

Fundamentals of Clinical Outcomes Assessment for Spinal Disorders: Clinical Outcome Instruments and Applications

Patrick Vavken^{1,2,3} Anne Kathleen B. Ganal-Antonio⁴ Julia Quidde³ Francis H. Shen⁵
Jens R. Chapman⁶ Dino Samartzis^{7,8}

¹ Department of Orthopaedic Surgery, Boston Children's Hospital, Harvard University Medical School, Boston, Massachusetts, United States

² Harvard Center for Population and Development Studies, Harvard School of Public Health, Boston, Massachusetts, United States

³ Department of Orthopedic Surgery, University of Basel, Basel, Switzerland

⁴ Department of Orthopedics, Makati Medical Center, Makati City, Philippines

⁵ Department of Orthopaedic Surgery, University of Virginia, Charlottesville, Virginia, United States

⁶ Swedish Neuroscience Institute, Swedish Medical Center, Seattle, Washington, United States

⁷ Department of Orthopaedics and Traumatology, The University of Hong Kong, Pokfulam, Hong Kong, SAR, China

⁸ The Laboratory and Clinical Research Institute for Pain, The University of Hong Kong, Pokfulam, Hong Kong, SAR, China

Address for correspondence Dino Samartzis, DSc, Department of Orthopaedics and Traumatology, The University of Hong Kong, Professorial Block, 5th Floor, 102 Pokfulam Road, Pokfulam, Hong Kong, SAR, China (e-mail: dsamartzis@msn.com).

Global Spine J 2015;5:329–338.

Abstract

Study Design A broad narrative review.

Objectives Outcome assessment in spinal disorders is imperative to help monitor the safety and efficacy of the treatment in an effort to change the clinical practice and improve patient outcomes. The following article, part two of a two-part series, discusses the various outcome tools and instruments utilized to address spinal disorders and their management.

Methods A thorough review of the peer-reviewed literature was performed, irrespective of language, addressing outcome research, instruments and tools, and applications.

Results Numerous articles addressing the development and implementation of health-related quality-of-life, neck and low back pain, overall pain, spinal deformity, and other condition-specific outcome instruments have been reported. Their applications in the context of the clinical trial studies, the economic analyses, and overall evidence-based orthopedics have been noted. Additional issues regarding the problems and potential sources of bias utilizing outcomes scales and the concept of minimally clinically important difference were discussed.

Conclusion Continuing research needs to assess the outcome instruments and tools used in the clinical outcome assessment for spinal disorders. Understanding the fundamental principles in spinal outcome assessment may also advance the field of “personalized spine care.”

Keywords

- ▶ spine
- ▶ outcomes
- ▶ instruments
- ▶ questionnaires
- ▶ personalized

received
February 7, 2014
accepted after revision
September 29, 2014
published online
March 13, 2015

DOI <http://dx.doi.org/10.1055/s-0034-1396046>.
ISSN 2192-5682.

© 2015 Georg Thieme Verlag KG
Stuttgart · New York

License terms



Introduction

Throughout recent years, clinical outcomes research in orthopedic surgery has grown immensely, whereby the spine discipline and specialists have gained a leadership role in this development. Generally, there are six categories of outcome measurements, some of which are more readily employed in spine surgery and research than others (► **Table 1**). However, an abundance of outcome measurement tools have been designed to address the spine-related disorders, including general functional status instruments, pain measurement tools, back-specific disability scales, and neck and pain disability scales. These tools have been tailored to not only address localized pathologic manifestations, but to also include the assessment of patient-specific factors that may affect health status and treatment outcomes. Such factors may entail patients' educational level, employment history and satisfaction, psychological elements, expectations, satisfaction levels, worker's compensation, and influences by third-party claims. Furthermore, the strength of an outcome measurement tool is derived from its specificity for a given condition or outcome, reproducibility, construct validity, responsiveness, and interpretability.¹ The following review provides several examples of the common outcome instruments used in spine research, focusing on their clinical applications and implications of outcomes research, without any claim on completeness. A more expanded list of various outcome instruments can be found in ► **Table 2**.

Health-Related Quality-of-Life Outcome Instruments

Short Form-36 Survey

The Short Form-36 (SF-36) is a 36-item questionnaire that was developed to measure general health status.²⁻⁴ It is a widely used, comprehensive, generic outcome tool that quantitatively measures physical and mental dimensions. The SF-36 has been extensively investigated to confirm its validity and reliability, and it has been translated into more than 40 languages as part of the International Quality of Life Assessment Initiative.⁵ The questionnaire takes ~5 to 10

minutes to complete and is based on eight domains: physical functioning, role limitations caused by physical health or by emotional problems, bodily pain, social functioning, general mental health, vitality, and general health perceptions. In essence, the SF-36 form is a reliable and valid tool that is commonly used because of its brevity, psychometric assessment potentials, and applicability to patients with different medical conditions and demographics.⁶ Moreover, the SF-36 is a self-administered questionnaire that is sensitive to differences in disease severity, and it distinguishes between sick and healthy populations. In addition, to further improve the efficacy associated with the SF-36 form and to decrease costs, a shorter version of the questionnaire was constructed in the mid-1990s, aptly named the SF-12.⁴ This shorter questionnaire employs the same eight-scale profile as the SF-36, but in less depth than the original questionnaire.

