



Published in final edited form as:

J Ment Health Policy Econ. 2011 December ; 14(4): 201–208.

Application of the National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS®) to Mental Health Research

William T. Riley,
National Heart, Lung, and Blood Institute

Paul Pilkonis, and
University of Pittsburgh

David Cella
Northwestern University

Abstract

Background—The Patient-Reported Outcomes Measurement Information System (PROMIS) is a National Institutes of Health initiative to develop item banks measuring patient-reported outcomes (PROs) and to create and make available a computerized adaptive testing system (CAT) that allows for efficient and precise assessment of PROs in clinical research and practice.

Aims of the Study—Based on the presentation from a symposium on “Evidence-based Outcomes in Psychiatry: Updates on Measurement Using Patient-Reported Outcomes (PRO)” at the 2011 American Psychiatry Association Convention, this paper provides an overview of PROMIS and its application to mental health research.

Methods—The PROMIS methodology for item bank development and testing is described, with a focus on the implications of this work for mental health research.

Results—Utilizing qualitative item review and state-of-the-art applications of item response theory (IRT), PROMIS investigators have developed, tested, and released item banks measuring physical, mental, and social health components. Ongoing efforts continue to add new item banks and further validate existing banks.

Discussion—PROMIS provides item banks measuring several domains of interest to mental health researchers including emotional distress, social function, and sleep. PROMIS methodology also provides a rigorous standard for the development of new mental health measures.

Implications for Health Care Provision—Web-based CAT or administration of short forms derived from PROMIS item banks provide efficient and precise dimensional estimates of clinical outcomes that can be utilized to monitor patient progress and assess quality improvement.

Implications for Future Research—Use of the dimensional PROMIS metrics (and co-calibration of the PROMIS item banks with existing PROs) will allow comparisons of mental health and related health outcomes across disorders and studies.

Self-report or patient-report measures have a long history in the mental health field. Questionnaire methods have been used in psychology for over a century (1). Self-report is particularly appropriate as a measurement methodology when the information of interest is known only to the patient or when other data sources exist but the logistics or costs to obtain these data are prohibitive (2). There are numerous mental health constructs that are known only to the patient (e.g. emotions, thoughts) or that are often too difficult to obtain via other sources (e.g. social participation); however, patient-reported outcomes (PROs) appear to be

better integrated into clinical research for medical conditions such as arthritis (4) and cancer (5) than for mental health conditions.

Due in part to concerns about the reliability and validity of self-report from those with cognitive deficits or emotional biases, clinical research in mental disorders has relied substantially on clinical rating scales (e.g. 5). Clinical ratings are based predominately on patient-report during the clinical interview; however, the association between self-report and clinical ratings is only moderate (e.g. 6), suggesting that data other than patient report influence clinical ratings. The additional data available to the clinician (e.g. patient cognitive status, direct observation) are generally assumed to reduce measurement error relative to self-report, but it is also plausible that clinical raters add error to the estimate, and measurement problems with clinical ratings are well documented. Zimmerman and colleagues, for example, summarized numerous problems with the Hamilton Depression Rating Scale (HDRS) including version control, overrepresentation of vegetative symptoms, lack of a consistent metric, and problems with anchor point descriptions (7). Recent improvements to the HDRS have attempted to address some of these problems (8), but it is clear that both clinical ratings and patient-reported outcomes are less than perfect estimates of the true score, and may be better viewed as complementary rather than competitive outcome measures.

Advances in biomarker research hold promise for developing objective indicators of mental disorders, or at least some of the endophenotypic substrates that contribute to these disorders. These efforts have provided important leads for better understanding the pathophysiology and for developing more targeted biological treatments for disorders such as schizophrenia (9) and depression (10,11). Despite the value of biomarkers for elucidating the neurobiological pathways that contribute to mental disorders, their promise as surrogate endpoints for clinical outcomes is far from realized. In the development of a methodological and statistical framework for evaluating surrogate endpoints in mental health research, Molenberghs and colleagues (2010) emphasized that acceptance of a surrogate biomarker requires not only correlation with the clinical endpoint, but also that treatment of the surrogate endpoint predicts the effect on the clinical endpoint. One implication of this framework is that improving the reliability and validity of current clinical endpoint measures, including clinical rating and patient report measures, will facilitate the eventual identification and validation of surrogate endpoint biomarkers. Therefore, even among those who envision a future in which all mental health outcome measures are biologically-based indices, improvement in our current clinical endpoint measures, including PROs, is clearly needed.

The Patient-Reported Outcomes Measurement Information System (PROMIS)

Since 2004, the National Institutes of Health has funded the Patient-Reported Outcomes Measurement Information System (PROMIS). The goal of this effort is to apply state-of-the-art qualitative item development methodology and modern psychometric theory to health outcomes measurement development. This trans-NIH Roadmap initiative was developed in response to the diverse array of outcome measures used in clinical research, and the difficulty comparing and integrating data across studies using different measures of the same construct. The objectives of the PROMIS initiative are to:

1. Develop and test large banks of items measuring PROs;
2. Create a computerized adaptive testing system for efficient, psychometrically robust assessment of PROs in clinical research on a wide range of clinical diseases;

3. Create a publicly available system that can be added to and modified periodically, and that will allow researchers to access a common repository of PROs and computerized adaptive testing (13).

