

Measure validation is an ongoing process: the Functional Assessment of Cancer Therapy-Breast Symptom Index as a case example

Sofia F. Garcia^{1,2}, Elizabeth A. Hahn^{1,2}

¹Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA; ²Robert H. Lurie Comprehensive Cancer Center of Northwestern University, Chicago, IL, USA

Corresponding to: Sofia F. Garcia, Ph.D. Northwestern University Feinberg School of Medicine, 625 N. Michigan Ave., Suite 2700, Chicago, IL 60611, USA. Email: sofia-garcia@northwestern.edu.



Submitted Sep 28, 2012. Accepted for publication Oct 30, 2012.

DOI: 10.3978/j.issn.2224-5820.2012.10.07

Scan to your mobile device or view this article at: <http://www.amepc.org/apm/article/view/1198/1803>

Patient-reported outcomes (PROs) such as measures of symptoms status or health-related quality of life (HRQL) have demonstrated their value to clinical practice and research over the last three decades. Routine assessment of HRQL has been shown to enhance clinical care by improving problem identification and management, patient-provider communication, and patient satisfaction with care (1-3). PROs are particularly relevant in palliative medicine when symptom relief and maintenance of function are primary care objectives. HRQL and symptom measures are also increasingly being used to evaluate treatment efficacy in clinical practice and research, including trials of new drugs and other medical products. This proliferation of PROs and their high-stakes uses has led to the establishment of guidelines for their development, modification, validation and applications (4-7). A recent article by Lee *et al.* in the *Journal of Pain and Symptom Management* illustrates how even measures such as the Functional Assessment of Cancer Therapy-Breast Symptom Index (FBSI), which have been developed from validated parent measures, benefit from additional validation after modification and when used with new populations (8). Their work exemplifies how researchers can fill important gaps in the ongoing assessment of a PRO measure's validity.

The FBSI is a prime example of a PRO measure that has undergone sequential efforts to establish its validity. The FBSI development process emphasized content validity across all of its stages, paying particular attention to incorporating patient and expert input via qualitative research. Items in the FBSI were drawn from the Functional

Assessment of Cancer Therapy-General (FACT-G), the FACT-Breast (FACT-B) and other components of the Functional Assessment of Chronic Illness Therapy (FACIT) measurement system. FACIT measures were developed by incorporating patient- and clinician-identified symptoms and concerns, as well as literature reviews, in order to build in content validity (9-11).

The FACT-G (now in Version 4) is a 27-item compilation of general questions divided into four primary HRQL domains (Physical Well-Being, Social/Family Well-Being, Emotional Well-Being, and Functional Well-Being) and has been well validated in cancer populations, other chronic illness conditions and in the general population (using a slightly modified version) (9,12). The FACT-B (Version 4) is a 37-item measure that contains the four general FACT-G subscales along with the Breast Cancer-Specific subscale that assesses symptoms/concerns of particular relevance to breast cancer (e.g., body image, arm swelling and tenderness). The FACT-B has demonstrated good reliability, validity, sensitivity to change, and ease of use (10). It has been translated from English into over 50 languages using the sequential approach for the development of PROs intended for international use that is employed for all FACIT translations. This approach involves iterative forward-backward translations, extensive qualitative item review and evaluation by bilingual health professionals, as well as qualitative input from patients (13-15). The psychometric measurement properties of several translated versions of the FACT-B have been evaluated using different methods (internal consistency and test-retest reliability;

responsiveness to change; convergent, divergent and known groups validity; factor analysis and structural equation modeling; differential item functioning) and they have been found to be generally equivalent to the English-language FACT-B, reliable, responsive to change, valid and suitable for use in international studies (16-21).

The FBSI was developed to both respond to requests from clinical and regulatory communities for symptom-based measures that can be interpreted more clearly than multi-dimensional HRQL measures as well as to produce a measure with decreased administration time and response burden (1-3 minutes versus 5-10 for the FACT-B). Efforts to ensure the content validity of the FBSI included a survey of National Comprehensive Cancer Network physician and nurse experts asked to identify priority symptoms in evaluating breast cancer treatments (22). In a preliminary validation study conducted within a larger clinical trial, the FBSI demonstrated acceptable reliability and validity, and a minimally important difference score (2-3 points) was identified that can be used to better interpret scores (23). However, that study included not the full eight-item FBSI but an abridged six-item version that was available at the time of the trial. Therefore, the recent study by Lee *et al.* is the first to evaluate the eight-item FBSI while also building upon previous studies examining the validity of the Chinese version of the FACT-B and its comparability to the original English version (8,18,19).

