded instead with how the *item content* could be expanded to fully capture their definition of the domain. None of the respondents reported that the physical function definition needed to be narrowed to accurately reflect the bank content. (For initial definitions reviewed by the content experts, see Table 3).

Fatigue (Bank B)

Of the six experts, three provided “fatigue”, two provided fatigue plus descriptors (e.g. “fatigue assessment scale”, “fatigue frequency and severity”), and one provided “PROMIS Item Bank B” as the domain name. After unblinding, four of the six indicated that the name reflected the item bank “very much” (1 “quite a bit”, 1 “somewhat”). Blinded definitions included, “assessment of symptoms of subjective fatigue and excessive tiredness, “fatigue severity and fatigue interference”, and “supposed to assess fatigue but confuses tiredness, exhaustion, and sleepiness in this construct.” One expert indicated that the item content was not adequately reflected in the fatigue definition, but repeated the concern noted above about the definition being too broad. One respondent noted that the fatigue definition needed to be narrowed to accurately reflect the bank content, specifically noting that duration of fatigue is not as well represented in the bank as frequency of fatigue.

Pain Behavior (Bank C)

The four experts provided “pain behavior,” “pain responses scale,” “pain effects,” and “pain-related affective distress” as the domain name. After unblinding, two of the four indicated that the name reflected the item bank “very much” (1 “quite a bit”, 1 “a little bit”). Blinded definitions included, “how an individual with pain responds to pain,” “observable behavior associated with pain,” “pain-related affective distress,” “broad set of behaviors that patients may express when in pain.” Two of the experts indicated that the item content was not adequately reflected in pain behavior definition. One noted that “pain behavior” is not familiar to many pain specialists, and another indicated that the definition needed to represent a balance of expressing, avoiding, minimizing, and reducing pain. None of the respondents reported that the pain behavior definition needed to be narrowed. One expert recommended less prominence of “pain behavior as communication” since communication is not always understood broadly to include inadvertent, unintentional, and/or unrecognized communication.

Pain Impact (Bank D)

Of the four experts, three provided “pain interference”, and one provided “pain-related interference with functioning” as the domain name. After unblinding, one of the four indicated that the name reflected the item bank “very much” (3 “quite a bit”). Blinded definitions included, “various aspects of pain-related interference with functioning,” “degree to which pain interferes with various aspects of a patients [sic] life,” “how pain interferes

with various activities and states,” “pain interference with daily activities.” One expert indicated that the item content was not adequately reflected in the definition but stated that “the definition seems broad enough.” Two respondents indicated that the pain impact definition needed to be narrowed to accurately reflect the content of the bank, noting only one sleep item in the bank.

Depression (Bank E)

Of the nine experts, five provided “depression” as the domain name. The others provided “dysphoric mood”, “components of depression”, “depression symptoms”, and “depression, discouragement, demoralization.” After unblinding, five indicated that the name reflected the item bank “very much” (2 “quite a bit”, 2 “somewhat”). Blinded definitions included, “depressive symptoms and mood but without vegetative symptoms,” symptoms of depressed mood and the trait of negative affectivity,” and “physical, emotional, and social aspects of depression with emphasis on feeling states, withdrawal from others, and pessimism about the future.” Four experts indicated that the item content was not adequately reflected in the definition, noting that the bank emphasized negative affect much more than positive affect and that the definition could better reflect this emphasis. Two respondents indicated that the definition needed to be narrowed to accurately reflect the content of the bank, with one noting that there was only one item related to indecisiveness despite a reference to “information-processing deficits” in the definition.

Anxiety (Bank F)

Of the nine experts, six provided “anxiety” in some derivation (e.g. “general fear and anxiety”, “anxiety and fear,” “anxiety symptoms”) as the domain name. Three respondents included “panic” or “phobic” in the anxiety name provided. After unblinding, five indicated that the name reflected the item bank “very much” (3 “quite a bit”, 1 “somewhat”). Blinded definitions included, “anxiety, negative emotions, worry, and panic/fear,” “emotional/ affective and physical components of anxiety,” and “anxious/fearful mood and somatic symptoms of anxious arousal.” Seven indicated that the item content was not adequately reflected in the definition, and comments focused primarily having only one behavioral avoidance item (“I avoided public places and activities”) in the bank. Two respondents indicated that the definition needed to be narrowed to accurately reflect bank content, and reiterated the underrepresentation of behavioral avoidance.

Anger (Bank G)

All nine respondents provided some derivation of “anger” (e.g. “anger and irritability”, “anger and hostility”, “anger and frustration tolerance”) as the domain name. After unblinding, eight indicated that the name reflected the item bank “very much” (1 “quite a bit”). Blinded definitions included, “basic emotion of anger/hostility,” “tendency toward angry, frustrative affect”, and “perceptions of anger, irritation, resentment, and frustration.” One expert indicated that the item content was not adequately reflected in the definition, and that angry behavior appeared underemphasized relative to covert anger. None of the respondents reported that the definition needed to be narrowed.