The first project period (2004-2009) or “PROMIS I” consisted of a cooperative agreement with six project sites and a statistical coordinating center. The PROMIS I cooperative network developed, evaluated, and released item banks for adults measuring Physical Function, Pain Interference, Pain Behavior, Fatigue, Sleep Disturbance, Sleep-related Impairment, Depression, Anxiety, Anger, Social Role Function, and Satisfaction with Social Role Function. In addition to these cooperative network projects, individual projects included development and evaluation of pediatric item banks that paralleled many of the adult item banks, evaluation of the relationship of various reporting periods, effects of different modes of administration, and a range of other studies. Perhaps most significant for the dissemination and implementation of these item banks, the Statistical Coordinating Center, with input from the network, developed an Internet-based item bank administration and scoring system, Assessment Center, that allows clinical researchers and other end-users to select and automate administration and scoring, including the capability of computer adaptive testing in which the next item administered is based on the responses to the prior items (14).

The second project period (2009-2013) or “PROMIS II” significantly expanded the PROMIS network to 12 sites and 3 centers (network, statistical, technology) to further validate the item banks developed in PROMIS I and develop new item banks for additional constructs. The PROMIS II network has continued the work begun in PROMIS I on evaluating the concurrent validity, construct validity, and sensitivity to change in various patient groups. The PROMIS II network also has begun development and testing of a variety of additional domains including Gastrointestinal Distress, Chronic Disease Management Self-efficacy, and a number of pediatric domains including Experience of Stress, Positive Affect, and Life Satisfaction. Details of the PROMIS initiative can be found at www.nihpromis.org.

The PROMIS Mental Health item banks released for use currently include three negative emotion item banks - Depression, Anxiety, and Anger - for adults and children, Positive and Negative Psychosocial Illness Impact Item Banks, and Applied Cognition (General Concerns and Abilities) item banks (15). Alcohol item banks including Alcohol Use, Positive and Negative Consequences, and Positive and Negative Expectancies are completed and soon to be released. Although technically considered part of the Physical Health component, the sleep banks (Sleep Disturbance and Sleep-related Impairment) and the sexual function banks have considerable overlap with mental health, as do a number of banks in the Social Health component. A list of currently released item banks and scales are provided in Table 1, and many of these item banks have relevance to mental health research.

The PROMIS Item Bank Development and Testing Methodology

In addition to developing assessment tools for clinical research, the PROMIS network also has had a role in refining standards for the development and evaluation of patient-reported outcome measures. The network has taken a thoughtful and empirically-based perspective to develop and refine the steps in the PROMIS item bank development and evaluation process, cognizant that these steps may be cited as a standard for future development of IRT-derived item banks measuring health and illness constructs. Details of this methodology have been published elsewhere, (e.g. 16-19), but the following description and Figure 1 provide an overview of the PROMIS item bank development and testing process.

The PROMIS methodology for item bank development begins with the development of a domain framework based on literature review, expert input, and the analysis of existing datasets. The PROMIS domain framework has evolved as new item banks are added and data on relationships of the various domains are further explored. The domain framework uses the World Health Organization model of physical, mental, and social well-being as the primary components of health (20). Within each of these components are subcomponents. Within mental health, for example, there are affective (e.g. emotional distress), cognitive (e.g. concentration, memory), and behavioral (e.g. substance abuse) subcomponents. Within these subcomponents reside the unidimensional domains or subdomains that are represented by their respective item banks. The framework is a nested structure in which domains and subdomains can be generally subsumed under their higher order components and subcomponents, but the framework does not assume exclusivity within any higher order construct. For example, both sexual function and sleep function reside within the physical health component, but there are mental health, and in the case of sexual function, also social health relationships to these banks (21).

After generating the conceptual framework for the construct of interest, an item pool is developed based on literature review of existing measures (e.g. 22) and generation of additional item content from patient focus groups (e.g. 23). The resulting pool of items is organized into the various facets of the construct to derive representative item content for each identified facet of the construct. The resulting item content is revised and formatted for consistency of item stems and response options. These items are reviewed and further revised to improve readability and to facilitate future translatability to other languages. The items are subjected to cognitive interviews with participants representing likely respondents to assess the item comprehension and to revise or eliminate items that produce respondent confusion or responses inconsistent with the intent of the item (17). Intellectual property review is performed to identify any derived items that closely resemble a unique copyrighted item. In these cases, permission for use is sought and items are excluded if permission is not obtained.