The Sickness Impact Profile

The Sickness Impact Profile (SIP) was first published in the mid-1970s and was further revised in the 1980s.⁷ This outcome tool was also constructed to evaluate the functional health consequences of health care. It is a behavior-based measurement tool that presents a set of items in a yes-or-no, dichotomous fashion. The SIP evaluates the type of work that chronically ill patients can perform, as well as how these individuals will respond in their work environments because of their medical conditions. A total of 136 items are included in the assessment tool, which takes ~20 to 30 minutes to complete. Although the SIP attempts to offer a descriptive profile of changes in a patient's behavior caused by his or her illness, it fails to capture the potential dynamics between an individual's health and his or her work. The weakened sensitivity of such a tool is possibly attributable to the inadequate depiction of a general set of work activities, as well as the potential categorical limitations inherent in a yes-or-no question format. However, the SIP and other measurement tools have become the foundation, as well as the impetus, for more precise role-specific measurements that relate to the working life. Overall, it is a reliable and valid instrument that has been used to investigate various

Table 1 Types of outcome measures

Type of measure	Description
Dimension-specific	Focus is on a particular aspect of health (i.e., Beck's Depression Inventory)
Disease-/population-specific	Measures several health domains and focuses on aspects of health that are relevant to particular health problems
Generic	Measures outcomes across diseases and different patient populations
Individualized	The importance of certain aspects of the respondent's life are measured and weighted to produce a single score (i.e., patient-generated index scores)
Role-specific	A more specific generic tool that captures aspects of working life (i.e., Occupational Role Questionnaire)
Utility	Developed for economic evaluation, entails preferences for health states, and yields a single index (i.e., EuroQol EQ-5D)

Table 2 Various condition and outcome measurement tools (list is not comprehensive)

Categories	Measurements
Pain scales	<ul style="list-style-type: none"> • Verbal rating scale • Visual analog scale • Numerical rating scale • Wisconsin Brief Pain Questionnaire • Memorial Pain Questionnaire • McGill Pain Questionnaire • Patient Outcome Questionnaire • Medical Outcomes Study • Descriptor Differential Scale • Integrated Pain Scale • Pain Perception Profile • West Haven-Yale Multidimensional Pain Inventory • Brief Pain Inventory • Unmet Analgesic Needs Questionnaire • City of Hope Mayday Pain Resource Center Pain Audit Tools • City of Hope Mayday Pain Resource Center Patient Pain Questionnaire • Dallas Pain Questionnaire • Northwick Park Neck Pain Questionnaire • Neck Pain and Disability Scale
Disability: lower back questionnaires	<ul style="list-style-type: none"> • Oswestry Disability Index Questionnaire • Million Disability Questionnaire • Roland-Morris Disability Questionnaire • Waddell Disability Index • Low Back Pain Type Specifications
Disability: cervical questionnaires	<ul style="list-style-type: none"> • Neck Disability Index • Neck Pain and Disability Scale • Headache Disability Index
Psychometric questionnaires	<ul style="list-style-type: none"> • Illness Behavior Questionnaire • Psychosocial Pain Inventory • Waddell Non-Organic Low Back Pain Signs • Modified Somatic Perception Questionnaire • Somatic Amplification Rating Scale • Modified Zung Depression Index • Minnesota Multiphasic Personality Inventory • Health Status Questionnaire • Fear Avoidance Beliefs Questionnaire
Patient satisfaction questionnaires	<ul style="list-style-type: none"> • Patient Satisfaction Questionnaire • Low Back Pain Patient Satisfaction • Group Health Association of America Consumer Satisfaction Survey • Chiropractic Satisfaction Questionnaire
Combined assessment scales	<ul style="list-style-type: none"> • Edmonton Symptom Assessment System • Symptom Distress Scale • Memorial Symptoms Assessment Scale • Symptom Scale • Voices • Rotterdam Symptom Checklist • Support Team Assessment Schedule • National Hospice Study • Care Cooperative Charts • Hospice Quality of Life Index • McGill Quality of Life Index • Quality of Well-Being Scale • European Organization for Research and Treatment of Cancer Quality of Life Questionnaire 30 • VITAS Quality of Life Index • SF-36, SF-12 • Health Status Questionnaire • RAND 36 • Sickness Impact Profile • Nottingham Health Profile • Scoliosis Follow-Up Questionnaire • Scoliosis Research Society (SRS-22, - 30)

(Continued)

Table 2 (Continued)

Categories	Measurements
	<ul style="list-style-type: none"> • Cervical Spine Outcomes Questionnaire • North American Spine Society Lumbar Spine Outcome Assessment Instrument

Adapted from Samartzis D, Dominique DA, Perez-Cruet MJ, Fehlings MG. Clinical outcome analyses. In: Perez-Cruet MJ, Khoo LT, Fessler RG, eds. *An Anatomical Approach to Minimally Invasive Spine Surgery*. St. Louis: Quality Medical Publishing, Inc.; 2006:103–130.⁶

pathologic conditions of the spine, and it is available in multiple languages.²

EQ-5D

The group of quality-of-life outcome measurements was expanded in 1990 by the EQ-5D, published by EuroQoL. This multinational group was founded in 1987 with the aim of creating a non-disease-specific measurement tool for health-related quality of life (HRQOL), which was to be simple, self-administered, and brief, so as not to be a burden for the respondent. From this, the EQ-5D was developed and constantly improved. Respondents are asked to describe their state of health for 16 items, arranged in two groups of eight 20-cm visual analog scales (VASs) anchored at “best imaginable health state” (i.e., 100) and “worst imaginable health state” (i.e., 0). Being Euro-centric, the EQ-5D has been translated into and validated for several European languages and populations. Also, based on such validation studies, weights have been calculated for individual health states to allow the interpolation of scores for health states not directly assessed by the EQ-5D.

