An experience with the Patient-Reported Outcomes Measurement Information System: Pros and cons and unanswered questions

Terry A. Badger, University of Arizona
Margaret Heitkemper, University of Washington
Kathryn A. Lee, University of California San Francisco
Deborah W. Bruner, Emory University

Journal Title: Nursing Outlook
Volume: Volume 62, Number 5
Publisher: Elsevier | 2014-09-01, Pages 332-338
Type of Work: Article | Preprint: Prior to Peer Review
Publisher DOI: 10.1016/j.outlook.2014.06.009
Permanent URL: https://pid.emory.edu/ark:/25593/tvj4r

Final published version: http://dx.doi.org/10.1016/j.outlook.2014.06.009

Copyright information:
© 2014 Elsevier Inc.
This is an Open Access work distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Accessed January 6, 2020 7:07 AM EST
The Experience of PROMIS: Pros and Cons and Unanswered Questions

Terry Badger, PhD, RN, PMHCNS-BC, FAAN1 [Professor], Margaret Heitkemper, PhD, RN, FAAN2 [Professor], Kathryn Lee, RN, CBSM, PhD, FAAN3 [Professor], and Deborah Watkins Bruner, PhD, RN, FAAN4 [Professor]

1Department of Psychiatry, University of Arizona, Tucson, AZ
2Dept. of Biobehavioral Nursing and Health Systems, University of Washington, Seattle, WA
3University of California, San Francisco, CA
4Emory University, Atlanta, GA

Abstract

The goal of PROMIS© (Patient-Reported Outcomes Measurement Information System) is to create efficient, reliable, and valid assessments of adult and child health. The nursing science literature in which PROMIS measures are used is rapidly expanding. Investigators have been encouraged to consider the integration of PROMIS measures into both descriptive studies and clinical trials. To do this has created opportunities and challenges for investigators. This paper highlights three projects to illustrate the perspectives of nurse scientists who incorporated PROMIS measures into their research. The first project describes advantages of PROMIS to allow for comparisons of a study population with a national sample and to compliment legacy measures. The second project examines issues in the translation of tools for region-specific Hispanic populations. The third project provides a perspective on utilization of PROMIS measures to capture cancer-related fatigue and to develop new components of a sexual function scale. As indicated by these three examples, nurse scientists can contribute an important role in moving the PROMIS initiative forward. Results from these types of projects also move symptom science forward within a more interdisciplinary approach to common measures of interest.

Keywords

PROMIS; irritable bowel; pain; fatigue; sexual dysfunction; Spanish translation; depression; symptoms

© 2014 Mosby, Inc. All rights reserved.

Please address all correspondence to: Margaret Heitkemper, PhD, Dept. of Biobehavioral Nursing and Health Systems, University of Washington, Box 357266, Seattle, WA 98195-7266, (206)-221-2408 (voice), (206)-543-4771 (fax), heit@uw.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Background

This paper provides information shared at the October 2013 Council for the Advancement of Nursing Science (CANS) Conference Innovative Approaches to Symptom Science: Measurement and Analysis. The authors of this paper are investigators who incorporated either one or more PROMIS tools into their NIH-funded research projects. Abstracts of these studies are presented in Tables 1 and 2.

PROMIS Advantages

The use of the PROMIS measures by bio-behavioral nurse scientists who are focused on understanding the biological underpinnings of symptom reports provides several advantages. A key advantage is the ability to compare symptoms reported by one’s sample to a national data base (one of the intended purposes of the PROMIS development). Project 1, “The Pathways to Abdominal Pain,” provides an example of this advantage (Table 1).