To not overly burden test respondents, the resulting item pool is further reduced to a testable subset of items that best represent all of the facets of the construct. In the case of the initial PROMIS testing, a combination of full-bank and block (subsets of items from multiple banks) testing was performed to evaluate the numerous item pools within and across domains (16). The entire testing sample for the initial PROMIS effort included 21,133 participants to insure a minimum of 500 participants responding to each tested item. The general population sample was obtained from an Internet polling panel (YouGov Polimetrix) and augmented by clinical samples from the various PROMIS sites to ensure adequate coverage of the more severe range of the construct being assessed. In addition to IRT calibration using the full sample, a scale-setting subsample representative of the U.S. population was used for norming purposes and to set the T-score metric (mean = 50, SD = 10) (24).

Following testing, the items were analyzed using a range of classical and modern item analysis to examine monotonicity, scalability, and dimensionality (18). IRT calibrations were determined using the graded response model and IRT model fit was assessed. Differential item function (DIF) for gender, age, and education also was performed and items with DIF were excluded (18). These analyses were then reviewed by content and psychometric teams to select the items to be included in the banks.

Since in most cases the resulting item banks consisted of only approximately half of the items tested, the resulting item banks were further reviewed by experts to assess if the resulting item banks continued to adequately represent the domain names and definitions,

and further revisions of the domain names and definitions were performed based on this feedback (25). Based on the psychometric properties of the individual items and CAT simulations, static short forms were developed as an alternative to CAT administration.

In addition to the content validity methods outlined above, existing legacy measures included in the initial calibration testing were used to assess concurrent validity. Ongoing validity studies are testing construct validity and sensitivity to change of the PROMIS item banks.

Implications of the PROMIS Methodology and Results

Comparability of Measures via IRT Methodology

The application of Item Response Theory (IRT) is central to the PROMIS measure development effort. IRT is a latent variable modeling approach to measurement development and analysis, and is generally consistent with classical test theory and development, but offers a number of benefits over classical test development including: a) the ability to scale both people and items on the same metric, b) methods to better characterize and minimize measurement error, c) integer scaling, and d) the flexibility of item administration that allows for computer adaptive testing (CAT) in which subsequent item administration is based on the information obtained from the previously administered items (26,27). PROMIS is clearly not the first to apply Item Response Theory (IRT) to mental health constructs. IRT has a long history among mental health measures, and a number of classically-developed instruments have been evaluated using IRT methodology (28,29) to elucidate the psychometric properties of the items and to facilitate more efficient and precise administration via CAT (30,31). PROMIS is unique, however, in the pooling of item content across all known measures of a health-related construct and the utilization of IRT methodologies to create a single item bank that can be cross-calibrated with existing legacy measures of the construct.

The linking or cross-calibrating of existing and future instruments to the PROMIS item banks is critical to understanding why the development of yet another measure solves the problem of having too many different measures. A common strategy to address comparability and data integration or harmonization across studies has been to identify a common or consensus measure of a given construct that can then be used in most research. There are numerous recent efforts that have utilized this strategy in research (e.g. 32) and in clinical practice (e.g. 33). Assuming that everyone agreed to use the consensus measure in all future research, the identification of a consensus measure does address measurement comparability across future studies, but does not address the lack of comparability across prior studies or between prior and future studies. More importantly, as measurement science advances and better measures are developed, it is unclear if the new measure should become the consensus measure, sacrificing comparability with prior research, or if the current consensus measure should be retained, sacrificing improved reliability and validity.

One method to address these issues is to have a common scale or metric that all instruments or measures use to express their results. Blood pressure measurement is illustrative of this type of approach. Systolic and diastolic blood pressures are measured in mmHg. This scale, or metric, is the result of traditional sphygmomanometers that used the height of a column of mercury in mm to reflect the circulating pressure (34). Mercury is not used in current aneroid and electronic blood pressure devices, but these newer instruments still calibrate their raw output to the standard mmHg scale. By separating the scale from the method or instrument used to estimate a person's value on that scale, researchers have been able to compare blood pressure measurement across devices, between studies, and over time. The various instruments and methods for measuring blood pressure differ in their precision, but

all produce an estimate on the mmHg scale. This approach also allows researchers to select a more efficient, albeit less precise method to estimate blood pressure should their research question not require the same level of precision as a hypertension trial. Practitioners also benefit from a single mmHg scale for expressing blood pressure that can be easily applied to treatment decision algorithms regardless of the method and instrumentation used to determine mmHg. By allowing the method and instrumentation to vary and evolve while holding constant the scale on which all blood pressure measurement is expressed, blood pressure measurement can be harmonized and integrated across prior, current, and future studies that measure blood pressure.

The PROMIS initiative takes a similar approach of setting a scale and then cross-calibrating prior legacy measures and all future measures to that scale. Instead of selecting a chosen measure from among the existing measures of a given construct, the IRT-based PROMIS item bank development process has created a common metric or scale on which the PROMIS item bank or any existing measure of the same construct can be cross-calibrated. PROMIS investigators are currently working to cross-calibrate a variety of existing PRO instruments with the PROMIS item bank scale so that these existing instruments can be compared on the same scale. More importantly, as PRO measurement science advances, more precise and efficient measures can be developed and expressed on the same scale, maintaining comparability across studies and providing a standard scale for clinical decision making that can be retained as new measurement instruments or procedures are introduced.

