The Patient-Reported Outcomes Measurement Information System (PROMIS) Seeks to Improve and Standardize Measures of Five Generic Health-Related QOL Domains

John E. Ware Jr.
University of Massachusetts Medical School

Follow this and additional works at: http://escholarship.umassmed.edu/qhs_pp

Part of the Biostatistics Commons, Epidemiology Commons, and the Health Services Research Commons

Repository Citation
http://escholarship.umassmed.edu/qhs_pp/613

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Quantitative Health Sciences Publications and Presentations by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
The Patient-Reported Outcomes Measurement Information System (PROMIS) seeks to improve and standardize measures of five generic health-related QOL domains.

John E. Ware, Jr., PhD*
Health Assessment Lab and QualityMetric Incorporated Lincoln, RI, USA

The Patient-Reported Outcomes Measurement Information System (PROMIS), which seeks to improve and standardize tools of clinical research across multiple NIH Institutes, is a 5-year project that is part of the US National Institutes of Health (NIH) Roadmap Initiative. More information about the NIH Roadmap Initiative is available on the Internet (go to http://nihroadmap.nih.gov). Briefly, PROMIS began in late 2004, with the formation of a multicenter cooperative network of US academic institutions and the NIH. Its purpose is to improve the tools for measuring patient-reported health outcomes in clinical research. Specifically, PROMIS collaborators are constructing and making accessible item banks that are useful in measuring key health status concepts across a wide variety of chronic conditions. The concepts are five generic health-related QOL domains: emotional distress, fatigue, pain, physical functioning, and social role participation. The approach is to develop a set of publicly-available computerized adaptive tests (CAT) for use in measuring these domains for the clinical research community. The PROMIS network includes clinicians, clinical researchers, and measurement experts and is organized around six primary research site (PRS) grants – Duke University, Stanford University, Stony Brook University, University of North Carolina at Chapel Hill, University of Pittsburgh, and University of Washington – and a Statistical Coordinating Center (CORE, Evanston Northwestern Healthcare and Northwestern University). NIH scientists also participate in the PROMIS network, and a Scientific Advisory Board oversees PROMIS and evaluates its progress toward its goals. The initial goals of PROMIS are documented in detail in recent newly-developed items measuring five generic health-related QOL domains (emotional distress, fatigue, pain, physical functioning, and social role participation); (2) Collect questionnaire data from large samples of adults and children including those who are well and those in the general population with selected chronic diseases; (3) Analyze questionnaire and other data to calibrate and study items in each bank; (4) Build an electronic Web-based repository to administer and report health assessments using computerized adaptive testing (CAT) methods; (5) Conduct feasibility studies to evaluate CAT-based and “static” forms derived from the PROMIS item banks along with other research projects that will benefit PROMIS; and (6) Develop a plan to establish a public-private partnership to sustain the item bank repository, improve data collection, add domains and items, test the items in new populations, maintain the system in the public domain, and expand the applications of the system in clinical research and practice.

Development of the PROMIS item banks followed a common protocol across PRS teams, within the context of an established domain hierarchy developed by the PROMIS network. The item pools include items from legacy questionnaires, (e.g., Health Assessment Questionnaire and the SF-36 Health Survey in the case of Physical Functioning), along with newly created items hypothesized to improve each bank. Item pools were developed first by a systematic search for existing items in currently available scales. Expert item review and revision was conducted to review the wording of each item and revise it as appropriate within PROMIS conventions, as well as to write additional items needed to expand the item pool. Focus groups and cognitive interviews were used to provide patient input into the domain definition and item wording process.

In addition to the qualitative studies referred to above, large-scale PROMIS data collection (N = 20,000+ has also been completed and initial data analyses have already yielded preliminary item calibrations that will be posted on the PROMIS website for all five item banks along with information on self-reported sociodemographic and clinical characteristics of study participants. Analyses of the PROMIS item banks are following a common protocol, including evaluation of...
data quality (e.g., logic and range checking) and descriptive item statistics (e.g., frequencies, means), tests of item response theory (IRT) model assumptions (e.g., unidimensionality, local independence, monotonicity), evaluation of model fit, and tests for differential item functioning.

There are good reasons to expect that this first wave of large-scale PROMIS data collection will increase our understanding of the strengths and weaknesses of widely-used legacy tools and will also enable noteworthy advances in the measurement of at least some key health domains. My confidence rests in part on the recently completed applications of IRT methods to a pre-PROMIS bank of 136 physical function items selected from nine widely-used questionnaires and administered (in various permutations) to nearly 18,000 US general population adults. Using methods similar to those we are using to analyze the first wave of PROMIS data and previously used to evaluate headache-related disability measures, my colleagues and I successfully cross-calibrated 70 of the 136 items from nine independently developed physical functioning questionnaires and administered (in various permutations) to nearly 18,000 US general population adults. Using methods similar to those we are using to analyze the first wave of PROMIS data and previously used to evaluate headache-related disability measures, my colleagues and I successfully cross-calibrated 70 of the 136 items from nine independently developed physical functioning questionnaires and administered (in various permutations) to nearly 18,000 US general population adults. Using methods similar to those we are using to analyze the first wave of PROMIS data and previously used to evaluate headache-related disability measures, my colleagues and I successfully cross-calibrated 70 of the 136 items from nine independently developed physical functioning questionnaires and administered (in various permutations) to nearly 18,000 US general population adults.

Results will be published this year in the physical functioning questionnaires. The inclusion and cross-calibration of questionnaire items from legacy tools within PROMIS item banks will greatly facilitate the interpretation of improved PROMIS metrics. Because IRT models do not yield scores that are directly interpretable, their interpretation must be determined from experience. The concurrent cross-calibration strategy of maintaining direct linkages between legacy and improved PROMIS score estimates will provide a basis for comparing new results with those from literally thousands of prior clinical studies and will facilitate much needed meta-analyses of those studies. As a result, the accumulation of interpretation guidelines will continue without interruption rather than starting all over again. An additional payoff from this approach will be a better understanding of why some scales hypothesized to measure the same domain sometimes yield different results in clinical studies.

Thus, the PROMIS approach of combining modern psychometric (i.e., IRT) and computerized dynamic methods to achieve PRO assessments that are both more practical and more precise is likely to be successful. Accordingly, the challenges faced by PROMIS, in my opinion, may be more political and administrative than scientific. What is the best way to create and promote measurement standards while making them readily available? To be well understood and accepted, standards must be well documented. To protect their scientific validity, standards and the trademarks used to label them must be copyrighted and registered so that all will know when the standards have and have not been properly reproduced. Much like the temperature scales constructed by Celsius and Fahrenheit were standardized and cross-calibrated hundreds of years ago with substantial advantages, the cross-calibration and standardization of legacy tools and improved PROMIS metrics today will substantially enable their proper use and meaningful interpretation going forward.