Project 1 Background

Project 1 was an ARRA-funded project with the aim of collecting data in children and adults with irritable bowel syndrome (IBS). IBS is a common functional gastrointestinal (GI) disorder that affects 10–17% of the American population. Abdominal discomfort/pain relieved by an alteration in bowel function (diarrhea, constipation, mixed is a key diagnostic criteria (part of the Rome III) guidelines (Drossman, Corazziari, Delvaux, Spiller, Talley, Thompson, Whitehead, 2006). Previous studies by the investigative team utilized ‘legacy’ measures including a retrospective measure of GI distress and a daily diary to assess twenty-six symptoms (e.g., gastrointestinal, somatic, psychological distress) on a scale of ‘0’ not present to ‘4’ extreme (Heitkemper, Cain, Shulman, Burr, Poppe, Jarrett, 2011 & Cain KC, Jarrett ME, Burr RL, Rosen S, Hertig VL, Heitkemper MM, 2008).

Abdominal Pain/Discomfort Symptom Measures

When the investigators began this project, PROMIS measures were ‘new’ and offered the added advantage of comparing reports of pain in this population to that of a national sample. For example, abdominal pain/discomfort is a core symptom specified in the Rome III criteria for IBS. Numerous cross sectional studies performed by us and others have validated that daily reports of abdominal pain/discomfort are higher in persons with IBS when compared to age- and gender-matched healthy controls. In addition, several older survey studies demonstrated that patients with IBS report more discomfort and disability when compared to patients with other chronic illnesses including diabetes and inflammatory bowel disease (Haagsma, Siersema, De Wit, Havelaar, 2010; Leong, Barghout, Virnbaum, Thibeault, Ben-Hamadi, Frech, Ofman, 2003; Tkalcić, Hauser, Stimac, 2010; Lea, Whorwell, 2001; Gralnek, Hays, Kilbourne, Naliboff, Mayer, 2000). However, for the most part these studies utilized disparate measures of abdominal pain to capture symptom intensity or frequency as well as quality of life (QOL) across populations. Incorporating PROMIS measures into Project 1 at the onset provided the opportunity to compare the pain experience in our sample with the pain experience in a national sample. The short forms of pain interference and behaviors were used in combination with legacy measures.
Using the PROMIS website and manual for scoring we determined that the IBS group had a PROMIS-Pain Behavior score of 54.1 (SD=8.2) and a PROMIS-Pain Interference score of 56.8 (SD=7.7) relative to the population standard score of 50. Thus, as a group our sample of IBS patients had higher pain behaviors and greater pain interference compared to the general population. In addition, there were positive correlations among PROMIS pain behavior and PROMIS pain interference measures with our legacy measures of daily diary abdominal pain and abdominal pain recall ($r=0.430$, $r=0.485$, respectively) and QOL ($r=−0.550$).

**Additional Issues to Consider from Experience with Project 1**

One benefit of incorporating the PROMIS measures into studies of IBS patients is that permits comparisons of different symptoms (e.g., pain, fatigue) with larger combined datasets. For example, the prevalence of other visceral and somatic pain conditions including fibromyalgia, chronic fatigue syndrome, interstitial cystitis as well as headache and musculoskeletal pain are higher in women with IBS compared to healthy gender-matched controls.\(^{1}\) With multiple investigators from a variety of laboratories incorporating PROMIS measures for pain, sleep, sexual dysfunction, and fatigue, Big Data can emerge to answer such questions as, “do patients with IBS plus comorbidities report more fatigue, pain, or sexual dysfunction relative to those who only report IBS?” Such information can be important to designing therapeutic trials that go beyond a single symptom or condition. For example, the patient with IBS who reports high levels of sleep disturbance may be more likely to benefit from incorporation of sleep hygiene. Whether data from the PROMIS measures can add to the information needed for hypothesis development for exploration of root causes of symptom distress remains to be determined.

Can PROMIS measures replace legacy measures such as a daily diary? Similar to some other legacy tools (IBS Disease Questionnaire), which capture retrospective symptom frequency and severity, the PROMIS measures do not inform us about the temporal relationship of the symptoms in the way daily diaries do (the primary legacy measure of our studies). For example, we know that GI symptoms for many women with IBS are exacerbated or amplified at specific times during the menstrual cycle.\(^{2,3}\) Similarly poor sleep reported on a diary is related to an increase in GI symptoms (also reported on a diary) the next day.\(^{4}\) Quantitative information such as timing of symptoms with respect to meals and stress would be best provided by diaries or technology that allows for real-time data collection.