Sleep Disturbance (Bank J)

Of the five experts, three provided some derivation of “insomnia or sleep disturbance,” and two provided “sleep quality” as the domain name. After unblinding, one of the five indicated that the name reflected the item bank “very much” (3 “quite a bit”, 1 “somewhat”). Blinded definitions included “all aspects of insomnia symptoms including concern about sleep and cognitive arousal,” “nature, extent, and severity of difficulties sleeping at night with an attempt to capture the psychological underpinnings of these difficulties,” and “sleep quality

– the cognitive, emotional, and restorative aspects of the sleep experience.” None of the experts reported that the item content was not adequately reflected in the definition or that the definition needed to be narrowed to accurately reflect item content.

Wake Disturbance (Bank K)

Of the five experts, three provided some derivation of “sleep-related daytime functioning,” and two provided “daytime impairment” or “daytime functioning” as the domain name. After unblinding, one of the five indicated that the name reflected the item bank “very much” (2 “quite a bit”, 2 “somewhat”). Blinded definitions included “negative consequence of disturbed sleep focusing on attention, cognition, and mood,” “impact of the loss of sleep and disturbed sleep on the ability to conduct daily activities and mental health,” and “nature, extent, and severity of daytime functioning that may be impaired following a night with sleep difficulties.” Two indicated that the item content was not adequately reflected in the definition, but the qualitative comments reflected item content that they believed should be in the item bank such as physical performance and interpersonal relationships. None of the respondents indicated that the definition needed to be narrowed.

Global Health Items Feedback

Ten responded “Quite a bit” and 22 responded “Very much” (Mean rating = 4.7, SD = 0.5) for how well the Global Health definition reflected the item content. Most comments were positive (e.g. “Fine as is”, “The name and definition are excellent”, “Seems quite on target and germane”). One expert suggested that “overall health” or “overall health and well-being” might be a better name, and one expert suggested that the definition should more clearly indicate that some of the mental and social items do not exclusively relate to the health impact. Two experts noted that spirituality or spiritual health was missing from the content of the global items.

Ratings of name and definition fit by domains are summarized in Table 2.

Domain Name and Definition Revisions

Consideration of the expert feedback by PROMIS domain groups and steering committee resulted in two domain name changes. The sleep domain group shared expert reviewer concerns with the name “Wake Disturbance”, and changed the domain name to “Sleep-related Impairment.” The pain domain group and the PROMIS SC believe that “Pain Impact” is a broader and more appropriate name for this domain, but the name was changed to “Pain Interference” in response to expert feedback and the acknowledgment that most of the pain research community associates “pain interference” with the content reflected in this bank.

All other PROMIS domain names remained unchanged, although there was considerable debate about changing “depression” to “depressive symptoms” since items related to somatic or vegetative symptoms had been removed from the bank due to poor psychometric fit. The primary concern was that the domain name “depression” might infer that the bank is measuring depression as a psychiatric diagnosis. Technically, “depression” is not the formal name of any psychiatric diagnosis [17], and in common usage the term “depression” describes both clinical and subclinical states of sadness and dysphoria. Based on these considerations, the PROMIS SC decided to retain the domain name “depression” and address the absence of somatic or vegetative symptom content in the domain definition.

PROMIS domain groups revised their domain definitions based on expert feedback. Definitions were further revised for consistency and were approved by the domain chairs and PROMIS SC. The initial and revised definitions are provided in Table 3.

Discussion

To identify potential effects of psychometric pruning on content validity, external experts provided feedback on the congruence of the PROMIS domain names and definitions with the item bank content. This feedback indicated that the domain names and definitions remained generally representative of bank content following psychometric pruning, but that minor to moderate domain name and definition revisions were necessary to better represent item bank content, including specifying underrepresented or missing content that experts expected to be present. These findings illustrate that the review and revision of domain names and definitions following psychometric pruning and item calibration is an important step in supporting the content validity of PROs. Although the item development process is the primary source of content validity evidence, subsequent item pruning can result in loss of item content in the resulting bank. Therefore, modifying the conceptual definition to be consistent with the resulting bank content appears to be an important additional component in establishing content validity, particularly for IRT-derived banks that involve considerable item pruning. Recent PRO guidance includes this conceptual model modification step, but we believe this is the first documentation of a standardized procedure for modifying domain names and definitions of IRT-derived item banks to ensure they accurately reflect the item bank content.

