



Melissa R. Dunbar and Zoher Ghogawala

The importance of the patient perspective in US healthcare is broadly recognized. There is a growing realization that patient experience, satisfaction, and outcome are essential to improving overall quality in spine care. While process measures and 30-day complications data have helped focus attention on improving quality in medicine, there are unique aspects to spinal care that demand the monitoring of long-term patient-reported outcomes.

The most important aspect of spinal care is restoration of function and productivity and reduction of pain. Patients are the best source to provide that feedback. There is a need for practical validated patient-reported outcomes tools in our quest to optimize outcomes within spine care.

Characteristics of Patient-Reported Outcomes Tools

Patient-reported outcome (PRO) measures should have three characteristics. They should be reliable, valid, and responsive [1–3]. Reliability refers to reproducibility. There is interobserver (degree to which different observers obtain similar results) reliability and intraobserver (degree to which the same observer gets the same result on repeated testing) reliability. There is also the concept of test-retest reliability, which examines how well an instrument performs when tested between two separate time points.

M. R. Dunbar

Department of Neurosurgery, Lahey Hospital & Medical Center, Burlington, MA, USA

Z. Ghogawala (✉)

Department of Neurosurgery, Lahey Hospital & Medical Center, Burlington, MA, USA

Department of Neurosurgery, Tufts University School of Medicine, Boston, MA, USA

e-mail: Zoher.Ghogawala@Lahey.org

Typically, the reliability of a PRO is assessed with a kappa (κ) statistic, which measures agreement between observers. An outcomes tool is considered to be highly reliable if $\kappa > 0.8$, is moderately reliable if κ statistic falls between 0.6 and 0.8, and is not reliable if κ is below 0.6 [4]. There is also assessment of how an individual test domain performs in relation to the overall composite result [5]. This is measured using the Cronbach α , which assesses whether individual domains of an outcomes tool correlate with the final composite score [6, 7].

Validity represents to what extent the measure actually assesses the measure of interest. For example, the Oswestry Disability Index (ODI) is widely used because it has been validated as being an accurate measure of disability from spinal disorders [8]. Typically, newer measures are judged to be accurate and valid when they correlate with previously validated and widely used instruments like the ODI.

Similarly, responsiveness is assessed by correlating performance to a previously validated tool such as ODI or SF-36 [9]. Recently, the Scoliosis Research Society-22 (SRS-22) instrument has been validated to be more responsive than ODI for detecting improvement in patients treated surgically for adult spinal deformity [10, 11].

Disease-Specific Versus General Health-Related Quality of Life

A PRO may focus upon assessment of a specific disease's impact on a patient or may focus upon general health-related quality of life. Instruments like ODI have been validated and used for the assessment of treatments for low back pain. Instruments like the SF-36 or SF-12 survey examine overall health-related quality of life. The distinction is vital in understanding applicability of PROs to specific clinical and research questions.

For example, two recently published RCTs on the benefit of lumbar fusion in patients with lumbar spondylolisthesis reached different conclusions [12]. While there are multiple differences between the trials in terms of study population and design, one used ODI as the primary outcome measure and concluded that there was no advantage to adding a fusion to a decompression when treating patients with lumbar spondylolisthesis [13], while the other trial used SF-36 as the primary outcome measure and reached the opposite conclusion [14]. The SF-36, and more recently the SF-12, have been demonstrated as being reliable, valid, and responsive for assessing patients with both cervical and lumbar spinal disorders [15].

Minimum Clinical Important Difference

When assessing a response to treatment, it is important to know whether the change in a validated PRO is clinically meaningful or not. Copay and colleagues have rigorously examined a large group of patients with spinal disorders to calculate the minimal clinically important difference (MCID) for commonly used PROs. They reported MCID for ODI as 12.8 points and MCID for SF-36 physical component summary (PCS) at 4.9 points [16]. Researchers often use MCID values to calculate

sample size for clinical trials by asking how many patients would be required to detect a meaningful clinical difference between two study populations when comparing the effectiveness to two or more treatments [17].

Cost-Utility Analysis

Cost-utility analysis is a specific type of cost-effectiveness evaluation that permits the direct comparison of cost and outcome for different treatments. The outcome for these analyses requires the assessment of preference-based health-related quality of life such as the EQ-5D (EuroQol group) [18]. Preference-based health-related quality of life is reported as a score between 0 (death) and 1 (perfect health). The score is then multiplied by time (years) to generate quality-adjusted life years (QALYs) so that cost can be compared in terms of cost/QALYs gained by one treatment versus another. The validated PRO, SF-36, can be converted to QALYs by transforming 11 questions from SF-36 into the preference-based SF-6D [19].

PROMIS

One of the challenges with the collection of PROs in registry quality efforts is the burden placed on patients. Completion of validated PROs is often time-consuming, and many registries have suboptimal completion rates over time. In addition, many validated PROs require licenses and considerable expense for healthcare researchers to use on a routine basis.

The NIH-funded Patient-Reported Outcomes Measurement Information System (PROMIS[®]) has been developed and validated recently by multiple groups for both cervical and lumbar spinal disorders [20, 21]. There is no fee associated with or license required for the use of PROMIS[®] for noncommercial use. PROMIS[®] assesses a breadth of domains and allows for a great range of responses which enables the measures to be responsive both for those in the healthy, general population as well as those suffering from chronic conditions.