A good example from the literature on the use of HRQOL outcomes is the 2010 study by Danielsson et al.⁸ The authors examined 77 patients with adolescent scoliosis. Thirty-seven patients were treated with a brace, and 40 patients had no immobilization. Radiographic and clinical examinations were performed during the follow-up for both groups. In addition to the usual end points, such as progression and spine-specific outcomes, quality of life was measured with the SF-36. The results showed no difference in quality of life as the SF-36 and the Scoliosis Research Society (SRS)-22 questionnaire demonstrated no significant differences between the two groups. Specifically, there were no statistical differences between the groups in the effect of bracing or on the quality of life of these young patients.

Neck-Related Outcome Instruments

Neck Disability Index

The Neck Disability Index (NDI) is a 10-item, one-dimensional questionnaire designed to assess neck pain and disability.⁹ The items are organized by type of activity and then by six different assertions corresponding to progressive levels of functional capability. Based on the response, the NDI is scored as a percentage of maximal pain and disability. The strength of the tool lies in its established validity among different patient populations, as well as against multiple measurements of function, pain, and clinical signs or symptoms.

Concerns regarding the NDI include a possible “ceiling effect”; that is, patients with severe disease may reach a plateau at a maximum score, and the scale cannot reflect any further decline in function.¹⁰ Furthermore, the scale can be affected by incompletely answered questionnaires, because it focuses on automobile driving, which is not often applicable to the elderly or societies where certain activities are not common.¹¹ These missing data fields make data interpretation and comparisons difficult.

Neck Pain and Disability Scale

The Neck Pain and Disability Scale (NPDS) is a multidimensional comprehensive tool used to measure neck pain and associated functional status.¹² The NPDS measures neck problems, pain intensity, the effect of neck pain on emotion and cognition, and the degree to which neck pain interferes with daily activities. This instrument was developed as an extension of the NDI.⁹ It consists of 20 items, and each question has a 10-cm line that is similar to a VAS. The items are scored from 0 (normal) to 5 (worse) based on the scale, and the total score is the sum of all the items. In comparison studies, the NPDS demonstrates good reproducibility, construct validity, and factorial structure.^{9,13} The NPDS has been validated in multiple languages, and all versions have shown good psychometric properties. Although patients prefer the simplicity of its VAS, this assessment tool has been associated with some limitations in the literature, because it must be completed directly by the study subject, rather than by study personnel.⁹

In a recently published prospective study from Kang et al,¹⁴ the NDI and NPDS were used on 72 patients who underwent an anterior cervical spine surgery. The NDI and NPDS (and other questionnaires) were evaluated before and 1 year after surgery to examine the outcome of a single-level anterior cervical discectomy and fusion, among other things. The results showed an improved NDI (34.2 to 9.9) and NPDS (44.8 to 16). With the help of these questionnaires, the authors compared the function, pain, and clinical signs and symptoms specifically for the neck.

Low Back-Related Outcome Instruments

In 2011, Chapman et al¹⁵ identified 75 different outcome measures evaluating chronic low back pain, raising awareness of the plethora of outcome questionnaires and the nonstandardization of assessment between centers. However, for low back pain in general, the most commonly used outcome instruments have been the Oswestry Disability Index (ODI) and the Roland-Morris Disability Questionnaire (RMDQ).

Oswestry Disability Index

The ODI was first reported in 1980 and enjoys the distinction of being one of the earliest disease-specific spine questionnaires.¹⁶ The 10-section questionnaire assesses pain level and how it affects the activities of daily living, such as sleeping, self-care, social life, sex, and traveling. Various versions of the ODI have been reported, and it has been translated into multiple languages with respective validity.^{17,18} Version 2.0 of the ODI has been recommended by Fairbank and Pynsent,¹⁷ and it is more specific to the patients' status in line with the current day of assessment as a reference point. Some sections have been modified or omitted in certain studies due to inapplicability or other domains were measured by other means. A standard scoring method can be used in all the versions of the ODI; however, variations in the ODI should allow for scoring adjustments.¹⁷ The categorical data are converted to an ordinal number, and the sum is taken. Though this can be viewed as a continuum, there is no linear correlation with disability.¹⁷ A minimum of 15-point change in score has been recommended to represent a clinically significant change.¹⁷

Roland Morris Disability Questionnaire

The RMDQ was designed to assess physical disability due to low back pain and consists of 24 binary questions, which can result in a total score of 24 (the higher the score, the worse the outcome). This questionnaire was originally designed for use in primary care, for the elderly to monitor their care, and for research purposes.¹⁹ The RMDQ was adopted from the SIP and the phrase "because of my back pain" was added at the end of each statement.²⁰ Because it is short, simple to complete, and readily understood, it is also widely used and validated in different languages.²¹ Several modifications to the RMDQ have been proposed but have not been adapted because these modifications have not demonstrated a vast improvement of the original version.²² Studies have demonstrated that the improvement of initial scores lower than 4 points and the deterioration in patients with initial scores greater than 20 points cannot be detected with a high degree of confidence.²³ However, the minimum level of detectable change is up to 5 points.