Instead of modifying domain definitions to match item bank content, we could have modified item bank content to match domain definitions. The latter approach is the accepted standard during item development, and the qualitative responses of some experts in this study suggested that they would have preferred to modify item content even when the prescribed task was to modify the definition. Patient and expert feedback on the importance of the omitted items to the domain of interest could have been obtained as per ISPOR PRO Task Force recommendations [4]; however, as noted earlier, this approach is only the first step in a time and labor intensive process of adding new item content to an existing item bank. In the interim, we chose to solicit feedback from content experts to ensure that the domain names and definitions clearly conveyed the content of the banks to clinical researchers and practitioners. Refining the item content based on the conceptual definition and refining the conceptual definition based on psychometric findings are complementary and iterative.

When major deviations from the hypothesized conceptual model are found, it may be appropriate to focus first on generating and testing additional item content before attempting to revise the conceptual definitions to match the retained items. In our case, the factor structure of the PROMIS social domain items did not fit well with our hypothesized conceptual framework, so instead of revising the concept, we chose first to generate and test additional items in this domain. For most item bank development, however, seeking expert feedback to revise the conceptual definitions following item banking can ensure optimal fit between the domain definition and retained item bank content.

Several improvements to this domain name and definition review procedure should be considered. First, a small percentage of participants had prior experience with the PROMIS initiative, and including only “independent” experts could minimize response bias. However, even those without prior PROMIS experience were likely exposed to PROMIS information. Therefore, it is unclear how many participants were truly “blind” to the current domain names and definitions, but obtaining domain name and definition input before revealing them appears to have reduced their influence given the range of responses provided. Second, we arbitrarily set a minimum of four expert responses per domain area, but the number of expert responses required for this task is unclear. Setting the number based on a saturation threshold similar to PRO patient focus groups and interviews

procedures [5] may have produced clearer direction for revisions. Third, in addition to asking experts about revising definitions based on item content, we also could have asked about important content omitted from the bank, thus providing direction for future additional item development. Fourth, to obtain feedback efficiently on 11 item banks and the global items, we chose online feedback, but interviews or other interactive methods (e.g. focus groups) could provide richer and more detailed feedback. Finally, in the absence of prior literature, we chose a subjective appraisal of the expert feedback to determine if and how domain names and definitions should be revised. Using the rating information presented in this study, future research may be able to set apriori criteria for determining if a domain name or definition should be revised.

Content expert feedback resulted in improved PROMIS domain names and definitions that more closely match the item content of the calibrated item banks; however, these revised conceptual definitions are likely narrower than the conceptual definitions of some researchers, clinicians, and patients. Consistent with an iterative PRO development model (2-4, 18), the mismatch of a resulting item bank to the initial conceptual definition is an opportunity to better define the content measured by a psychometrically sound item bank and to revise the conceptual framework based on the empiric data. The combination of psychometric pruning of item banks and content expert feedback to revise the names and definitions of these banks provides the basis for iterative conceptual model refinement, narrowing some conceptual definitions while illuminating attributes or facets that might be better conceptualized as a related but separate domain.

Acknowledgments

The Patient-Reported Outcomes Measurement Information System (PROMIS) is a National Institutes of Health (NIH) Roadmap initiative to develop a computerized system measuring patient-reported outcomes in respondents with a wide range of chronic diseases and demographic characteristics. PROMIS was funded by cooperative agreements to a Statistical Coordinating Center (Northwestern University, PI: David Cella, PhD, U01AR52177) and six Primary Research Sites (Duke University, PI: Kevin Weinfurt, PhD, U01AR52186; University of North Carolina, PI: Darren DeWalt, MD, MPH, U01AR52181; University of Pittsburgh, PI: Paul A. Pilkonis, PhD, U01AR52155; Stanford University, PI: James Fries, MD, U01AR52158; Stony Brook University, PI: Arthur Stone, PhD, U01AR52170; and University of Washington, PI: Dagmar Amtmann, PhD, U01AR52171). NIH Science Officers on this project have included Deborah Ader, PhD, Susan Czajkowski, PhD, Lawrence Fine, MD, DrPH, Laura Lee Johnson, PhD, Louis Quatrano, PhD, Bryce Reeve, PhD, William Riley, PhD, Susana Serrate-Sztejn, MD, and James Witter, MD, PhD. This manuscript was reviewed by the PROMIS Publications Subcommittee prior to external peer review. See the web site at www.nihpromis.org for additional information on the PROMIS cooperative group.