Unidimensional Measures and Multi-dimensional Constructs

One challenge in the development of item banks measuring mental health constructs such as depression and anxiety is that a core IRT assumption is unidimensionality of the underlying construct. Depression and anxiety, however, are highly related domains (35). Bifactor models have been developed to address this issue (36); however, when bifactor and unidimensional factor models were used on the PROMIS depression and anxiety item pools, comparable fits were found, resulting in selection of the more parsimonious model of separate but highly related banks for depression and anxiety (15).

The other unidimensional challenge of the emotional distress item banks is that, while these banks are unidimensional and were developed to assess outcomes, not to diagnose, the diagnostic classifications associated with these banks are multidimensional in nature. Anxiety and depression have been conceptualized as parts of a tripartite model (35) with various facets including affective, cognitive, somatic, and behavioral dimensions. The PROMIS investigators attempted to balance the IRT demands of unidimensionality with the potential multidimensional content assumed to be integral to these constructs, and this effort for the emotional distress item banks is well described elsewhere (15). The resulting PROMIS Depression and Anxiety item banks fit the unidimensional model well, but may underrepresent some facets or attributes traditionally considered part of the diagnostic construct. The Depression item bank, for example, is comprised primarily of items assessing affective (I feel sad) and cognitive (I feel worthless) facets of depression. Somatic items such as psychomotor functioning, appetite, and sleep did not have sufficient fit to be included in the Depression item bank (15), a finding consistent with other IRT analyses of depression items (e.g. 37).

A practical benefit of this psychometric result is that the PROMIS Depression item bank can be used to measure depression in medical populations with minimal confounding from somatic items that may be the result of a medical condition other than depression. Ongoing studies are evaluating the relationship of the PROMIS Depression bank to existing legacy self-report and interview measures (e.g. HAM-D, PHQ-9) that include these somatic items.

These data should help determine if the PROMIS Depression bank performs comparably to existing measures that include somatic items.

The failure of these somatic items to fit within the PROMIS Depression item bank also has important theoretical and conceptual implications. For example, sleep disturbance, while highly associated with the affective and cognitive facets of depression, may not be part of the core depressive syndrome. Substantial evidence indicates that sleep disturbance frequently predates the onset of depression, often by years (38) and has been shown to persist after the other depressive symptoms have remitted (39). Therefore, sleep disturbance may be highly related to depression but may not be a core symptom. Logically, one would not expect sleep, appetite, and psychomotor functioning to be discriminating symptoms of depression. In addition to being symptoms for a variety of medical conditions other than depression, these are also “Goldilocks symptoms” in the sense that a patient can meet the diagnosis of Major Depression by having “too much”, “too little”, or “just right” levels of these somatic symptoms. Somatic symptoms are still clearly of clinical importance, and PROMIS has developed a separate item bank measuring sleep disturbance, but the failure of these items to fit within the PROMIS depression item bank suggests that affective and cognitive symptoms may be more core to a major depressive syndrome than somatic symptoms.

Another potential benefit of constructing unidimensional mental health constructs is that these unidimensional constructs may clarify phenotypes that can be associated with genotypes and other pathophysiological studies of mental health mechanisms. The National Institute of Mental Health (NIMH) has an ongoing initiative, the Research Domain Criteria Project (RDoC) to develop a research classification of mental disorders based on dimensions of observable behavior and neurobiological measures (40). It is intended to serve as a framework to guide classification of patients that can be directly linked to genomics, neuroscience, and behavioral science to facilitate etiological research. The major domains being developed by RDoC are Negative Valence Systems, Positive Valence Systems, Cognitive Systems, Systems for Social Processes, and Arousal/Regulatory Systems. (<http://www.nimh.nih.gov/research-funding/rdoc/nimh-research-domain-criteria-rdoc.shtml>). Although RDoC has largely excluded patient report from current considerations, the PROMIS effort to define single dimensional item banks, and the development of a unidimensional depression item bank, is consistent with the RDoC effort. The RDoC effort is also intended to be complementary, not competitive, with the clinical diagnosis effort of Diagnostic and Statistical Manual – V (DSM-V).

PROMIS and the DSM-V

The DSM-V field testing includes the testing of cross-cutting dimensional measures which include selected PROMIS item banks. The aim of this effort is to augment the categorical diagnostic criteria with dimensional measures to provide additional information that assists the clinician in assessment, treatment planning, and treatment monitoring (41; see Narrow et al. in this issue for a detailed description of this effort).