Direct comparison between the ODI and RMDQ showed correlation between the two scales, though the ODI seems to be more sensitive in detecting change in a more severe symptoms versus minor disability.²¹ Both questionnaires use the time scale of "now" compared with the average of symptoms in the previous week, or the previous month as is the case with the SF-36. Roland and Fairbank recommend use of the ODI in persistent severe disability, whereas the RMDQ is recommended in patients with relatively little disability, because the RMDQ includes more subtle complaints (back pain, discomfort) and the ODI focuses on major problems (activities of daily living, hygiene, social function, sexual function) that might not be as relevant for the early stages of spine diseases.²¹

Spinal Stenosis Questionnaire

Stucki et al²⁴ developed a self-administered questionnaire to address disability and health-related parameters in patients

diagnosed with lumbar spinal stenosis. It includes three scales with seven questions on symptom severity, five on physical function, and six on satisfaction. Symptom severity addresses pain, pain frequency, back pain, leg pain, numbness, weakness, and balance disturbance. The physical function questions address walking distance and ability to walk for pleasure, for shopping, and for daily functions at one's residence. Patient satisfaction addressed outcomes from low back surgery, pain relief after operation, walking ability after the operation, ability to do housework or yard work or job after surgery, lower limb strength, and balance. The Spinal Stenosis Questionnaire has been validated in several languages against SF-36, ODI, and other measurement scales. The reliability has been measured in test-retest assessments and consistently scored above 90% agreement.²⁵⁻²⁸

With the help of the ODI, a Korean group examined 90 patients who underwent a direct lumbar interbody fusion with minimum follow-up of at least 6 months.²⁹ In addition to the morphological aspects and fusion rates, the clinical outcome was also measured on the basis of the ODI, which significantly improved after the surgery. As such, combined with the other results, the direct lumbar interbody fusion was proven to be an effective surgical procedure with satisfactory results.

Spine Deformity Outcome Instruments

In 1999, Haer et al³⁰ reported a disease-specific HRQOL instrument to assess the baseline condition and treatment effect in patients with spinal deformity. This instrument consisted of 24 items divided into seven equally weighted domains (SRS-24) evaluating patient satisfaction and performance of adolescents with idiopathic scoliosis. The parameters were pain, general self-image, postoperative self-image, general function, overall level of activity, postoperative function, and satisfaction. The study showed that the greatest differences were in pain and general activity level, and patient satisfaction had a high correlation with pain, followed by self-image.

Scoliosis Research Society Questionnaires (SRS-22, -23, and -30)

The SRS-23 questionnaire was the result of the changes made on the original SRS HRQL questionnaire (i.e., SRS-24).³¹ The five domains included in this modified SRS (i.e., MSRS) outcome instrument were function/activity, pain, self-image/appearance, mental health, and satisfaction of management. The original questionnaire was shortened from 24 to 23 questions, and the answers were expanded from two to five, giving more information. This MSRS showed increased internal consistency in all five domains compared with the original SRS outcome measure.³¹ When each MSRS domain was correlated with the SF-36, correlation was strong in all domains, establishing concurrent validity for the SRS Questionnaire.³¹

The MSRS was originally for adolescent idiopathic scoliosis; however, it has been established that it can also be used for adult deformity as a self-assessment tool for measuring

health status and outcomes.³² Adult scoliosis has a great effect on the quality of life of the patient with regards to pain, function, appearance, and mental health. Radiographic measurements do not correlate with the health status.³² The MSRS demonstrated a high-degree of internal consistency and reproducibility in adult scoliosis and was validated with SF-36 showing good correlation.³²

The SRS-22 is a result of modifications made on the MSRS (SRS-23), including deleting an item in the self-image/appearance domain and moving another item to the pain domain.³³ The SRS-22 has shown good to excellent internal consistency among the five domains, with excellent test-retest intraclass correlations and relatively high correlation between relevant SRS-22 and SF-36 domains.³³ There might be some bias concerning treatment satisfaction, because the operating surgeon is known to the patient and the treatment is an accepted technique that is known to have good results.³³

The use of the SRS-22 in adult deformity has also been established, comparing it to the ODI and SF-12.³⁴ Compared with the ODI assessing patients preoperatively to 2 years postoperatively, the SRS-22 showed greatest change involving the self-image, pain, and total score domains. This suggests that the SRS-22 is more sensitive to change caused by primary surgery, followed by ODI, then the SF-12. Furthermore, comparing the SRS-22 with the latter questionnaires, adult scoliosis patients seem to have significant improvement in pain, self-image, and function after surgical treatment.

Modifications were made on the SRS-23 in 2003, including adding historical recall questions and rephrasing/relocating various questions, giving rise to the SRS-30.³⁵ Baldus et al³⁵ established the population medians, means, confidence intervals, standard deviations, and percentiles for the domains of the SRS-30 questionnaire in adults without scoliosis. These values are regarded as reference points, allowing the clinicians and investigators to interpret and compare domain scores of individuals with spine deformity to those without deformity. Thus, the SRS-30 is the most current, most sensitive scoliosis outcome tool that has been validated.

Use of the SRS-30 has been demonstrated in a recently published study by Dorward et al.³⁶ One hundred twenty-eight patients with different spinal deformities underwent posterior fusion to spinal deformity surgery with posterior column osteotomies. The patients were observed at 2-year-follow-up. Besides the morphological differences, the clinical outcomes were measured using the SRS-30 and the ODI. Postoperative SRS-30 and ODI scores improved significantly and demonstrated a favorable clinical outcome.