References

1. Lohr K, for the Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research*. 2002; 11:193–205. [PubMed: 12074258]
2. Rothman ML, Beltran P, Cappelleri JC, Lipscomb J, Teschendorf B, the Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. Patient-Reported Outcomes: Conceptual Issues. *Value in Health*. 2007; 10(Suppl. 2):S66–S75. [PubMed: 17995476]
3. U.S. Department of Health and Human Services. Patient-reported outcome measures: Use in medical product development to support labeling claims. Rockville, MD: 2006. Food and Drug Administration. Guidance for industry. [http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidance s/ucm071975.pdf](http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidance%20for%20Industry/UCM071975.pdf)
4. Rothman M, Burke L, Erickson P, Kline Leidy N, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: The ISPOR Good Research Practices for evaluation and documenting content validity for the use of existing instruments and their modification PRO Task Force report. *Value in Health*. 2009; 8:1075–1083. [PubMed: 19804437]

5. Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: Developing best practices based on science and experience. *Quality of Life Research*. 2009; 18:1263–1278. [PubMed: 19784865]
6. Fries JF, Bruce B, Cella D. The promise of PROMIS: The new sciences behind patient-reported outcomes. *Clinical and Experimental Rheumatology*. 2005; 23:S53–S57. [PubMed: 16273785]
7. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS). Progress of an NIH Roadmap cooperative during its first two years. *Medical Care*. 2007; 45(Suppl. 1):S3–S11. [PubMed: 17443116]
8. DeWalt DA, Rothrock N, Yount S, Stone AA. Evaluation of item candidates: The PROMIS qualitative item review. *Medical Care*. 2007; 45(Suppl. 1):S12–S21. [PubMed: 17443114]
9. Castel LD, Williams KA, Bosworth HB, Eisen SV, Hahn EA, Irwin DE, et al. Content validity in the PROMIS social health domain: a qualitative analysis of focus group data. *Quality of Life Research*. 2008; 17:737–749. [PubMed: 18478368]
10. Christodoulou C, Jungaenel DU, DeWalt DA, Rothrock N, Stone AA. Cognitive interviewing in the evaluation of fatigue items: Results from the patient-reported outcomes measurement information system (PROMIS). *Quality of Life Research*. 2008; 17:1239–1246. [PubMed: 18850327]
11. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care*. 2007; 45(Suppl. 1):S22–S31. [PubMed: 17443115]
12. Amtmann A, Cook KF, Jensen MP, Chen W, Choi S, Revicki D, et al. Development of a PROMIS Item Bank to Measure Pain Interference. *Pain*. (under review). Revised and resubmitted to.
13. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D. Item Banks for Measuring Emotional Distress from the Patient-Reported Outcomes Measurement Information System (PROMIS): Depression, Anxiety, and Anger. *Psychological Assessment*. (under review). Submitted to.
14. Cella D, Riley W, Stone AA, Rothrock N, Reeve BB, Yount S, et al. Initial item banks and first wave testing of the Patient Reported Outcomes Measurement Information System (PROMIS) network: 2005-2008. *Journal of Clinical Epidemiology*. (in press).
15. Embretson, SE.; Reise, SP. *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates, Publishers; Mahway, NJ: 2000.
16. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Quality of Life Research*. 2009; 18:873–880. [PubMed: 19543809]
17. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Fourth Edition. American Psychiatric Association; Washington, DC: 1994.
18. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, the Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. What is sufficient evidence for the reliability and validity of patient-reported outcome measures. *Value in Health*. 2007; 10(Suppl. 2):S94–S105. [PubMed: 17995479]

Table 1

Content Expert Sociodemographic and Expertise Characteristics by Domains Reviewed

	Physical Function (A)	Fatigue (B)	Pain (C,D)	Emotional Distress (E, F, G)	Sleep (J,K)
N	6	6	4	9	5
Female (n)	2	1	1	3	1
Minority (n)	0	1	0	1	0
Ph.Ds. (n)	2	5	2	9	5
Experience Range (yrs)	5 to 50	11 to 25	18 to 30	20 to 45	25 to 40

Note: Five additional experts provided feedback on the social domain names and definitions; however, due to further social domain item development, their social domain feedback is not reported here but their feedback was included for the global items. Of these 5 social domain experts, 4 were female, 0 were minority, all were Ph.D.s, and years of experience ranged from 20-35.

Table 2

Ratings of Domain Name and Definition Fit by Item Banks

Domain (n rating)	Domain Name Fit with Content		Domain Definition Fit with Content		
	Mean (SD)*	Fraction rating "Very Much"	Mean (SD)*	Fraction rating "Very Much"	Fraction indicating content present not reflected in definition
Physical Function	4.8 (0.41)	5/6	4.7 (0.52)	4/6	2/6
Fatigue	4.5 (0.84)	4/6	4.2 (0.98)	3/6	1/6
Pain Behavior	4.0 (1.40)	2/4	3.8 (0.50)	0/4	2/4
Pain Impact	4.2 (0.50)	1/4	4.5 (0.58)	2/4	1/4
Depression	4.3 (0.87)	5/9	4.4 (0.73)	5/9	4/9
Anxiety	4.4 (0.73)	5/9	4.6 (0.73)	6/9	7/9
Anger	4.9 (0.33)	8/9	4.8 (0.44)	7/9	1/9
Sleep Disturbance	4.0 (0.71)	1/5	4.6 (0.55)	3/5	0/5
Wake Disturbance	3.8 (0.84)	1/5	4.8 (0.50)	3/5	2/4