Because only the short forms were used in this project, the issue of subject burden did not surface. However, use of the composite tools with 46 or 57 items, in addition to legacy tools, remains to be evaluated for subject burden in this population. Of note is that the gastrointestinal symptom PROMIS measure is currently in development, and the PROMIS website contains current information on these and other tools under development ([prosettastone.org](http://prosettastone.org)).
In summary, our initial belief was that comparisons with a national database would be important to an adequate description of the sample and to eventual data sharing.

**PROMIS Challenges**

In Project 2, the nurse scientist was involved in testing the effectiveness of two culturally-sensitive, expedient psychosocial interventions to improve quality of life for Latinas with breast cancer and their supportive partners. While Spanish language versions of PROMIS measures are available and widely used, there are challenges for nurse scientists to consider. This challenge is highlighted in Project 2.

**Project 2 Background**

Project 2 involved a sample of Latinas in southwestern US. They were enrolled in a study of low income women with breast cancer and their supportive partners. Most of the Latinas who participated in this study had families that had lived in the southwestern US for several generations, and the majority had cultural ties with the Sonora region of Mexico. These ties influenced the ability of the participants to understand some translations of the PROMIS items. Table 2 provides an abstract of the study.

**Depressive Symptoms and Quality of Life Measures**

For several constructs (e.g., depression), legacy measures were used in addition to the PROMIS measures. The rationale for the use of legacy measures for constructs such as depression was to enable investigators to compare across their own studies since these legacy tools had been used for prior studies. The advantages of these types of comparisons have been previously illustrated in Project 1.

**Language Challenges**

The major challenge encountered in Project 2 was adopting the PROMIS Spanish language versions of the symptom measures. The Spanish PROMIS instruments were developed using a universalist approach to translation (nihpromis.org/measures/translation); an approach where one word is selected to be suitable for all Spanish speakers across all regions within the US. The instruments were translated-back-translated with approximately 4000 US Spanish speakers (<10% of the sample were Hispanic) (Pilkonis, Choi, Reise, Stover, Riley, Cella, 2011).

The universalist approach to translation does not include region-specific language differences because it is thought to introduce bias, and decrease standardization (nihpromis.org/measures/translation), making it more difficult to compare results across studies. However, Badger and her Project 2 team found that some universalist item translations did not capture the intent of the question, or that items were exact translations from English and thus not understood by this sample.

This experience raised the concern about consistent linguistic and cultural equivalency across the PROMIS items. This concern may necessitate greater attention to determining tool validity in specific US populations. This project employed bilingual bicultural data collectors who were all born and lived locally, with cultural ties to Sonora Mexico. Very
quickly these data collectors began to question some of the item translations, as did many participants who found some items confusing. After pilot testing the instruments with 25 dyads, it was clear that some changes were needed to ensure cultural and linguistic equivalence for this Latina sample. Data collectors were routinely interpreting what a particular word or phrase meant for some items, indicating that a wording change was needed to capture valid and reliable data.

The universalist translation was clear for a majority of items, yet some items were not culturally and linguistically equivalent for this regional Latina population. Table 3 illustrates the modifications to several items from the Spanish PROMIS measures that were required in order to establish adequate reliability and validity in this study population.

**Additional Issues to Consider from Experience with Project 2**

An additional challenge to using the PROMIS with this low income population was the Computer Adaptive Testing (CAT) data collection method. With the CAT methodology, the subsequent question in the on-line survey will vary based on the participant’s response to the prior question. Although the CAT system can reduce participant burden and does provide a precise measure using a battery of items, the issue of the suitability of this method for all populations should be carefully evaluated. For example, in Project 2, static measures were selected due to the participants’ lack of access to computers, and due to data collection by telephone.