Summary and Implications

The PROMIS initiative provides mental health researchers with measurement tools that can be utilized in a range of research and clinical settings. For highly precise measurement of primary outcomes in clinical trials, the static short forms or CAT can be used to provide precise estimates of depression, anxiety, anger, sleep disturbance, and other related health constructs. For survey research, small subsets of items from a bank can be used to provide a less precise estimate, but one that can be compared to longer and more precise forms of the bank. The reduced respondent burden, particularly of CAT administration in which highly

precise estimates can be obtained with relatively few items, and the ability to administer these items remotely via the web, also makes the PROMIS item banks useful for monitoring outcomes in clinical practice. Patients can complete the measures online at specified time points during treatment and the results can be reviewed immediately by their mental health professional who can adapt treatment as needed. Results from the DSM-V field trials utilizing PROMIS short forms to assess cross-cutting dimensional constructs should provide additional data on the usefulness of these measures in clinical practice and their relationship to categorical diagnoses.

The inclusion of PROMIS measures in clinical care not only benefits clinical care, but also has significant benefit for health services and health economics research. Mental health services and economics research has been hindered by the lack of dimensional outcome measures in clinical records (42). As a result, outcomes are often based on clinical notes indicating the presence or absence of a particular diagnosis. Adoption of PROMIS item banks as outcome indices in clinical practice would provide precise dimensional measures of change over time documented in health records. Even if other outcome measures are used in clinical practice, co-calibration with the PROMIS item bank will allow these data to be compared across practice settings. PROMIS is a measurement tool that has considerable potential to improve outcome measurement in mental health research and clinical settings.

Acknowledgments

The Patient-Reported Outcomes Measurement Information System (PROMIS) is an NIH Roadmap initiative to develop a computerized system measuring PROs in respondents with a wide range of chronic diseases and demographic characteristics. PROMIS II was funded by cooperative agreements with a Statistical Center (Northwestern University, PI: David Cella, PhD, 1U54AR057951), a Technology Center (Northwestern University, PI: Richard C. Gershon, PhD, 1U54AR057943), a Network Center (American Institutes for Research, PI: Susan (San) D. Keller, PhD, 1U54AR057926) and thirteen Primary Research Sites which may include more than one institution (State University of New York, Stony Brook, PIs: Joan E. Broderick, PhD and Arthur A. Stone, PhD, 1U01AR057948; University of Washington, Seattle, PIs: Heidi M. Crane, MD, MPH, Paul K. Crane, MD, MPH, and Donald L. Patrick, PhD, 1U01AR057954; University of Washington, Seattle, PIs: Dagmar Amtmann, PhD and Karon Cook, PhD, 1U01AR052171; University of North Carolina, Chapel Hill, PI: Darren A. DeWalt, MD, MPH, 2U01AR052181; Children's Hospital of Philadelphia, PI: Christopher B. Forrest, MD, PhD, 1U01AR057956; Stanford University, PI: James F. Fries, MD, 2U01AR052158; Boston University, PIs: Stephen M. Haley, PhD and David Scott Tulskey, PhD (University of Michigan, Ann Arbor), 1U01AR057929; University of California, Los Angeles, PIs: Dinesh Khanna, MD and Brennan Spiegel, MD, MSHS, 1U01AR057936; University of Pittsburgh, PI: Paul A. Pilkonis, PhD, 2U01AR052155; Georgetown University, PIs: Carol M. Moinpour, PhD (Fred Hutchinson Cancer Research Center, Seattle) and Arnold L. Potosky, PhD, U01AR057971; Children's Hospital Medical Center, Cincinnati, PI: Esi M. Morgan DeWitt, MD, MSCE, 1U01AR057940; University of Maryland, Baltimore, PI: Lisa M. Shulman, MD, 1U01AR057967; and Duke University, PI: Kevin P. Weinfurt, PhD, 2U01AR052186). NIH Science Officers on this project have included Deborah Ader, PhD, Vanessa Ameen, MD, Susan Czajkowski, PhD, Basil Eldadah, MD, PhD, Lawrence Fine, MD, DrPH, Lawrence Fox, MD, PhD, Lynne Haverkos, MD, MPH, Thomas Hilton, PhD, Laura Lee Johnson, PhD, Michael Kozak, PhD, Peter Lyster, PhD, Donald Mattison, MD, Claudia Moy, PhD, Louis Quatrano, PhD, Bryce Reeve, PhD, William Riley, PhD, Ashley Wilder Smith, PhD, MPH, Susana Serrate-Sztejn, MD, Ellen Werner, PhD and James Witter, MD, PhD. This manuscript was reviewed by PROMIS reviewers before submission for external peer review. See the Web site at www.nihpromis.org for additional information on the PROMIS initiative.