* 1-5 rating (1 = not at all, 2 = a little bit, 3 = somewhat, 4 = quite a bit, 5 = very much)

Table 3

Initial and Revised PROMIS Domain Definitions

Domain Name	Initial Definition	Revised Definition
Global Health	<p>Global health refers to evaluations of health in general. The global health items include global ratings of the five primary PROMIS domains (physical function, fatigue, pain, emotional distress, social health) and general health perceptions that cut across domains. Global items allow respondents to weigh together different aspects of health to arrive at a “bottom-line” indicator of their health. These items have been found to be consistently predictive of important future events such as health care utilization and mortality. The PROMIS global health items include the most widely used single self-rated health item (global01). Previous research has shown that the former item taps physical health and mental health about equally but it reflects physical health more than mental health, especially for those with lower levels of income. PROMIS includes a single item that provides a pure rating of physical health (global03) and another item for mental health (global04). Also included is an overall quality of life item (global02). The remaining items provide global ratings of physical function (global06), fatigue (global08), pain (global07), emotional distress (global10), and social health (global05 and global09).</p>	<p>The PROMIS Global Health items assess health in general (i.e. overall health). The global health items include global ratings of the five primary PROMIS domains (physical function, fatigue, pain, emotional distress, social health) as well as perceptions of general health that cut across domains. Global items allow respondents to weigh together different aspects of health to arrive at a “bottom-line” indicator of their health. Similar global health items have been found predictive of future health care utilization and mortality. The PROMIS Global Health items include the most widely used single self-rated health item (“In general, would you say your health is ...”). Previous research has shown that this item taps physical and mental health about equally but reflects physical health more than mental health among respondents at lower income levels. PROMIS Global Health items include specific ratings of physical health and mental health, as well as a rating of overall quality of life. The remaining items provide global ratings of physical function, fatigue, pain, emotional distress, and social health. There is no reporting period specified for these items; current status is inferred. The PROMIS Global Health items can be administered as individual items or combined to produce separate physical and mental health summary scores (see Hays, Bjorner, Revicki, Spritzer, & Cella, 2009).</p>
Physical Function	<p>Physical Function is defined as one’s ability to carry out various activities that require physical capability, ranging from self-care (activities of daily living) to more vigorous activities that require increasing degrees of mobility, strength, or endurance^{21, 11, 12, 24}. Physical function items, when considered as an outcome endpoint for clinical research in chronic illness, generally have a “capability” stem and a corresponding “capability” set of response items (e.g., “Are you able to...normally, with some difficulty, with moderate difficulty, with great difficulty, unable to do”), and are given in the present tense. This specifically excludes some items that may have great utility in other settings, as with “performance” items that ask whether an activity was actually conducted during a specified time frame (with a “Did you?” type of stem). Such items require capability but also opportunity and motivation. The use of capability stems also excludes the concept of satisfaction (e.g., “How satisfied are</p>	<p>The PROMIS Physical Function item bank assesses one’s ability to carry out activities that require physical actions, ranging from self-care (activities of daily living) to more complex activities that require a combination of skills, often within a social context. “Physical Function” is inclusive of the term “disability” and includes the full spectrum of physical functioning from severe impairment to exceptional physical abilities. The PROMIS Physical Function items assess capability to perform a variety of physical activities, and often begin with the stem “Are you able to ...” Items assessing performance of these activities (the frequency with which physical activities were performed within a specified timeframe), may have great utility for some purposes, but are not included in the physical function item bank. Performance requires not only capability but also opportunity and motivation. The use of capability stems in the PROMIS Physical Function item bank also</p>