In summary, the PROMIS measures provide reliable and valid measures, allowed for comparisons between a Latina sample of women with breast cancer and a national sample, and were highly correlated with the legacy measures used in Project 2. Based on the experiences with PROMIS measures in Project 2, nurse scientists moving forward should be encouraged to consider using these measures in their research. This initiative will allow for meaningful comparison across studies on phenomena of interest and ultimately lead to better questions that can be answered with Big Data merged from many investigators. Although the Project 2 research team made some regional-specific changes to the Spanish measures in order to increase the participant’s ability to respond without explanation was likely to introduce error, it should be noted that most of the Spanish measures performed quite well and required no changes.

**Potential New Opportunities for Nurse Scientists using PROMIS measures**

Understanding the pathophysiology and management of symptoms in patients with chronic illness is a common pursuit of nurse scientists, and necessitates the use of valid and reliable instruments. Project 3 involves populations of cancer patients and the unique symptom experiences that have high potential for future programs of research for nurse scientists.

**Project 3 Background**

Cancer-related fatigue (CRF) is the most common cancer-related symptom regardless of cancer site or treatment. For the most part CRF is unrelated to sleep and activity. Fatigue is also a common side effect of other medical conditions including depression, heart disease and insomnia, which may also be present in the person with cancer. CRF is difficult to
assess and treat due to its non-specific nature and inconsistent measurement. In studies that ask only about self-reported CRF presence or severity, the prevalence is usually high (70–99%). (Curt, Breitbart, Cella, 2000) In contrast, the prevalence is moderate to high (30–70%) in studies requiring CRF to exceed a threshold of severity, duration, or functional impairment and even lower (15–30%) in studies that defined CRF as a syndrome (Barsevick, Irwin, Hinds, Miller, Berger, Jacobsen, Cella, 2013; Servaes, Prins, Verhagen, Bleijenberg. 2002; Andrykowski Donovan, Laronga, Jacobsen. 2010).

**Fatigue measures**

The PROMIS-Fatigue (PROMIS-F) measure was one of the first PROMIS scales to be developed. The intent was to develop a tool that would standardize fatigue assessments across adult populations. The PROMIS-F was tested using comparisons with common validated legacy measures including the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) and the Medical Outcomes Short Form (SF-36v2- Vitality subscale). The instruments were highly inter-correlated and performed equally well in detecting the observed associations. PROMIS-F also demonstrated good precision in another study with reliability greater than 0.9 in both healthy control and clinic patient populations (Christodoulou, Schneider S, Junghaenel DU, Broderick JE, Stone AA. 2013). However, many investigators have built programs of research on validated legacy measures of symptoms like fatigue. Thus, the uptake of PROMIS-F measures by researchers has not been widely embraced. For example, in a recent systematic review of 40 CRF instruments, none included the PROMIS-F. Interestingly, five of the most commonly used instruments for CRF each had nearly identical psychometric properties (Seyidova-Khoshknabi, Davis, & Walsh, 2011).

Despite the lack of rapid incorporation into studies of CRF, the PROMIS-F seems to hold promise for improving the consistent measurement of CRF. Recommendations from a 2010 Clinical Trials Planning Meeting on CRF convened by the National Cancer Institute included support for use of PROMIS-F. The meeting goal was to examine issues, including measurement, to advance CRF science. Participants included representatives of academia, community oncology, government, the pharmaceutical industry, and patient advocates. Recommendations that emerged from the meeting suggested that researchers adopt the PROMIS-F as the standard measure for CRF (Barsevick et al., 2013). However, it is important to note that this recommendation was made without a single clinical trial using PROMIS-F with the cancer population yet published. This suggests that further validation of PROMIS measures, particularly within the context of CRF, would be welcomed among interdisciplinary researchers.

