

## Psychometrics: Essential for Valid, Reliable, and Responsive Measurement in the Development of Patient-reported Outcome Instruments in Plastic Surgery

**Vivek Panchapakesan, MD, FRCSC**

Division of Plastic Surgery  
William Osler Health System  
Toronto, ON  
Canada

**Toni Zhong, MD, MHS, FRCSC**

University Health Network  
University of Toronto  
Toronto, ON  
Canada

**Christopher R. Forrest, MD, MSc, FRCSC, FACS**

Hospital for Sick Children  
University of Toronto  
Toronto, ON  
Canada

**Andrea L. Pusic, MD, MHS, FRCSC**

Memorial Sloan-Kettering Cancer Center  
New York, N.Y.

**Achilles Thoma, MD, MSc, FRCSC**

McMaster University  
Hamilton, ON  
Canada

Sir:

**W**e would like to take this opportunity to respond to Dr. Eric Swanson's<sup>1</sup> article on the role of psychometrics in plastic surgery. As a collective, our philosophy is different from what has been described by Dr. Swanson, and we would like to illustrate some fundamental concerns we have with his article.

1. Throughout his article, he continually refers to "rigorous methodology," often citing his own work.<sup>2-5</sup> Dr. Swanson correctly discusses techniques such as randomization, inclusion of consecutive patients, and use of prospective design as "rigorous" methodology. However, these techniques pertain to studies of intervention and the *utilization* of a measure in an outcome study; they do not pertain to the actual *develop-*

*ment* of an outcome measure, and specifically, patient-reported outcome (PRO) instruments. An entirely different set of methodological safeguards exists for this and is detailed elsewhere.<sup>6-8</sup> A Food and Drug Administration scientific advisory committee<sup>6</sup> and an international consensus group of quality of life research experts<sup>7,8</sup> have set the standards for this methodology, which is rigorous in its own right. One would not use this methodology for evaluating or conducting a clinical trial; similarly, one cannot use clinical trial methodology and apply it in evaluating the *development* of a PRO instrument.

2. With regard to validity, Dr. Swanson is right in pointing out that validity is to be determined by users of the measure; however, it is erroneous to call validity a "self-serving designation that adds nothing of value to the title." Consider the introduction of a new surgical technique. For the technique to be accepted by the surgical community, it must first be adopted, attempted, and the outcomes assessed by independent investigators to determine if indeed the technique is useful and applicable. However, for the technique to even be considered in the first place, the inventor must demonstrate some success of the technique with an initial series of patients. This analogy applies to the development of a measurement instrument as well. Some degree of initial validity, reliability, and responsiveness must be demonstrated for the measure to be adopted and further tested by other researchers. In presenting a new instrument in an article, therefore, an investigator is obliged to report this. In fact, for use of a measurement instrument in outcome studies, evidence of validity and reliability must first be demonstrated and previously published<sup>6,9,10</sup>; the use of unpublished instruments is notoriously fraught with bias.<sup>11</sup>

Copyright © 2014 The Authors. Published by Lippincott Williams & Wilkins on behalf of The American Society of Plastic Surgeons. PRS Global Open is a publication of the American Society of Plastic Surgeons. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 License, where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially.

Plast Reconstr Surg Glob Open 2014;2:e280; doi:10.1097/GOX.000000000000197; Published online 23 December 2014.

**Disclosure:** The BREAST-Q and FACE-Q is owned, in part, by Memorial Sloan-Kettering Cancer Center (MSKCC). Based on MSKCC inventor-sharing policy, Dr. Pusic receives a portion of licensing revenues when the BREAST-Q or FACE-Q is used in industry-sponsored clinical trials. Neither of the other authors has any financial disclosures. The Article Processing Charge was paid for by the authors.

3. Whether we call an outcome instrument a survey, a test, a questionnaire, and so on is really just semantics. What we are really looking at is *measurement*. A ruler must be constructed in a standardized manner such that the points and numbers along the ruler actually measure what they intend to measure (validity) and does so consistently when used in the same or similar circumstances (reliability). One would have no use for a ruler where the distance between 1 and 2 cm was not the same as between 2 and 3 cm as this would undermine both validity and reliability. Similarly, when measuring abstract concepts such as physical, mental, and social well-being, what patients are able to do and how they feel, the measuring device itself is crucial, and how that device was developed underlies its validity and reliability. This leads to a discussion of psychometrics.