Domain Name	Initial Definition	Revised Definition
	you with your current level of function?”). Such questions address subjective appraisals of oneself that incorporate concepts such as coping or adjustment. Because many persons with a chronic disease will have more than one chronic condition and cannot distinguish the fraction of a problem attributable to each one, physical function items attempt to quantitate the sum of these effects, leaving the teasing out of relative contributions to the analysis stage. Physical function is conceptually multidimensional, with four related subdomains: mobility (lower extremity function), dexterity (upper extremity function), axial (neck and back function), and ability to carry out instrumental activities of daily living (IADL).	excludes satisfaction items (e.g., “How satisfied are you with your current level of functioning?”). Such questions address subjective appraisals of oneself that incorporate concepts such as coping or adjustment. Additionally, because many persons with a chronic disease will have more than one chronic condition and often are unable to distinguish the proportion of physical limitation attributable to each condition, the PROMIS physical function items assess physical capabilities and limitations without causal attribution. Physical function is conceptually multidimensional, with four related subdomains: mobility (lower extremity function), dexterity (upper extremity function), axial (neck and back function), and ability to carry out instrumental activities of daily living. There is no reporting period specified for these items; current status is inferred.
Pain Intro	Pain is an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage. Pain is what the patient says it is – that is, the “gold standard of pain assessment is self-report. Pain is divided conceptually into components of quality (referring to the nature, characteristics, intensity, frequency, and duration of pain), impact upon physical, mental, or social activities, and behaviors one engages in to avoid, minimize, or reduce pain.	Pain is an unpleasant sensory or emotional experience associated with actual or potential tissue damage, or described in terms of such damage. Pain is what the patient says it is – that is, the “gold standard” of pain assessment is self-report. Pain is divided conceptually into components of quality (e.g. the nature, characteristics, intensity, frequency, and duration of pain), behaviors (e.g. verbal and nonverbal actions that communicate pain to others) and interference (e.g. impact of pain on physical, mental, and social activities).
Pain Behavior	Pain behaviors are behaviors that usually communicate to others that a person is experiencing pain. The include observable displays such as sighing or crying, as well as verbal reports of pain or pain severity behaviors such as resting, guarding, facial expressions, asking for help, and taking medications.	The PROMIS Pain Behavior item bank assesses external manifestations of experiencing pain. These actions or reactions can be verbal or nonverbal, involuntary or deliberate. Pain behaviors usually communicate to others that a person is experiencing pain. They include observable displays such as sighing or crying, and pain severity behaviors such as resting, guarding, facial expressions, and asking for help, as well as verbal reports of pain. The item bank uses a “past 7 days” reporting period.
Pain Interference (previously Pain Impact)	Pain Impact refers to the consequences of pain on relevant aspects of persons’ lives and may include impact on social, cognitive, emotional, physical, and recreational activity as well as sleep and enjoyment of life.	The PROMIS Pain Interference item bank assesses the consequences of pain on relevant aspects of persons’ lives and may include the impact of pain on social, cognitive, emotional, physical, and recreational activities as well as sleep and enjoyment in life (Note that Pain Interference bank includes only one sleep item). The item bank uses a “past 7 days” reporting period.
Fatigue	Fatigue at its highest level is defined as an overwhelming, debilitating, and sustained sense of exhaustion	The PROMIS Fatigue item bank assesses fatigue from mild subjective feelings of tiredness to an

Domain Name	Initial Definition	Revised Definition
	<p>that decreases one's ability to carry out daily activities, including the ability to work effectively and to function at one's usual level in family or social roles. Similar subjective feelings, yet fewer behavioral impacts, are associated with lower levels of fatigue. Fatigue is divided conceptually into the experience of fatigue (such as its intensity, frequency, and duration), and the impact of fatigue upon physical, mental, and social activities.</p>	<p>overwhelming, debilitating, and sustained sense of exhaustion that is likely to decrease one's ability to carry out daily activities, including the ability to work effectively and to function at one's usual level in family or social roles. Fatigue is divided conceptually into the experience of fatigue (such as its frequency, duration, and intensity), and the impact of fatigue upon physical, mental and social activities. The item bank uses a "past 7 days" reporting period.</p>
Emotional Distress Intro	<p>Emotional distress commonly refers to unpleasant feelings or emotions that are experienced privately and, therefore, are good candidates for assessment as patient-reported outcomes. Emotional distress is comprised typically of aspects of anxiety, depression, and anger. Anxiety, depression, and anger represent risk factors that have been associated with both the incidence and progression of disease. The mechanisms by which these associations arise are not well understood, but they can be organized into two general families: direct effects via physiological pathways (e.g., the association between depression and risk factors for cardiovascular disease such as blood lipids and inflammation, which may be produced by shared causal variables) and indirect effects via the impact on health-related behaviors (e.g., increased use of tobacco and alcohol as a consequence of negative emotions). Given the overlap among the symptoms of anxiety, depression, and anger, a number of conceptual models have been proposed to account for the shared versus unique variance captured in measures of negative affect. Watson and Clark⁶,²⁶ proposed a hierarchical structure to explain the relationships between self-reported symptoms of anxiety, depression, and anger. First, they described a second-order, nonspecific factor reflecting high levels of negative affect—or "general distress"—common to all these emotions. Anger tends to have smaller loadings on the general factor than anxiety and depression, but it still is a strong marker of the dimension. In addition, Watson and Clark's model included first-order factors that are specific to, and help to differentiate, the three.</p>	<p>Emotional distress typically refers to unpleasant feelings or emotions. Emotional distress is often reflected in reports of anxiety, depression, and anger. Given the overlap among the experiences and symptoms of anxiety, depression, and anger, a number of conceptual models have been proposed to account for the shared versus unique variance captured in measures of negative affect or emotional distress. For example, Watson (2005) proposed a hierarchical structure with first order factors of anxiety, depression, and anger subsumed under a second-order, nonspecific factor reflecting high levels of negative affect or "general distress". Anger tends to have smaller loadings than anxiety and depression on the general distress factor, but it still is an important component of general distress.</p>
Anger	<p>Anger is distinguished by attitudes of hostility and cynicism and is often associated with experiences of frustration impeding goal-directed behavior. Specific components relate to verbal and nonverbal evidence of interpersonal antagonism. The PROMIS item</p>	<p>The PROMIS Anger item bank assesses angry mood (e.g., irritability, frustration), negative social cognitions (e.g., interpersonal sensitivity, envy, disagreeableness), verbal aggression, and efforts to control anger. Anger is distinguished by attitudes of hostility</p>