Pain Outcome Instruments

Visual Analog Scale

The VAS consists of a straight line with the end points denoting the extreme limits, such as “no pain” and “pain as bad as it could be.”³⁷ When descriptive terms such as “mild,” “moderate,” “severe,” or a numerical scale are added to the line, a graphic rating scale is created. Patients are simply asked to show their pain level between the two end points of the line. The line is generally ~10 to 15 cm in length, because

studies have shown this length is the easiest for patient use and it results in the smallest measurement error.³⁸ The distance from the “no pain” end point represents the patient’s pain score. Studies have quantified the amount of change needed to be significant with the VAS. A change of 20% for chronic low back pain and 12% for acute low back pain has been shown to be clinically significant.³⁸ One possible disadvantage of scoring the traditional VAS is the time needed, because measurements must be made individually. To simplify the process, a mechanical VAS is available; it has a sliding tool that patients move corresponding to their pain. Overall, the VAS is a sensitive, reliable, and easy assessment tool to use for the evaluation of pain in patients with back disorders.^{38,39}

In a recent report by Lee et al,⁴⁰ the authors utilized the VAS as a measurement tool in their investigation of the quality of life and the cervical sagittal alignment in patients with ankylosing spondylitis in comparison with a healthy group. The study included 102 patients with ankylosing spondylitis and a control comparative group of 50 people. Radiographic parameters, VAS score assessing neck pain, the NDI, and the NPAD were compiled. Correlation analysis verified significant relationships between radiographic malposition and the quality of life.

The McGill Pain Questionnaire

Considered a standardization benchmark for the evaluation of pain, the McGill Pain Questionnaire (MPQ) is a reliable, valid, and sensitive tool for the assessment of pain relief and treatment.⁴¹ The MPQ has three major measurement components: a pain-rating index, several words chosen to describe pain, and a scale of 1 to 5 that represents pain intensity. The pain-rating index relies on a numeric grading of words describing the sensory, affective, and evaluative aspects of pain. Each descriptor is assigned a rank value based on its position in the word set. The sum of the rank values yields the pain-rating index. In addition, a short form of the MPQ (SF-MPQ) has been developed, with 15 questions, 11 of which address sensory dimensions and 4 are related to affective dimensions. The intensity scale in the SF-MPQ has been reduced to 4 points, and the pain rating index is incorporated as a VAS.

What to Do with Results from Outcome Assessment Studies

The findings of outcome analyses are often published as stand-alone articles in scientific journals. Moving scientific findings from basic research into clinical practice is at the heart of the translational research.

Clinical Research/Clinical Trials

Most outcome assessment is used for clinical research, aimed to publish in a scientific journal. Currently, 3 to 8% of all studies published in the orthopedic journals are controlled, meaning that most studies are still observational or cases series/case reports.^{42,43} However, there has been a move to higher-quality science recently, and spine surgery and research is no exception.⁴³

A particularly important subform of clinical research is clinical trials, which assess new interventions or medications. Clinical trials can be subdivided into three phases. In phase 1, safety and efficacy are established and whether the new treatment performs better than what is currently available without causing additional complications is investigated. A phase 1 trial requires the approval of the new treatment or medication from the Food and Drug Administration, or a similar institution if outside the United States. Phase 2 is designed to focus on evaluating efficacy or how this treatment will perform in a real-life setting. At the same time, the cost-effectiveness of the new treatment is analyzed to see if the additional cost will provide additional benefit compared with what treatments are already available. Finally, phase 3 is designed to ensure that a new treatment is a sustainable solution.

Economic Analyses

A particularly pressing issue in present-day clinical medicine is cost-effectiveness. Orthopedic procedures are among the most expensive medical treatments, and spine procedures can be found among the most cost-intensive orthopedic treatments.^{44,45} The rationale of economic analyses is to help in the comparison of two or more treatments to find the intervention that produces the most value, defined as the highest clinical impact-to-cost ratio, which is also called “efficiency.” In other words, economic analyses are not a “cold-blooded” exercise aiming at withholding treatment for patients to increase profit margins for hospitals and insurance providers, but to allocate the very limited resource of “money” in such a way as to treat the largest possible number of patients with the most effective treatments.^{44,45} Properly conducted outcome analyses are invaluable as a source for valid and reliable estimates of the clinical impact that can be put into relation with the incurred costs.

Most economic analyses come in one of four forms. (1) Cost analysis is the most basic form of economic analysis (i.e., a summation of the costs of a treatment). However, although this seems trivial, quite the contrary is true. Costs are divided into direct and indirect costs; direct costs include obvious factors, such as the cost of implants and/or medication, rent for the operating room, fees for surgeons and ancillary services, among others. Indirect cost are more complicated and include such items as loss of income or the extra cost involved with treatment, such as transport to and from the

hospital or doctor’s office, cost of care for dependents during treatment, and so on. Naturally, all errors in measurement of cost will be perpetuated in further analyses. (2) Although cost-effectiveness analysis (CEA) is commonly used as a synonym for economic analysis, it really is one specific subgroup comparing the cost to clinical effectiveness, as measured in such outcome tools as presented in this review. To compare two treatments, the outcomes are measured, divided by the costs to obtain outcome/dollar values, and then directly compared with see how much more or less outcome is “bought” per dollar. (3) Cost utility analysis (CUA) is very similar to CEA, but uses quality-of-life outcome tools instead of generic ones. By assessing the quality-of-life status and multiplying it by the amount of time spent in this state, the so-called quality-adjusted life-year (QaLY) is estimated. For example, if a patient spends 1 year at only 75% (0.75) of his perfect quality-of-life score (e.g., full SF-36 score), then this 1 calendar year counts as $(0.75 \text{ QoL} \times 1 \text{ year}) = 0.75 \text{ QaLY}$. If he returns to full health the next year, then both calendar years count as $(0.75 \text{ QoL} \times 1 \text{ year}) + (1 \text{ QoL} \times 1 \text{ year}) = 1.75 \text{ QaLY}$. This can then be inserted in a very similar equation as used for CEA. The advantage is that different treatments and diseases that cannot be assessed by the same disease-specific outcome scale can be compared, such as spine fusion and total hip replacement. (4) Cost-benefit analyses compares cost (in dollars) with benefit (in dollars) afforded by a treatment or intervention. The advantage is that the results of such analyses are very readily understood by the illustration of a plus or minus dollar sum, compared with “QaLY/dollar” or “points-ODI/dollar.” The disadvantage is that clinical outcomes have to be converted to a dollar sum. Numerous validated methods exist for this limitation, but even a concise description of these is outside of the scope of this review. ▶Table 3 illustrates the various methods of economic evaluation.