References

1. Gault RH. A history of the questionnaire method of research in psychology. *Ped Sem.* 1907; 14:366–383.
2. Baldwin, W. Information know one else knows: the value of self-report. In: Stone, AA.; Tukkan, JS.; Bachrach, CA.; Jobe, JB.; Kurtzman, HS.; Cain, VS., editors. *The Science of Self-Report.* LEA; London: 2000. p. 3-7.
3. Aletaha D, Landewe R, Karonitsch T, Bathon J, Boers M, Bombardier C, Bombardieri S, Choi H, Combe B, Dougados M, Emery P, Gomez-Reino J, Keystone E, Koch G, Kvein TK, Martin-Mola E, Matucci-Cerinic M, Michaud K, O'Dell J, Paulus H, Pincus T, Richards P, Simon L, Siegel J,

- Smolen JS, Sokka T, Strand V, Tugwell P, van der, Heijde D, van Riel P, Vlad S, van Vollenhoven R, Ward M, Weinblatt M, Wells G, White B, Wolfe F, Zhang B, Zink A, Felson D. Reporting disease activity in clinical trials of patients with rheumatoid arthritis: EULAR/ACR collaborative recommendations. *Ann Rheum Dis.* 2008; 67:1360–1364. [PubMed: 18791055]
4. Lipscomb J, Reeve BB, Clauser SB, Abrams JS, Watkins Bruner D, Burke LB, Denicoff AM, Ganz PA, Gondek K, Minsian LM, O'Mara AM, Revicki DA, Rock EP, Rowland J, Sgambati M, Trimble EL. Patient-reported outcomes assessment in cancer trials: taking stock, moving forward. *J Clin Oncology.* 2007; 25:5133–5140.
 5. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* 1960; 23:56–62. [PubMed: 14399272]
 6. Paykel ES, Prusoff BA, Klerman GL, DiMascio A. Self-report and clinical interview ratings of depression. *J Nerv Mental Dis.* 1973; 156:166–182.
 7. Zimmerman M, Posternak MA, Chelminski I. Is it time to replace the Hamilton Depression Rating Scale as a primary outcome in treatment studies of depression? *J Clin Psychopharm.* 2005; 25:105–110.
 8. Williams JB, Kobak KA, Bech P, Engelhardt N, Evans K, Lipsitz J, Olin J, Pearson J, Kalali A. The GRID-HAMD: Standardization of the Hamilton Depression Rating Scale. *Int Clin Psychopharmacol.* 2008; 23:120–129. [PubMed: 18408526]
 9. Carter CS, Barch DM, Bulmore E, Breiling J, Buchanan RW, Butler P, Cohen JD, Geyer M, Gollub R, Green MF, Jaeger J, Krystal JH, Moore H, Nuechterlein K, Robbins T, Silverstein S, Smith EE, Strauss M, Wykes T. Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia II: Developing imaging biomarkers to enhance treatment development for schizophrenia and related disorders. *Biol Psychiatry.* 2011; 70:7–12. [PubMed: 21529781]
 10. Bartova L, Berger A, Pezawas L. Is there a personalized medicine for mood disorders? *Eur Arch Psychiatry Clin Neurosci.* 2010; 260(Suppl 2):S121–126. [PubMed: 20957381]
 11. Luechter AF, Cook IA, Hamilton SP, Narr KL, Toga A, Hunter AM, Faull K, Whitelegge J, Andrews AM, Loo J, Way B, Nelson SF, Horvath S, Lebowitz BD. Biomarkers to predict antidepressant response. *Curr Psychiatry Rep.* 2010; 12:553–562. [PubMed: 20963521]
 12. Molenberghs G, Burzykowski T, Alonso A, Assam P, Tilahun A, Buyse M. A unified framework for the evaluation of surrogate endpoints in mental-health clinical trials. *Stat Methods Med Res.* 2010; 19:205–236. [PubMed: 19608602]
 13. Ader DN. Developing the Patient-Reported Outcomes Measurement System. *Med Care.* 2007; 45:S1–S3. [PubMed: 18027399]
 14. Gershon RC, Rothrock NE, Hanrahan RT, Jansky LJ, Harniss M, Riley W. The development of a clinical outcomes survey research application: Assessment Center. *Qual Life Res.* 2010; 19:677–685. [PubMed: 20306332]
 15. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D, PROMIS Cooperative Group. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS): depression, anxiety, anger. *Assessment.* 2011; 18:263–283. [PubMed: 21697139]
 16. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M, PROMIS Cooperative Group. The Patient-Reported Outcome Measurement Information System (PROMIS): progress of an NIH Roadmap Cooperative Agreement during its first two years. *Med Care.* 2007; 45:S3–S11. [PubMed: 17443116]
 17. DeWalt DA, Rothrock N, Yount S, Stone AA, PROMIS Cooperative Group. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care.* 2007; 45:S12–S21. [PubMed: 17443114]
 18. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DJ, Hambleton RK, Liu H, Gershon R, Reise SP, Lai JS, Cella D, PROMIS Cooperative Group. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care.* 2007; 45:S22–31. [PubMed: 17443115]
 19. 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 JF, Gershon R, Hahn E, Lai JS, Pilkonis P, Revicki D,