Domain Name	Initial Definition	Revised Definition
	<p>bank for anger focuses on angry mood (e.g. irritability, reactivity), negative social cognition (e.g. interpersonal sensitivity, envy, vengefulness), verbal aggression, and efforts necessary to control angry mood. In general, our PROMIS item banks emphasize the cognitive and affective components of these concepts. Both psychometric considerations (e.g. skewed distributions for high threshold behavioral items, the need to fit item response theory to coherent unidimensional concepts) and considerations regarding validity (e.g. potential confounding between somatic symptoms of emotional distress and markers of physical disease) have led us to this emphasis.</p>	<p>and cynicism and is often associated with experiences of frustration impeding goal-directed behavior. Specific components relate to verbal and nonverbal evidence of interpersonal antagonism. Physical aggression items were excluded from the PROMIS Anger item bank based on psychometric properties and poor fit of these items to the other items in the bank. The item bank uses a "past 7 days" reporting period.</p>
Anxiety	<p>Symptoms that best differentiate anxiety are those that reflect autonomic arousal and the experience of threat. The PROMIS item bank for anxiety focuses on fear (e.g. fearfulness, feelings of panic), anxious misery (e.g. worry, dread), hyperarousal (e.g. tension, nervousness, restlessness) and somatic symptoms related to arousal (cardiovascular symptoms, dizziness).</p>	<p>The PROMIS Anxiety item bank assesses fear (e.g., fearfulness, feelings of panic), anxious misery (e.g., worry, dread), hyperarousal (e.g., tension, nervousness, restlessness), and somatic symptoms related to arousal (e.g., racing or pounding heart, dizziness). Symptoms that best differentiate anxiety are those that reflect autonomic arousal and the experience of threat. Only one behavioral avoidance item (e.g. "I avoided public places and activities") is included in the PROMIS Anxiety item bank. Other behavioral avoidance items were excluded based on psychometric properties and poor fit with the item bank. Therefore, this item bank does not comprehensively tap behavioral fear avoidance. The item bank uses a "past 7 days" reporting period.</p>
Depression	<p>Symptoms specific to depression are those that reflect low levels of positive affect. In addition, depression is often characterized by the experience of loss and feelings of hopelessness, helplessness, and worthlessness. The PROMIS item bank for depression focuses on negative mood (e.g. sadness, guilt), decrease in positive affect (e.g. loss of interest), information-processing deficits (e.g. problems in decision-making, negative views of self (e.g. self-criticism, worthlessness), and negative social cognition (e.g. loneliness, interpersonal alienation).</p>	<p>The PROMIS Depression item bank assesses negative mood (e.g., sadness, guilt), negative views of the self (e.g., self-criticism, worthlessness), negative social cognition (e.g., loneliness, interpersonal alienation), and decreased positive affect and engagement (e.g., loss of interest, loss of meaning and purpose). Depression is reflected in high levels of negative affect and low levels of positive affect. It is often characterized by the experience of loss and feelings of hopelessness, helplessness, and worthlessness. Somatic symptoms items (e.g. changes in appetite, sleep, psychomotor functioning) were excluded from the PROMIS Depression item bank based on psychometric properties and poor fit of these items to the other items in the bank. Therefore, the PROMIS Depression item bank does not reflect the full range of symptoms commonly considered in a diagnosis of Major Depressive Disorder, but the exclusion of somatic items from</p>