Evidence-Based Clinical Orthopedics

Evidence-based medicine has been defined as “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients.”⁴⁶ Three pieces of information should be included in evidence-based medicine decision making in clinical medicine: patient preferences, clinical circumstance, and research evidence. Diligent outcome assessment studies will provide such research evidence and allow choosing a

Table 3 Methods of economic analysis

	Input	Outcome measure	Outcome tool utilized
Cost-effectiveness analysis	Money units (\$)	Natural outcome measure (life-years saved, infections prevented, etc.)	Disease-specific outcome scores
Cost-utility analysis	Money units (\$)	QoL outcomes measures (QALYs)	Short Form-36, EuroQoL-5D, etc.
Cost-benefit analysis	Money units (\$)	Money unit (\$)	Willingness to pay, human capital, revealed preferences

Abbreviations: QALYs, quality-adjusted life years; QoL, quality of life.

course of action that is consistent with the patient's wishes as well as being efficacious and cost-effective.

Clinical Applications

Outcomes studies mark a substantial detour from the way that medicine is practiced based on practitioner-based observations ("paternalistic model") toward a patient-centric value-based transaction. Consequently, patient-reported quality-of-life outcomes tests ("consumer-based value model") have seen a tremendous increase in their relevance as scientific standards for clinical research have dramatically increased. Consequently, the use of outcomes assessments has left the research environs and is rapidly becoming a standard for the routine measurement of patient outcomes beyond the traditional medical/surgical observations. The development of increasingly hardwired mechanisms to gather these patient-derived outcomes scores is an urgent need for clinical practices with high-cost implications, such as seen in spine care. Doing this in a cost-effective and accountable fashion is challenging and can be very expensive. Limiting outcomes testing to validated core data sets and employing systematic planned data gathering rather than incidental occurrence-based sampling will become desirable features of a future health care delivery model.

The other perspective of patient health-reported quality-of-life assessment tools lies in the prospect of clinical decision making on a front-end basis, rather than using it for outcomes analyses in a mostly post hoc fashion. Our current decision making in spine care remains heavily based on phenotypical assessment strategies, such as physical examination, imaging interpretation, and rather cursory use of some patient symptom reporting system. With growing awareness of the influence of genetic, psychological, and sociologic factors on a large component of spinal disorders as well as the outcomes of treatment, we can expect to increasingly test strategies to account for these components in the front-end part of a future, more comprehensive decision-making process. For example, depression scoring and quantification of distress responsiveness has a substantial impact on patient well-being and subsequently patient health-reported quality-of-life. It would make sense to modify the intervention strategies under consideration of such insights gained. Similarly, employment status, socioeconomic health, and demographics play a role in the resource utilization, such as readmission, return to functional status, and readmission rates. These factors deserve consideration.

Finally, we are now aware of the genetic basis of many diseases. The propensity for premature symptomatic disk degeneration, osteoporosis, and deformities are all examples of spinal disorders with well understood roots in genetic abnormalities.⁴⁷⁻⁵¹ Similarly, pain response mechanisms may have a genetic foundation.⁴⁸ Rapid progress in understanding the causal correlations will lead toward the creation of inter-relational data repositories that will—once taken into cumulative consideration—allow for more individualized decision making in health care delivery under utilization of an artificial intelligence. Although this prospect might appear frightening to some, the more positive interpretation is that of

allowing for better and more directed "personalized" health care delivery. And that prospect would seem to be of great appeal to all involved. As such, the outcomes tools assessments are just the beginning of a more sophisticated understanding of health care transactions.

Problems and Potential Sources of Bias in the Use of Outcome Scales

Several potential problems arise with the use of clinical outcomes scores in the outcome research. Although we cannot list all of them in this text, we strive to include some of the most meaningful and often most disregarded ones.

The most pertinent bias is internal and external validity. Internal validity describes the ability of a test to assess a desired outcome and is often explained as using a thermometer, not a scale, to measure temperature. Validity is often mentioned along with, but is not synonymous with, reliability, which is the test-retest consistency of an outcome tool. To stay with the former example, a scale might have excellent reliability but poor validity in a study of temperatures. External validity, in turn, describes if results of a study can be applied to a population outside of the study. With highly specific and rigorously controlled inclusion criteria, this is oftentimes questionable.

Another bias related to a test's ability to measure an outcome is the ceiling effect.⁴² With improved medical technology, treatment outcomes have improved substantially and success rates above 80% are not uncommon. However, with such remarkable successes, further improvements are often only small in scale (i.e., a ceiling effect has been reached). Because many clinical scores are designed somewhat coarse for the sake of general applicability and easier understanding, their "resolution" cannot detect the small changes that are possible with ceiling effects.