Lee *et al.* selected a good set of standard methods to evaluate the reliability and validity of the eight-item FBSI (8). Given that test-retest reliability is often excluded from validation studies, its inclusion in the study provides useful information for future users of the FBSI. It should be noted that the reliability coefficients reported by Lee *et al.* do not meet standards for individual measurement. That is to say that their results indicate that the English and (simplified) Chinese versions of the eight-item FBSI can be used to measure groups of patients reliably but are not appropriate for individual screening or decision-making. However, they did find both the English and Chinese versions of the eight-item FBSI to demonstrate known-groups validity when comparing the scores of patients with and without evidence of disease as well as those receiving or not receiving chemotherapy or radiation. The English version of the FBSI was responsive to changes in patients' performance status whereas the Chinese version was responsive only to declines in performance status. The investigators also performed receiver operating characteristic curve analyses to compare the FBSI's discriminative ability and

responsiveness to change (in terms of performance status, evidence of disease and treatment status) to those of the FACT-B and found the two measures to be comparable. Lee *et al.* also found that both language versions of the FBSI demonstrated convergent and divergent validity based on correlations with the FACT-B subscales.

Given that race/ethnicity differed between the English- and Chinese-language groups, the investigators conducted some exploratory regression analyses restricted only to the baseline data of the ethnic Chinese participants (8). Using that subgroup's data and entering all patient characteristics other than race as covariates, they found no significant difference between the English- and Chinese-language groups for the eight-item FBSI but did note a minor difference between the two language groups for one item. Lee *et al.* took an important first step in not confounding race/ethnicity when examining language-based differences in responses to the FBSI. However, it would have been interesting to know exactly how the English- and Chinese-language groups compared on sociodemographic and clinical variables when limiting the sample to only the ethnic Chinese participants. Such comparisons would be informative for developing a model-building strategy for regression analyses. Careful consideration of confounding through development of a multivariable model may provide some interesting insights into differences and similarities of the two language groups.

In addition, future work could use item response theory (IRT) methods to examine the cross-cultural equivalence of the different language versions of the FBSI by identifying any significant differential item functioning that would constitute measurement bias and by determining what items can be used in cross-cultural comparisons or when pooling data in international research (16). IRT methods offer solutions to two limitations faced by classical test theory approaches, in which items in a measure may function differently depending on the samples tested and in which a respondent's score may vary depending on the particular items in the measure. Being able to evaluate items' cross-cultural performance in a way that is independent of the exact grouping of items included in the measure tested is particularly important for measures such as the FBSI, which include items that have been tested within various measures (i.e., the FACT-G, FACT-B, and different versions of the FBSI).

As mentioned by Lee *et al.*, since their evaluation of the eight-item FBSI, the National Comprehensive Cancer Network-FBSI (NFBSI-16) was developed using methods consistent with recent regulatory guidance for

PROs as endpoints in clinical trials and it underwent a preliminary evaluation of its validity (24). Development of the NFBSI-16 emphasized patient input via interviews and surveys asking patients what symptoms they considered most important and then evaluated those results alongside those gathered from oncology clinicians surveyed during the development of the eight-item FBSI. The NFBSI-16, which includes all eight items from the original FBSI and eight additional items from FACIT measures (13 from the FACT-B), is structured as three subscales (Disease-Related Symptom, Treatment Side-Effect, and General Function and Well-Being). There is preliminary support for the NFBSI-16's internal consistency reliability, convergent validity (demonstrated by associations with the FACT-B and EQ-5D) and known-groups validity (using performance status). However, like other studies examining the validity of versions of the FBSI and its parent measures, the preliminary validation of the new NFBSI-16 had limitations. The generalizability of those study results is limited by the small and homogenous sample and the omission from the validation analyses of the three NFBSI-16 items not in the FACT-B. Therefore, there is ample room for future studies to validate the full-length NFBSI-16 with larger and more diverse samples.

As outlined above, the FBSI exemplifies the evolution of a measure to meet the needs of clinical, research and regulatory communities for reliable and valid PRO measures that are easy to interpret, quickly administered, and available in multiple languages. As PROs are progressively becoming a standard component of clinical research and practice, it is increasingly important to ensure that measures are reliable and valid for their intended applications. Efforts should be made to build content validity into PRO measures from their initial development, with a special emphasis placed on the inclusion of patient input. As a next step, studies evaluating PRO's measurement properties should benefit from following guidelines for establishing evidence of their reliability and validity (4-7). Additional validation efforts must be made as measures are translated into other languages or are otherwise modified, with special attention paid to establishing that shortened versions are as accurate as longer versions and that translated versions are cross-culturally equivalent. The paper by Lee *et al.* demonstrates how studies can successfully contribute to a growing body of evidence for a measure's validity (8). Clearly, as illustrated by the case of the FBSI, establishing a PRO's validity is a process that can span decades as a measure is modified and used with different populations.

Acknowledgements

Disclosure: The authors declare no conflict of interest.

