Clin Orthop Relat Res (2019) 477:669-671
DOI 10.1097/CORR.0000000000000631

Clinical Orthopaedics and Related Research®
A Publication of The Association of Bone and Joint Surgeons®

Editorial

Published online: 10 January 2019

Editorial: Importance of Validating the Scores We Use to Assess Patients with Musculoskeletal Tumors

Seth S. Leopold MD

In this month's issue of *Clinical Orthopaedics and Related Research®*, we are excited once again to present selected proceedings from our partner society, the Musculoskeletal Tumor Society (MSTS). *CORR®* has proudly served as the official journal of the MSTS for more than 20 years and this affiliation continues to provide *CORR*'s readers with a substantial quantity of the very best musculoskeletal oncology research published anywhere.

This exposure also has helped our regular readers to become expert assessors of this kind of challenging work. The diagnoses covered in these articles run from the uncommon to the rare, and there is tremendous heterogeneity in their presentations; all chondrosarcomas are not alike. The operations performed for these conditions vary widely even for the same diagnosis, depending on the tumor's location. And yet it is essential that we overcome these challenges; most malignancies are life threatening, and even some benign, non-life-threatening tumors have serious functional consequences.

Unlike other orthopaedic subspecialties, the primary goal of orthopaedic oncology generally is to save the patient's life, and, where possible, the involved limb. Many reconstructions in patients with tumors are not expected to restore normal function. This, along with the combination of uncontrollable variables I mentioned earlier, makes randomized, blinded trials on the most-important questions in tumor surgery all-but impossible.

Because of that, we need to get all we can out of studies that reside lower in the levels-of-evidence hierarchy.

Where studies of treatments are concerned, this principally means mitigating selection bias, transfer bias (followup that is insufficiently long or complete), and assessment bias. Unfortunately, since randomization is out of the question for most diagnoses, selection bias can be difficult to eliminate. Other, newer approaches—such as propensity matching [9, 10]—can help us generate causal inferences in this setting but cannot completely overcome the problem. Illness, infirmity, and death being common among patients with high-grade sarcomas or metastatic carcinomas results in transfer bias being a hallmark of many oncology trials.

The remaining opportunity for improvement is in assessment bias.

When evaluating patients' results after surgery, robust patient-reported outcomes instruments are best. Where those are unavailable, validated outcomes instruments are useful. But not all readers understand what "validated" means in this context. As a bare minimum, an outcomes instrument worthy of our attention should have the following attributes:

Construct validity

The outcomes instrument should assess the underlying element of health, disease, pain, or function that it's supposed to measure. This often is assessed by comparing a new instrument to older instruments believed to do this well, or to surveys that have good face validity (whether defined as intuitive appeal, or elements chosen by experts or consensus panels).

Reliability

It should be stable; that is, the score should not change if the same patient fills it out more than once over a span

The author certifies that neither he, nor any members of his immediate family, have any commercial associations (such as consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article.
The opinions expressed are those of the writer, and do not reflect the opinion or policy of *CORR®* or The Association of Bone and Joint Surgeons®.

S. S. Leopold, Editor-In-Chief, *Clinical Orthopaedics and Related Research®*, Philadelphia, PA, USA

S. S. Leopold MD (✉), *Clinical Orthopaedics and Related Research®*, 1600 Spruce Street, Philadelphia, PA 19013 USA, Email: sleopold@clinorthop.org

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research®* editors and board members are on file with the publication and can be viewed on request.

of weeks or months unless the patient's health has changed over that time. And if a patient's health does change, it should respond accordingly.

Robust psychometric properties

It shouldn't have severe floor or ceiling effects. In other words, the scores should not cluster towards the top or bottom of the allowable range of scores used by the measurement instrument, since that kind of clustering results in losing the ability to identify potentially important differences among patients with the health conditions in question. It helps also to know at what point after surgery one might reasonably expect the score to stabilize, representing maximal recovery.

A well-defined minimum clinically important difference (MCID)

We need to know how much of a difference—how many points on the new outcomes instrument—is enough for a patient to care about. Surgery is painful, expensive, and it carries serious risks; we would not want to recommend surgery merely because some study found a statistical difference using the outcomes instrument. We need to know, for instance, that the difference in score between two procedures being compared was enough for a patient to notice and believe was important [8].