**Sexual Dysfunction Measures**

In addition to fatigue, another common symptom that lacks a common definition and metric is sexual dysfunction. Sexual dysfunction after cancer treatment ranges from 40 to 100% depending on the tumor site and treatment as well as the metric used for assessment. It is a significant concern during the entire treatment trajectory, from diagnosis through survivorship. Development of a comprehensive self-report tool to measure sexual dysfunction in the cancer population is important in order to capture unique difficulties that
are relevant in this population, such as effects of particular chemotherapeutic agents, anti-androgens, or surgeries. Unlike CRF, cancer-related sexual dysfunction has received less attention in the literature and measures are even more variable and not as well developed or validated. Entire aspects of sexual function have no cancer-specific metric. For example, oral sex, emerging as an important area of research to improve our understanding of human papilloma virus (HPV)-positive head and neck cancers, is an area in need of developing and testing specific common metrics.

The PROMIS-Sexual Function (SexFS) measures have the potential to offer researchers and clinicians a comprehensive, reliable and valid set of tools to measure self-reported sexual function and satisfaction among men and women living with cancer (Flynn, Lin, Cyranowski, Reeve, Reese, Jeffery, Weinfurt, 2013). The internal consistency reliability (Cronbach alpha coefficients) for men and women range from 0.87–0.95, and test-retest reliability interclass correlation coefficients range from 0.71–0.87.

Unlike the PROMIS-F for fatigue, the PROMIS-SexFS offers new and previously unavailable modules/domains, such as a measure of oral sexual engagement and satisfaction. However, as with CRF, one challenge in adopting the PROMIS-SexFS measures is that there are commonly used legacy instruments for specific aspects of sexual function, such as the International Index of Erectile Function, considered a standard measure in both research and clinical practice.

In summary, PROMIS-measures in general and SexFX in particular, may challenge established investigators to abandon their use of standard legacy scales in favor of enhanced ability for comparisons across studies and populations. As these comparisons become available, and Big Data contain merged data for these types of symptoms, this challenge may be attenuated. As scores are mapped from PROMIS-SexFS to these more familiar legacy instruments for comparisons, it remains to be determined whether the PROMIS-SexFS is clinically relevant to the same degree as the legacy tools.

**Summary**

As nurse scientists, the authors concur that it is important to support moving the PROMIS initiative forward. The notion of having a common currency with regard to symptoms and symptom impact on quality of life is laudable, but more work still needs to be done. For most investigators, the utilization of PROMIS measures will be done in concert with the continued use of more familiar legacy tools and the incorporation of selected relevant short forms or composite tools. Among nurse scientists there are aspirations for a national symptom data base to be utilized by trainees and interdisciplinary colleagues. Incorporation of the PROMIS measures, along with other NIH tool kits such as the Neuro-QOL, may be a step towards developing Common Data Elements for nurse scientists. The ultimate goal is to enhance the translation of valid research data into clinical practice and better outcomes, and use of PROMIS measures will continue to have advantages as well as challenges.
References


Abstract of a study in which PROMIS pain measures were used in women with Irritable Bowel Syndrome.

### Table 1

A Comparison of PROMIS Pain Measures with Legacy Measures in Women with Irritable Bowel Syndrome (IBS)

**Introduction:** Patients with IBS are more vigilant to pain-associated stimuli. The purpose of this pilot study was 1) to compare women with IBS to a national data base on the PROMIS pain behavioral and interference measures; 2) to determine the relationship of PROMIS pain behaviors and interferences measures with legacy measures of pain and quality of life (QOL) and pain sensitivity using a test central pain modulation (CPM).