PROMIS[®] employs computer-adaptive testing (CAT) technology that reduces the number of questions that each patient must complete based on answers to previous questions. One recent study found that PROMIS[®]-physical function took 1.1 min for patients to complete as opposed to 4 min for completing SF-12 [20].

Anxiety/Depression

Multiple groups have found that PROs are affected by anxiety and depression. These two common mental disorders complicate the ability to compare outcomes of different spinal treatments. In the past, it has been difficult to account for anxiety and depression other than to acknowledge their overall negative impact upon PRO results. With the development of PROMIS[®] anxiety and depression domains, it has

been possible to learn more about the impact of these conditions in evaluating treatment for spinal disorders. For example, a recent study found that PROMIS® depression scores >50 were associated with worse PROMIS® physical function and ODI disability scores before and after lumbar decompression surgery; however, depressed patients had greater improvement in PROMIS® physical function after surgery despite having lower absolute scores postoperatively compared with non-depressed patients [22].

Summary

The importance of measuring PROs for assessing quality efforts in spine care has been established. A given PROs must be reliable, valid, and responsive. In addition, PROs are useful for cost-utility analyses. National and international registry efforts have underscored the need for PROs to be free of cost and simple to complete for patients. Recent developments in CAT have increased the feasibility of including PROs in everyday spinal practice.

References

1. Deyo RA, Andersson G, Bombardier C, et al. Outcome measures for studying patients with low back pain. *Spine*. 1994;19:2032S–6S.
2. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials*. 1991;12:142S–58S.
3. Kopec JA, Esdaile JM. Functional disability scales for back pain. *Spine*. 1995;20:1943–9.
4. Hadley MN, Walters BC, Grabb PA, et al. Guidelines for the management of acute cervical spine and spinal cord injuries. *Clin Neurosurg*. 2002;49:407–98.
5. Karanicolas PJ, Bhandari M, Kreder H, et al. Evaluating agreement: conducting a reliability study. *J Bone Joint Surg Am*. 2009;91(Suppl 3):99–106.
6. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull*. 1955;52:281–302.
7. Cronbach LJ. Test reliability; its meaning and determination. *Psychometrika*. 1947;12:1–16.
8. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy*. 1980;66:271–3.
9. Guilfoyle MR, Seeley H, Laing RJ. The short form 36 health survey in spine disease – validation against condition-specific measures. *Br J Neurosurg*. 2009;23:401–5.
10. Ghogawala Z, Resnick DK, Watters WC 3rd, et al. Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 2: assessment of functional outcome following lumbar fusion. *J Neurosurg Spine*. 2014;21:7–13.
11. Bridwell KH, Berven S, Glassman S, et al. Is the SRS-22 instrument responsive to change in adult scoliosis patients having primary spinal deformity surgery? *Spine*. 2007;32:2220–5.
12. Ghogawala Z, Resnick DK, Glassman SD, Dziura J, Shaffrey CI, Mummaneni PV. Randomized controlled trials for degenerative lumbar spondylolisthesis: which patients benefit from lumbar fusion? *J Neurosurg Spine*. 2017;26:260–6.
13. Forsth P, Olafsson G, Carlsson T, et al. A randomized, controlled trial of fusion surgery for lumbar spinal stenosis. *N Engl J Med*. 2016;374:1413–23.
14. Ghogawala Z, Dziura J, Butler WE, et al. Laminectomy plus fusion versus laminectomy alone for lumbar spondylolisthesis. *N Engl J Med*. 2016;374:1424–34.

15. Lee CE, Browell LM, Jones DL. Measuring health in patients with cervical and lumbosacral spinal disorders: is the 12-item short-form health survey a valid alternative for the 36-item short-form health survey? *Arch Phys Med Rehabil*. 2008;89:829–33.
16. Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry disability index, medical outcomes study questionnaire short form 36, and pain scales. *Spine J*. 2008;8:968–74.
17. Ghogawala Z, Resnick DK, Glassman SD, Dziura J, Shaffrey CI, Mummaneni PV. Achieving optimal outcome for degenerative lumbar spondylolisthesis: randomized controlled trial results. *Neurosurgery*. 2017;64:40–4.
18. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol group. *Ann Med*. 2001;33:337–43.
19. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ*. 2002;21:271–92.
20. Boody BS, Bhatt S, Mazmudar AS, Hsu WK, Rothrock NE, Patel AA. Validation of Patient-Reported Outcomes Measurement Information System (PROMIS) computerized adaptive tests in cervical spine surgery. *J Neurosurg Spine*. 2018;28:268–79.
21. Sharma M, Ugiliweneza B, Beswick J, Boakye M. Concurrent validity and comparative responsiveness of PROMIS – SF versus legacy measures in the cervical and lumbar spine population: longitudinal analysis from baseline to post surgery. *World Neurosurg*. 2018;115:e664.
22. Merrill RK, Zebala LP, Peters C, Qureshi SA, McAnany SJ. Impact of depression on patient-reported outcome measures after lumbar spine decompression. *Spine*. 2018;43:434–9.