Domain Name	Initial Definition	Revised Definition
		this bank eliminates the confounding effects of these items when assessing depression in patients with comorbid physical conditions. The item bank uses a “past 7 days” reporting period.
Sleep Intro	<p>Sleep and wakefulness are the two fundamental behavioral states of human beings. Sleep is a rapidly reversible, recurrent state of reduced (but not absent) awareness of and interaction with the environment. Wakefulness is a behavioral state of active engagement and interaction with the environment, including the perception and processing of stimuli and the production of cognitive, emotional, and behavioral responses. Sleep and wakefulness are both distinct from abnormal behavioral states such as delirium or coma. The generation of sleep and wakefulness is an endogenous phenomenon which is regulated by homeostatic and circadian physiological processes, but which can be influenced by internal (e.g., cognitive, emotional) and external (e.g., physical, environmental) stimuli. A considerable body of scientific data describes the neuroanatomy and neurophysiology of sleep and wakefulness. While the precise functions of sleep remain to be identified, there is little doubt that sleep is necessary for optimal mental and physical function during wakefulness. Alterations in the amount or quality of sleep have been associated with impaired alertness, cognitive and emotional function, and learning; disordered function of the central nervous system, cardiovascular, endocrine-metabolic, and immune systems; and even with increased mortality. As fundamental behavioral and brain states, sleep and wakefulness can be described at several levels of organization, including the activity of individual cells, neural systems, or the entire organism. Methods for measuring sleep at the organismic level in humans include physiological recording, functional neuroanatomic studies, and patient-reported outcomes (PROs). The PROMIS Sleep Disturbances and Wake Disturbances Scales are examples of the latter. Multiple types of assessments are possible within the broad domain of sleep-wake PROs. For instance, some self-report assessments are used to diagnose specific sleep disorders; others are used to assess habitual sleep-wake quantities and patterns; and still others measure an individual’s perceptions of the quality and global experience of sleep and wakefulness. The PROMIS Sleep Disturbances and Wake Disturbances Scales fall into the latter category. Both scales</p>	<p>Sleep and wakefulness are the two fundamental neurobehavioral states of human beings. Sleep is a rapidly reversible, recurrent state of reduced (but not absent) awareness of and interaction with the environment. Wakefulness is a behavioral state of active engagement and interaction with the environment, including the perception and processing of stimuli and the production of cognitive, emotional, and behavioral responses. As fundamental neurobehavioral states, sleep and wakefulness can be described on several levels, ranging from single neuronal activity to patient-reported outcomes (PROs) of sleep experience and quality. Multiple types of assessments are possible within the broad domain of sleep-wake PROs. Some self-report assessments are used to diagnose specific sleep disorders; others are used to assess habitual sleep-wake quantities and patterns; and still others measure an individual’s perceptions of the quality and global experience of sleep and wakefulness. The PROMIS Sleep Disturbance and Sleep-Related Impairment item banks fall into the latter category.</p>

Domain Name	Initial Definition	Revised Definition
	assess function and disturbances over a seven-day time frame.	
Sleep Disturbance	The PROMIS Sleep Disturbance Scale focuses on perceptions of sleep quality, sleep depth, and restoration associated with sleep; perceived difficulties with getting to sleep or staying asleep; and perceptions of the adequacy of and satisfaction with sleep. The Sleep Disturbance Scale does not include symptoms of specific sleep disorders, nor does it provide subjective estimates of sleep quantities (e.g. the total amount of sleep, time to fall asleep, or amount of wakefulness during sleep).	The PROMIS Sleep Disturbance item bank assesses perceptions of sleep quality, sleep depth, and restoration associated with sleep; perceived difficulties and concerns with getting to sleep or staying asleep; and perceptions of the adequacy of and satisfaction with sleep. The PROMIS Sleep Disturbance Scale does not include symptoms of specific sleep disorders, nor does it provide subjective estimates of sleep quantities (e.g., the total amount of sleep, time to fall asleep, or amount of wakefulness during sleep). The item bank uses a “past 7 days” reporting period.
Sleep-Related Impairment (previously Wake Disturbance)	The PROMIS Wake Disturbance Scale focuses on perceptions of alertness, sleepiness, and tiredness during usual waking hours; and on functional impairments during wakefulness that are associated with sleep problems or impaired alertness. The Wake Disturbance Scale does not directly assess cognitive, affective, or performance impairments. The Wake Disturbance Scale measures the level of waking alertness, sleepiness, and function within the context of overall sleep function.	The PROMIS Sleep-Related Impairment item bank assesses perceptions of alertness, sleepiness, and tiredness during usual waking hours, and the perceived functional impairments during wakefulness associated with sleep problems or impaired alertness. The Sleep-Related Impairment item bank measures the level of waking alertness, sleepiness, and function within the context of overall sleep-wake function, but does not directly assess cognitive, affective, or performance impairments. The item bank uses a “past 7 days” reporting period.