Psychometrics evolved from techniques of psychological measurement, and was adapted to health science research, in particular for the development of PRO instruments. PRO instruments are developed primarily from patient input through intensive qualitative methods. Psychometrics is used to evaluate the questions or items that emerge from this patient-centered input for construction of the measurement instrument. It is an integral part of the methodology for measurement instrument development adopted by national and international agencies no less than the National Institutes of Health,<sup>12</sup> the Food and Drug Administration,<sup>6</sup> and the International Society for Quality of Life Research<sup>13</sup> to name a few. Although the technical details (not “jargon”) of psychometrics may seem complex and foreign, reporting of these details is essential. Ten to fifteen years ago, the concepts, terminology, and statistical tests pertaining to randomized controlled trials (RCTs) and systematic reviews were foreign to most plastic surgeons. Even now, one may not understand completely the concepts of block permutations, the multiple forms of bias, statistical heterogeneity, or fixed versus random effects yet can accept that the end results of applying this rigorous methodology are studies that are the gold standards in the well-known hierarchy of evidence.<sup>14–16</sup>

Presentation of the methodological details in publications is essential to allow for any reader knowledgeable in the methodology to critically appraise the quality of a study. In fact, there are now widely accepted reporting guidelines for RCTs that require the presentation of numerous technical details of an RCT.<sup>9</sup> Similarly, presentation of the technical details of psychometric tests is essential to allow for critical appraisal and acceptance of

a measurement instrument.<sup>7,10</sup> Furthermore, we would argue that the readership be given more credit than to assume a discussion of psychometrics is incomprehensible to them.

To ignore the importance of psychometrics in measurement development in favor of other, standard, statistical tests for the sole reason that they are more familiar and easier to understand is an oversimplification and a misguided proposal. Would one propose that an RCT be conducted attempting to control for *all* biases or only the ones the average reader would understand? Discounting an entire discipline on the basis of a completely inapplicable methodology and body of research lacks scientific integrity. A knowledgeable understanding of PRO instruments and how they are developed and utilized is required before educated, *constructive*, criticism can be provided.

Further reading for those interested in learning more about measurement and psychometrics:

- A. Cano SJ, Hobart JC. The problem with health measurement. *Patient Prefer Adherence* 2011;5:279–290.
- B. McDowell I. *Measuring Health: A Guide to Rating Scales and Questionnaires*. 3rd ed. New York, NY: Oxford University Press; 2006.
- C. Streiner D. *Health Measurement Scales: A Practical Guide to Their Development and Use*. New York, NY: Oxford University Press; 2014.

Correspondence to Dr. Panchapakesan  
 Division of Plastic Surgery  
 William Osler Health System  
 316–89 Humber College Boulevard  
 Toronto, ON, Canada M9V 4B8  
 vivek.panchapakesan@williamoslerhs.ca

## REFERENCES

1. Swanson E. Validity, reliability, and the questionable role of psychometrics in plastic surgery. *Plast Reconstr Surg Glob Open* 2014;2:e161.
2. Swanson E. Levels of evidence in cosmetic surgery: analysis and recommendations using a new CLEAR classification. *Plast Reconstr Surg Glob Open* 2013;1:e66.
3. Swanson E. Letter to Editor Re: Thoma *et al*. Randomized controlled trial comparing health-related quality of life in patients undergoing vertical scar versus inverted T-shaped reduction mammoplasty. *Plast Reconstr Surg*. 2014;133:59e–60e.
4. Swanson E. Prospective outcome study of 225 cases of breast augmentation. *Plast Reconstr Surg*. 2013;131:1158–1166.
5. Swanson E. The FACE-Q: the importance of full disclosure and sound methodology in outcome studies. *Aesthet Surg J*. 2014;34:626–627.
6. U.S. Department of Health and Human Services Food and Drug Administration. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009. Available at: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>.

7. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010;19:539–549.
8. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol.* 2010;10:22.
9. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.
10. Calvert M, Blazeby J, Altman DG, et al.; CONSORT PRO Group. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA* 2013;309:814–822.
11. Marshall M, Lockwood A, Bradley C, et al. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *Br J Psychiatry* 2000;176:249–252.
12. PROMIS®. PROMIS® Instrument Development and Psychometric Evaluation Scientific Standards. 2012. Available at: [http://www.nihpromis.org/Documents/PROMISStandards\\_Vers2.0\\_Final.pdf](http://www.nihpromis.org/Documents/PROMISStandards_Vers2.0_Final.pdf).
13. ISOQOL. User's Guide to Implementing Patient-Reported Outcomes Assessment in Clinical Practice.
14. Sullivan D, Chung KC, Eaves FF III, et al. The level of evidence pyramid: indicating levels of evidence in Plastic and Reconstructive Surgery articles. *Plast Reconstr Surg.* 2011;128:311–314.
15. Group OLoEW. The Oxford 2011 Levels of Evidence. 2011. Available at: <http://www.cebm.net/index.aspx?o=5653>.
16. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1986;89(2 Suppl):2S–3S.