It is important to remember that the above parameters are not absolutes, and that the validity, reliability, and psychometric properties of an outcomes instrument may vary—sometimes widely—across cultures. Validation of outcomes instruments therefore must be done not just in the language in which the instrument was created, but in any languages and cultural settings where researchers and clinicians wish to use the instrument.

The two most-commonly used outcomes instruments in the work we see are the MSTS Score [3] and the Toronto Extremity Salvage Score (TESS) [2]. The TESS is a patient-reported instrument, and the MSTS requires that the patient be assessed by a professional, so it cannot be considered a patient-reported outcomes instrument. Requiring interaction with a healthcare professional to complete an outcomes score introduces assessment bias, and generally is considered a shortcoming.

The TESS was validated in English by its designers [2] but (to my knowledge) not since then; its test-retest reliability in that initial study was generally good to excellent, and it seemed responsive to change, though I note that its construct validity was "validated" against the MSTS score, which had not itself been validated at the time the study validating the TESS was done [2]; since then, the validity of the MSTS has been called into question [7, 13].

The MSTS score since has been validated in a few small, limited studies. One evaluated its construct validity against the DASH in a small number of patients with upper-extremity tumors with mixed-to-good results, although the size and scope of this study mitigates against its wide generalizability [14]. It also was evaluated against other quality-of-life instruments (including the SF-36) in small a population of patients in Korea, but that study did not perform test-retest reliability or assess the test's other properties (such as floor or ceiling effects, or responsiveness to treatment) [7]. That study raised questions about its breadth and suggested that important domains pertaining to patients' lives and their health were not well assessed by the MSTS score. There is limited evidence comparing scores to objective functional assessments; one study found generally good correlations between MSTS scores and gait-analysis parameters after distal femoral resections for malignant tumors that underwent reconstruction [5].

More-recent studies have found that the MSTS score did not assess health-related quality of life to an adequate degree [13] and suggested that it may be compromised by ceiling effects [4]. Another concern is that the MSTS score has been validated in only a small number of languages other than English, though these studies have found at least that test-retest reliability is robust in versions of the MSTS score that were tested in Japanese and Brazilian Portuguese [4, 12, 13]. The TESS generally has fared better in studies in which it is translated and culturally adapted [1, 6, 11], although these, too, are limited by the fact that they were in part anchored to the MSTS for assessment of validity.

Further concerns about the MSTS score include the facts that all of its domains include one or more intermediate grades that are not well defined, there are five physician-generated domains and one patient-generated domain that may or may not be appropriate to pool (though they typically are summed to generate the total score), and scores from a system that includes a categorical approach to grading patients' results often are presented as a percentage of the total possible score, which may or may not be statistically appropriate.

All told, we do not know nearly enough about the behavior of the scoring systems we use to assess our treatments for patients who have musculoskeletal tumors. The heterogeneity of patients and treatments that I mentioned earlier poses real limitations in our ability to refine these instruments; the same tumor may be treated with different kinds of resection, occur in different locations, and affect patients whose overall health varies widely. Even so, we can and should do better.

Editorial

For example, as far as I know, we have no information at all about the MCIDs for either the MSTS score or the TESS, which means we don't know how large an improvement on either scale is enough for a patient to notice or care about. While important in general [8], establishing MCIDs seems essential in patients with malignancies in whom our treatment choices include the very toxic (chemotherapy, radiation) and the highly morbid (major limb-reconstructive surgery, amputation).

While randomizing patients with cancer can be challenging, and loss to followup in clinical trials of these patients is inevitable, learning more about the instruments we use to assess orthopaedic-oncology interventions is entirely within our grasp.

I challenge our partners in the Musculoskeletal Tumor Society—as well as those in the International Society of Limb Salvage, which also publishes its proceedings in *CORR*—along with any other interested clinician scientists who are not members of either of those societies, to refine the outcomes instruments we use to assess patients with musculoskeletal tumors, validate them more convincingly, and determine MCIDs at least for the common diagnoses that musculoskeletal oncology surgeons care for. By doing so, they would fill an essential gap in our knowledge, and because of this, I would commit to fast-tracking the review process on any study about these important topics submitted to *CORR*.