- Rose M, Weinfurt K, Hays R, PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks. *J Clin Epidemiol.* 2010; 63:1179–1194. [PubMed: 20685078]
20. World Health Organization (WHO). [Accessed Nov. 4, 2011] <http://apps.who.int/gb/bd/PDF/bd47/EN/constitution-en.pdf>
 21. PROMIS. [Accessed Nov. 4, 2011] http://www.nihpromis.org/Documents/PROMIS_Full_Framework.pdf
 22. Klem ML, Saghabi E, Abromitis R, Stover A, Dew MA, Pilkonis PA. Building PROMIS item banks: librarians as co-investigators. *Qual Life Res.* 2009; 18:881–888. [PubMed: 19548118]
 23. Kelly MAR, Morse JQ, Stover A, Hofkens T, Huisman E, Shulman S, Eisen SV, Becker SJ, Weinfurt K, Boland E, Pilkonis PA. Describing depression: congruence between patient experiences and clinical assessments. *Br J Clin Psychol.* 2011; 50:46–66. [PubMed: 21332520]
 24. Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley W, Hays RD. Representativeness of the Patient-Reported Outcomes Measurement Information System Internet panel. *J Clin Epidemiol.* 2010; 63:1169–1178. [PubMed: 20688473]
 25. Riley WT, Rothrock N, Bruce B, Christodoulou C, Cook K, Hahn EA, Cella D. Patient-Reported Outcomes Measurement Information System (PROMIS) domain names and definitions revisions: further assessment of content validity in IRT-derived item banks. *Qual Life Res.* 2010; 19:1311–1321. [PubMed: 20593306]
 26. Natesan P. Review of Bayesian item response modeling: Theory and applications. *J Educ Beh Stat.* 2011; 36:550–552.
 27. Thomas ML. The value of item response theory in clinical assessment: A review. *Assessment.* 2011; 18:297–307.
 28. Bernstein IH, Rush AJ, Thomas CJ, Woo A, Trivedi MH. Item response analysis of the Inventory of Depressive Symptomatology. *Neuropsychiatr Dis Treatment.* 2006; 2:557–564.
 29. Pallant JF, Tennant A. An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol.* 2007; 46:1–18. [PubMed: 17472198]
 30. Fliege H, Becker J, Walter OB, Bjorner JB, Klapp BF, Rose M. Development of a computer-adaptive test for depression (D-CAT). *Qual Life Res.* 2005; 14:2277–2291. [PubMed: 16328907]
 31. Gibbons RD, Weiss DJ, Kupfer DJ, Frank E, Fagiolini A, Grochocinski VJ, Bhaumik DK, Stover A, Bock RD, Immekus JC. Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatr Serv.* 2008; 59:361–368. [PubMed: 18378832]
 32. Hamilton CM, Strader LC, Pratt JG, Maiese D, Hendershot T, Kwok RK, Hammond JA, Huggins W, Jackman D, Pan H, Nettles DS, Beaty TH, Farrer LA, Kraft P, Marazita ML, Ordovas JM, Pato CN, Spitz MR, Wagener D, Williams M, Junkins HA, Harlan WR, Ramos EM, Haines J. The PhenX toolkit: Get the most from your measures. *Am J Epidemiol.* 2011; 174:253–260. [PubMed: 21749974]
 33. Identifying Core Behavioral and Psychosocial Data Elements for the Electronic Health Record: Executive Summary. http://conferences.thehillgroup.com/OBSSR/EHR2011/documents/EHR_Meeting_May_2-3-2011--Executive_Summary.pdf [Accessed Nov. 4, 2011]
 34. Booth J. A short history of blood pressure measurement. *Proc Royal Soc Med.* 1977; 70:793–799.
 35. Watson D, Clark LA, Weber K, Assenheimer JA, Strauss ME, McCormick RA. Testing a tripartite model: I. evaluating the convergent and discriminant validity of anxiety and depression symptoms. *J Abnormal Psychol.* 1995; 104:3–14.
 36. Gibbons RD, Hedeker D. Full-information item bi-factor analysis. *Psychometrika.* 1992; 57:423–436.
 37. Kendel F, Wirtz M, Dunkel A, Lehmkuhl E, Hetzer R, Regitz-Zagrosek V. Screening for depression: Rasch analysis of the dimensional structure of the PHQ-9 and the HADS-D. *J Affective Dis.* 2010; 122:241–246.
 38. Riemann D, Voderholzer U. Primary insomnia: a risk factor to develop depression. *J Affect Disord.* 2003; 76:255–259. [PubMed: 12943956]
 39. Lustberg L, Reynolds CF. Depression and insomnia: questions of cause and effect. *Sleep Med Rev.* 2000; 4:253–262. [PubMed: 12531168]

40. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, Sanislow C, Wang P. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*. 2010; 167:748–751. [PubMed: 20595427]
41. American Psychiatric Association. [Accessed Nov 4, 2011] <http://www.dsm5.org/ProposedRevisions/Pages/Cross-CuttingDimensionalAssessmentinDSM-5.aspx>
42. McCrone P. Mental health economics. Current methodological issues. *Epidemiol Psychiatr Sci*. 2011; 20:239–243. [PubMed: 21922966]