References

1. Magruder-Habib K, Zung WW, Feussner JR. Improving physicians' recognition and treatment of depression in general medical care. Results from a randomized clinical trial. *Med Care* 1990;28:239-50.
2. Detmar SB, Muller MJ, Schornagel JH, et al. Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. *JAMA* 2002;288:3027-34.
3. Velikova G, Booth L, Smith AB, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *J Clin Oncol* 2004;22:714-24.
4. U. S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, et al. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims. Rockville, MD: U.S. Dept. of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health, 2009.
5. Rothman M, Burke L, Erickson P, et al. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. *Value Health* 2009;12:1075-83.
6. Wild D, Eremenco S, Mear I, et al. Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force report. *Value Health* 2009;12:430-40.
7. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002;11:193-205.
8. Lee CE, Ng R, Wong NS, et al. Measurement Properties of the Eight-Item Abbreviated Functional Assessment of Cancer Therapy-Breast Symptom Index and Comparison With Its 37-Item Parent Measure. *J Pain Symptom*

- Manage 2012. [Epub ahead of print].
9. Cella DF, Tulsy DS, Gray G, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol* 1993;11:570-9.
 10. Brady MJ, Cella DF, Mo F, et al. Reliability and validity of the Functional Assessment of Cancer Therapy-Breast quality-of-life instrument. *J Clin Oncol* 1997;15:974-86.
 11. Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. *Health Qual Life Outcomes* 2003;1:79.
 12. Webster K, Odom L, Peterman A, et al. The Functional Assessment of Chronic Illness Therapy (FACIT) measurement system: validation of version 4 of the core questionnaire. *Qual Life Res* 1999;8:604.
 13. Bonomi AE, Cella DF, Hahn EA, et al. Multilingual translation of the Functional Assessment of Cancer Therapy (FACT) quality of life measurement system. *Qual Life Res* 1996;5:309-20.
 14. Cella D, Hernandez L, Bonomi AE, et al. Spanish language translation and initial validation of the functional assessment of cancer therapy quality-of-life instrument. *Med Care* 1998;36:1407-18.
 15. Lent L, Hahn E, Eremenco S, et al. Using cross-cultural input to adapt the Functional Assessment of Chronic Illness Therapy (FACIT) scales. *Acta Oncol* 1999;38:695-702.
 16. Hahn EA, Holzner B, Kemmler G, et al. Cross-cultural evaluation of health status using item response theory: FACT-B comparisons between Austrian and U.S. patients with breast cancer. *Eval Health Prof* 2005;28:233-59.
 17. Belmonte Martínez R, Garin Boronat O, Segura Badía M, et al. Functional Assessment of Cancer Therapy Questionnaire for Breast Cancer (FACT-B+4). Spanish version validation. *Med Clin (Barc)* 2011;137:685-8.
 18. Ng R, Lee CF, Wong NS, et al. Measurement properties of the English and Chinese versions of the Functional Assessment of Cancer Therapy-Breast (FACT-B) in Asian breast cancer patients. *Breast Cancer Res Treat* 2012;131:619-25.
 19. Wan C, Zhang D, Yang Z, et al. Validation of the simplified Chinese version of the FACT-B for measuring quality of life for patients with breast cancer. *Breast Cancer Res Treat* 2007;106:413-8.
 20. Yoo HJ, Ahn SH, Eremenco S, et al. Korean translation and validation of the functional assessment of cancer therapy-breast (FACT-B) scale version 4. *Qual Life Res* 2005;14:1627-32.
 21. Pandey M, Thomas BC, Ramdas K, et al. Quality of life in breast cancer patients: validation of a FACT-B Malayalam version. *Qual Life Res* 2002;11:87-90.
 22. Cella D, Paul D, Yount S, et al. What are the most important symptom targets when treating advanced cancer? A survey of providers in the National Comprehensive Cancer Network (NCCN). *Cancer Invest* 2003;21:526-35.
 23. Yost KJ, Yount SE, Eton DT, et al. Validation of the Functional Assessment of Cancer Therapy-Breast Symptom Index (FBSI). *Breast Cancer Res Treat* 2005;90:295-8.
 24. Garcia SF, Rosenbloom SK, Beaumont JL, et al. Priority symptoms in advanced breast cancer: development and initial validation of the National Comprehensive Cancer Network-Functional Assessment of Cancer Therapy-Breast Cancer Symptom Index (NFBSI-16). *Value Health* 2012;15:183-90.

Cite this article as: Garcia SF, Hahn EA. Measure validation is an ongoing process: the Functional Assessment of Cancer Therapy-Breast Symptom Index as a case example. *Ann Palliat Med* 2012;1(3):207-210. DOI: 10.3978/j.issn.2224-5820.2012.10.07