Minimum Clinically Important Difference

Another important concept discussing the size of improvements measured by a clinical test is the minimum clinically important difference (MCID). The original meaning of the MCID was the "smallest difference in scores in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side-effects and excessive cost, a change in the patient's management."⁵² In the course of further investigations, it was simplified to define the threshold value of minimum treatment effectiveness, which stands for the amount of change that is considered to be meaningful and worthwhile by the patient.^{53,54} It therefore describes the smallest change in outcome that is important to the patient.^{55,56} It is seen as a possibility to overcome the shortcomings or deficiencies of the "statistically significant difference" to help clinicians evaluate the importance of seemingly statistically significant results.⁵⁴ Methods to calculate the MCID are classified into anchor-based methods, which compare the change in patient-reported outcome score to some other measure of change (considered an external criterion), or distribution-based methods, which compare the

change in patient-reported outcome scores to some measure of variability, such as the standard error of measurement, the standard deviation, the effect size, or the minimum detectable change.^{53,54} The latter leads to a difficult analysis and evaluation as all the different calculation methods result in different MCID values.

Conclusion

Spinal disorders affect every population worldwide. Treatment modalities of spine conditions can vary between centers and surgeons. Novel and elaborate methods are being introduced that contribute to a surgeon's armamentarium on a daily basis. When choosing a clinical outcome instrument, it is important to make sure that it will reflect the anticipated end point. It is prudent to consider its generalizability or its ability to be compared with other instruments or prior literature. Although it cannot be stated that one questionnaire or outcome instrument is superior to another, some are being used more frequently than others. As such, one should make sure to be aware which outcome instruments are used most frequently in one's own field of research. Also, it is important to ensure that the chosen instrument has been validated for the anticipated use. In addition, the perceived outcomes and financial issues may vary geographically. As such, there is a need to continue research assessing such outcome instruments and their outcomes between more developed to less developed countries.

Funding

This work was supported by grants by the Hong Kong Theme-Based Research Scheme (T12-708/12N) and the Hong Kong Research Grants Council (777111).

Disclosures

Patrick Vavken, none
Anne Kathleen B. Ganal-Antonio, none
Julia Quidde, none
Francis H. Shen, none
Jens R. Chapman, none
Dino Samartzis, none

References

- Liang MH, Lew RA, Stucki G, Fortin PR, Daltroy L. Measuring clinically important changes with patient-oriented questionnaires. *Med Care* 2002;40(4, Suppl):II45-II51
- Lurie J. A review of generic health status measures in patients with low back pain. *Spine (Phila Pa 1976)* 2000;25(24):3125-3129
- Németh G. Health related quality of life outcome instruments. *Eur Spine J* 2006;15(Suppl 1):S44-S51
- Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-483
- Aaronson NK, Acquadro C, Alonso J, et al. International Quality of Life Assessment (IQOLA) Project. *Qual Life Res* 1992;1(5):349-351
- Samartzis D, Dominique DA, Perez-Cruet MJ, Fehlings MG. Clinical outcome analyses. In: Perez-Cruet MJ, Khoo LT, Fessler RG, eds. *An Anatomical Approach to Minimally Invasive Spine Surgery*. St. Louis, MO: Quality Medical Publishing, Inc.; 2006:103-130
- Gilson BS, Gilson JS, Bergner M, et al. The sickness impact profile. Development of an outcome measure of health care. *Am J Public Health* 1975;65(12):1304-1310
- Danielsson AJ, Hasselius R, Ohlin A, Nachemson AL. Health-related quality of life in untreated versus brace-treated patients with adolescent idiopathic scoliosis: a long-term follow-up. *Spine (Phila Pa 1976)* 2010;35(2):199-205
- Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. *J Manipulative Physiol Ther* 1991;14(7):409-415
- Riddle DL, Stratford PW. Use of generic versus region-specific functional status measures on patients with cervical spine disorders. *Phys Ther* 1998;78(9):951-963
- Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF. Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. *Spine (Phila Pa 1976)* 2002;27(5):515-522
- Wheeler AH, Goolkasian P, Baird AC, Darden BV II. Development of the Neck Pain and Disability Scale. Item analysis, face, and criterion-related validity. *Spine (Phila Pa 1976)* 1999;24(13):1290-1294
- Wlodyka-Demaille S, Poiraudou S, Catanzariti JF, Rannou F, Fermanian J, Revel M. The ability to change of three questionnaires for neck pain. *Joint Bone Spine* 2004;71(4):317-326
- Kang SS, Lee JS, Shin JK, Lee JM, Youn BH. The association between psychiatric factors and the development of chronic dysphagia after anterior cervical spine surgery. *Eur Spine J* 2014;23(8):1694-1698
- Chapman JR, Norvell DC, Hermsmeyer JT, et al. Evaluating common outcomes for measuring treatment success for chronic low back pain. *Spine (Phila Pa 1976)* 2011;36(21, Suppl):S54-S68
- Fairbank JCT, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66(8):271-273
- Fairbank JCT, Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976)* 2000;25(22):2940-2952, discussion 2952
- Fairbank JC. Use and abuse of Oswestry Disability Index. *Spine (Phila Pa 1976)* 2007;32(25):2787-2789
- Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine (Phila Pa 1976)* 1983;8(2):141-144
- Roland M, Morris R. A study of the natural history of low-back pain. Part II: development of guidelines for trials of treatment in primary care. *Spine (Phila Pa 1976)* 1983;8(2):145-150
- Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine (Phila Pa 1976)* 2000;25(24):3115-3124
- Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research. A proposal for standardized use. *Spine (Phila Pa 1976)* 1998;23(18):2003-2013
- Stratford PW, Binkley J, Solomon P, Finch E, Gill C, Moreland J. Defining the minimum level of detectable change for the Roland-Morris questionnaire. *Phys Ther* 1996;76(4):359-365, discussion 366-368
- Stucki G, Daltroy L, Liang MH, Lipson SJ, Fossel AH, Katz JN. Measurement properties of a self-administered outcome measure in lumbar spinal stenosis. *Spine (Phila Pa 1976)* 1996;21(7):796-803
- Comer CM, Conaghan PG, Tennant A. Internal construct validity of the Swiss Spinal Stenosis questionnaire: Rasch analysis of a disease-specific outcome measure for lumbar spinal stenosis. *Spine (Phila Pa 1976)* 2011;36(23):1969-1976
- Tomkins CC, Battié MC, Hu R. Construct validity of the physical function scale of the Swiss Spinal Stenosis Questionnaire for the measurement of walking capacity. *Spine (Phila Pa 1976)* 2007;32(17):1896-1901
- Tomkins-Lane CC, Battié MC. Validity and reproducibility of self-report measures of walking capacity in lumbar spinal stenosis. *Spine (Phila Pa 1976)* 2010;35(23):2097-2102