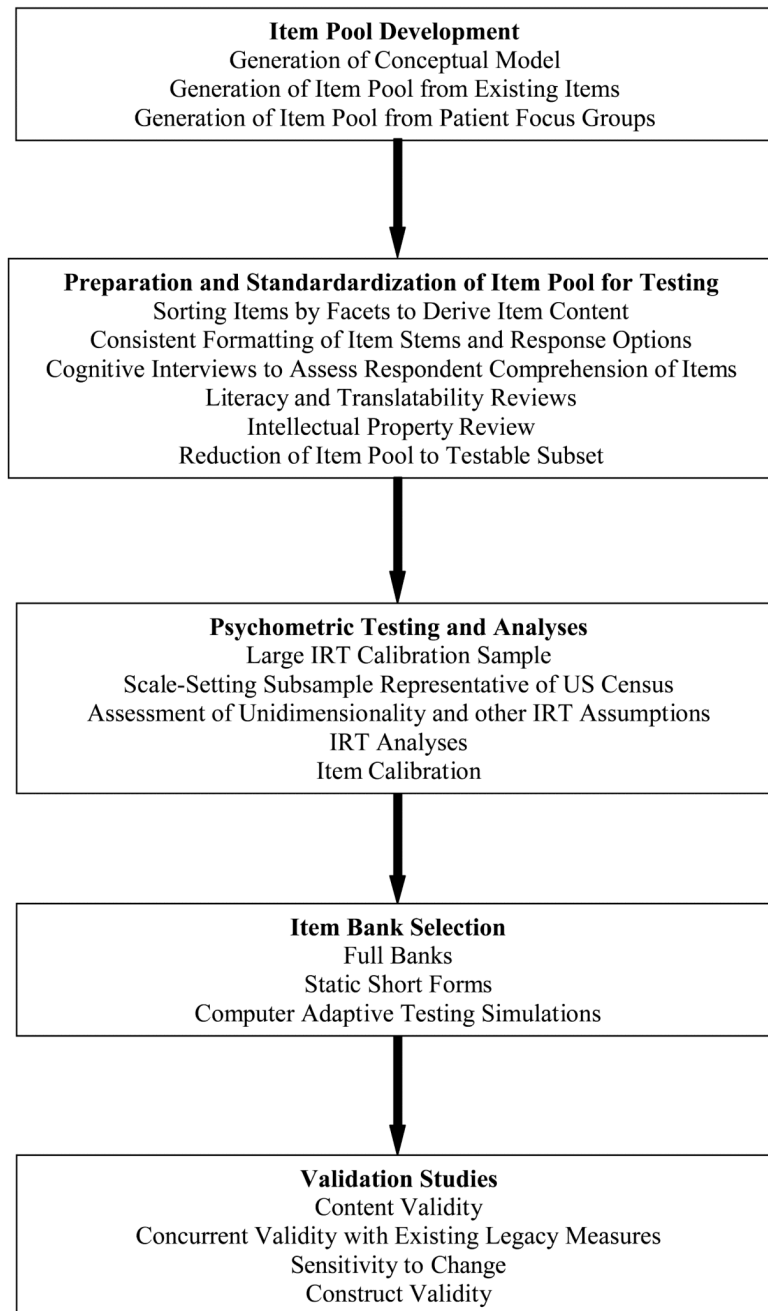


Figure 1. Flowchart Illustrating the PROMIS Item Bank Development Methodology

Table 1
Currently¹ Released PROMIS Item Banks and Scales for Adults and Children

Adult Item Banks

Physical Health

Pain Behavior
 Pain Interference
 Fatigue
 Physical Function
 Sexual Function*
 Sleep Disturbance
 Sleep-Related Impairment

Mental Health

Anxiety
 Depression
 Anger
 Psychosocial Illness Impact- Positive
 Psychosocial Illness Impact – Negative
 Alcohol Use
 Alcohol – Negative Consequences
 Alcohol – Positive Consequences
 Alcohol – Negative Expectancies
 Alcohol – Positive Expectancies
 Applied Cognition – General Concerns
 Applied Cognition – Abilities

Social Health

Ability to Participate in Social Roles and Activities
 Satisfaction with Social Roles and Activities
 Emotional Support
 Instrumental Support
 Informational Support
 Companionship
 Social Isolation

Child Item Banks

Physical Health

Fatigue
 Asthma Impact
 Upper Extremity Physical Function
 Mobility

Mental Health

Anxiety
 Depressive Symptoms
 Anger

Social Health

Peer Relationships

¹As of November 28, 2011

* The sexual function domain includes a number of item banks and pools including interest in sexual activity, vaginal discomfort, lubrication, erectile function, orgasm, global satisfaction with sex life, sexual activities, anal discomfort, therapeutic aids, and interfering factors.