**Methods:** Women (n=20) between the ages of 18 and 45 with a medical diagnosis of IBS (Rome III diagnosis) were compared to healthy controls (HC) women recruited from the community. Women completed the PROMIS-Pain Behavior and the PROMIS-Pain Interference short forms (8 items each) as well as a 14-day symptom diary. On the diary GI, psychological distress, and abdominal discomfort symptoms were rated on a 0–4 scale based on severity (none, mild, moderate, severe, very severe). Diary data were collapsed into percent of days with moderate–very severe report of symptoms and life interference. Pain Behavior was measured by the Pain Behavior short form which asks about common pain behaviors that can be observed, behaviors associated with pain severity, verbal reports of pain (rated from 1 ‘not at all’ to 5 ‘very much’), and one social item. Internal consistency for this study was \( r = .89 \). Pain Interference was measured by the PROMIS-Pain Interference Short Form which asks about the consequences of pain on five relevant aspects including: social, cognitive, emotional, physical and recreational activities over the past 5 days. Internal consistency for this study was \( r = .91 \). The CPM was tested with a counter-irritation approach in a laboratory between 8 and 10 am on a follicular phase day. The study protocol and findings are described in further detail in (Jarrett, Shulman, Cain, Deecheakawan, Smith, Richebe, Eugenio, Heitkemper, 2014).

**Results:** There were positive associations among PROMIS pain measures and legacy measures of abdominal pain (.416-.430). There were significant inverse relationships with the legacy QOL measure. During the CPM testing, there was a positive relationship between baseline pain sensitivity and PROMIS pain interference measure.

**Implications:** The PROMIS pain measures allow for characterization of women with IBS in comparison to a national population. In addition, there is good correlation with legacy measures in this population.
Abstract of a study in which PROMIS measures were used for Latinas with breast cancer.

<table>
<thead>
<tr>
<th>Interventions to Improve Quality of Life for Latinas with Breast Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong>: To test the effectiveness of two culturally-sensitive, expedient psychosocial interventions to improve quality of life (psychological, physical, social and spiritual well-being) for Latinas with breast cancer and their supportive partners.</td>
</tr>
<tr>
<td><strong>Design</strong>: This study is a randomized clinical trial design with measures obtained at baseline, after the 8-week intervention, and 2- and 4-months post intervention. Latinas and their supportive partners are randomly assigned to either a telephone delivered interpersonal counseling intervention or a health education intervention. All intervention sessions and measurements are conducted by telephone. This study is conducted in the southwestern US and all written materials are provided in either English or Spanish.</td>
</tr>
<tr>
<td><strong>Sample</strong>: Thus far, the sample includes 72 Latinas, with a mean age of 49 years, 52% were married/partnered with 80% having household incomes of ≤ $30K year. Sixty seven percent of the Latinas had ≤ High School education and 32% were unemployed but seeking and 42% were disabled/unable to work. In this sample, 42% had complete mastectomy and only 2% underwent breast reconstruction. Most have public insurance which does not allow for breast reconstruction. Forty percent of the women had stage 3 or 4 breast cancer. The majority (80%) did not participate in any supportive treatment (e.g., support groups). The supportive partners were primarily Latino/a, and have similar incomes, educational and employment histories. Of note, 64% of the supportive partners were female relatives.</td>
</tr>
<tr>
<td><strong>Results</strong> (relevant to PROMIS): PROMIS instruments had good reliability (α ≥ .92), but some Spanish translations were questioned.</td>
</tr>
</tbody>
</table>
Table 3
Translations of PROMIS Measures: English, Universalist, and Region-Specific Change Items.

<table>
<thead>
<tr>
<th>PROMIS</th>
<th>English Version</th>
<th>Universalist Translation</th>
<th>Region Specific Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>I felt uneasy</td>
<td>Me sentí intranquilo</td>
<td>Inquieta (word used for uneasy in local region)</td>
</tr>
<tr>
<td>Fatigue</td>
<td>How much were you bothered by your fatigue on average?</td>
<td>En qué medida la molestó el agotamiento en promedio?</td>
<td>En promedio que tanto le molesto el agotamiento.</td>
</tr>
<tr>
<td>Social Support</td>
<td>I have someone who will listen to me when I need to talk</td>
<td>Tengo quien me escuche cuando necesito hablar</td>
<td>Tengo quien a alguien me escuche cuando necesito hablar</td>
</tr>
</tbody>
</table>

In this example, “on average” was placed first.