Acknowledgments I am grateful to Mark C. Gebhardt MD, who serves as a Senior Editor at *CORR* and who provided many thoughtful suggestions that improved this essay. Likewise, I appreciate the insightful feedback and perspective from John H. Healey MD, a Deputy Editor at *CORR* and Liaison from *CORR* to the MSTS and the International Society of Limb Salvage. Finally, I thank Raphaël Porcher PhD, *CORR*'s Senior Editor and senior methodologist/statistician, for his keen insights on this complex topic.

References

1. Akiyama T, Uehara K, Ogura K, Shinoda Y, Iwata S, Saita K, Tanzawa Y, Nakatani F, Yonemoto T, Kawano H, Davis AM, Kawai A. Cross-cultural adaptation and validation of the Japanese version of the Toronto Extremity Salvage Score (TESS) for patients with malignant musculoskeletal tumors in the upper extremities. *J Orthop Sci*. 2017;22:127-132.
2. Davis AM, Wright JG, Williams JI, Bombardier C, Griffin A, Bell RS. Development of a measure of physical function for patients with bone and soft tissue sarcoma. *Qual Life Res*. 1996;5:508-516.
3. Enneking WF, Dunham W, Gebhardt MC, Malawar M, Pritchard DJ. A system for the functional evaluation of reconstructive procedures after surgical treatment of tumors of the musculoskeletal system. *Clin Orthop Relat Res*. 1993;286:241-246.
4. Iwata S, Uehara K, Ogura K, Akiyama T, Shinoda Y, Yonemoto T, Kawai A. Reliability and validity of a Japanese-language and culturally adapted version of the Musculoskeletal Tumor Society scoring system for the lower extremity. *Clin Orthop Relat Res*. 2016;474:2044-2052.
5. Kawai A, Backus SI, Otis JC, Healey JH. Interrelationships of clinical outcome, length of resection, and energy cost of walking after prosthetic knee replacement following resection of a malignant tumor of the distal aspect of the femur. *J Bone Joint Surg Am*. 1998;80-A:822-831.
6. Kim HS, Yun J, Kang S, Han I. Cross-cultural adaptation and validation of the Korean Toronto Extremity Salvage Score for extremity sarcoma. *J Surg Oncol*. 2015;112:93-97.
7. Lee SH, Kim DJ, Oh JH, Han HS, Yoo KH, Kim HS. Validation of a functional evaluation system in patients with musculoskeletal tumors. *Clin Orthop Relat Res*. 2003;411:217-226.
8. Leopold SS, Porcher R. Editorial: The minimum clinically important difference—The least we can do. *Clin Orthop Relat Res*. 2017:475:929-932.
9. Lonjon G, Boutron I, Trinquart L, Ahmad N, Aim F, Nizard R, Ravaud P. Comparison of treatment effect estimates from prospective nonrandomized studies with propensity score analysis and randomized controlled trials of surgical procedures. *Ann Surg*. 2014;259:18-25.
10. Lonjon G, Porcher R, Ergina P, Fouet M, Boutron I. Potential pitfalls of reporting and bias in observational studies with propensity score analysis assessing a surgical procedure: A methodological systematic review. *Ann Surg*. 2017;265:901-909.
11. Ogura K, Uehara K, Akiyama T, Iwata S, Shinoda Y, Kobayashi E, Saita K, Yonemoto T, Kawano H, Chuman H, Davis AM, Kawai A. Cross-cultural adaptation and validation of the Japanese version of the Toronto Extremity Salvage Score (TESS) for patients with malignant musculoskeletal tumors in the lower extremities. *J Orthop Sci*. 2015;20:1098-1105.
12. Rebolledo DC, Vissoci JR, Pietrobon R, de Camargo OP, Baptista AM. Validation of the Brazilian version of the Musculoskeletal Tumor Society rating scale for lower extremity bone sarcoma. *Clin Orthop Relat Res*. 2013;471:4020-4026.
13. Uehara K, Ogura K, Akiyama T, Shinoda Y, Iwata S, Kobayashi E, Tanzawa Y, Yonemoto T, Kawano H, Kawai A. Reliability and validity of the Musculoskeletal Tumor Society scoring system for the upper extremity in Japanese patients. *Clin Orthop Relat Res*. 2017;475:2253-2259.
14. Wada T, Kawai A, Ihara K, Sasaki M, Sonoda T, Imaeda T, Yamashita T. Construct validity of the Enneking score for measuring function in patients with malignant or aggressive benign tumors of the upper limb. *J Bone Joint Surg Br*. 2007;89:659-663.