- 28 Yi H, Wei X, Zhang W, et al. Reliability and validity of simplified Chinese version of Swiss Spinal Stenosis Questionnaire for patients with degenerative lumbar spinal stenosis. *Spine (Phila Pa 1976)* 2014;39(10):820–825
- 29 Lee YS, Park SW, Kim YB. Direct lateral lumbar interbody fusion: clinical and radiological outcomes. *J Korean Neurosurg Soc* 2014; 55(5):248–254
- 30 Haheer TR, Gorup JM, Shin TM, et al. Results of the Scoliosis Research Society instrument for evaluation of surgical outcome in adolescent idiopathic scoliosis. A multicenter study of 244 patients. *Spine (Phila Pa 1976)* 1999;24(14):1435–1440
- 31 Asher MA, Min Lai S, Burton DC. Further development and validation of the Scoliosis Research Society (SRS) outcomes instrument. *Spine (Phila Pa 1976)* 2000;25(18):2381–2386
- 32 Berven S, Deviren V, Demir-Deviren S, Hu SS, Bradford DS. Studies in the modified Scoliosis Research Society Outcomes Instrument in adults: validation, reliability, and discriminatory capacity. *Spine (Phila Pa 1976)* 2003;28(18):2164–2169, discussion 2169
- 33 Asher M, Min Lai S, Burton D, Manna B. The reliability and concurrent validity of the scoliosis research society-22 patient questionnaire for idiopathic scoliosis. *Spine (Phila Pa 1976)* 2003; 28(1):63–69
- 34 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 (Phila Pa 1976)* 2007;32(20): 2220–2225
- 35 Baldus C, Bridwell K, Harrast J, et al. The Scoliosis Research Society Health-Related Quality of Life (SRS-30) age-gender normative data: an analysis of 1346 adult subjects unaffected by scoliosis. *Spine (Phila Pa 1976)* 2011;36(14):1154–1162
- 36 Dorward IG, Lenke LG, Stoker GE, Cho W, Koester LA, Sides BA. Radiographic and clinical outcomes of posterior column osteotomies in spinal deformity correction. *Spine* 2014; February 27 (Epub ahead of print)
- 37 Freyd M. The graphic rating scale. *J Educ Psychol* 1923;43:83–102
- 38 Haefeli M, Elfering A. Pain assessment. *Eur Spine J* 2006;15 (Suppl 1):S17–S24
- 39 Von Korff M, Jensen MP, Karoly P. Assessing global pain severity by self-report in clinical and health services research. *Spine (Phila Pa 1976)* 2000;25(24):3140–3151
- 40 Lee JS, Youn MS, Shin JK, Goh TS, Kang SS. Relationship between cervical sagittal alignment and quality of life in ankylosing spondylitis. *Eur Spine J* 2014; August 12 (Epub ahead of print)
- 41 Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain* 1975;1(3):277–299
- 42 Vavken P. Rationale for and methods of superiority, noninferiority, or equivalence designs in orthopaedic, controlled trials. *Clin Orthop Relat Res* 2011;469(9):2645–2653
- 43 Vavken P, Culen G, Dorotka R. [Clinical applicability of evidence-based orthopedics—a cross-sectional study of the quality of orthopedic evidence]. *Z Orthop Unfall* 2008;146(1):21–25
- 44 Drummond M, Sculpher MJ, Torrance G, O'Brien B, Stoddart G. Cost-effectiveness Analysis. Methods for the Economic Evaluation of Health Care Programmes. 3rd ed. Oxford, UK: Oxford University Press; 2005:1103–1136
- 45 Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA* 1996;276(15):1253–1258
- 46 Bhandari M, Giannoudis PV. Evidence-based medicine: what it is and what it is not. *Injury* 2006;37(4):302–306
- 47 Fan YH, Song YQ, Chan D, et al. SNP rs11190870 near LBX1 is associated with adolescent idiopathic scoliosis in southern Chinese. *J Hum Genet* 2012;57(4):244–246
- 48 Karppinen J, Shen FH, Luk KD, Andersson GB, Cheung KM, Samartzis D. Management of degenerative disk disease and chronic low back pain. *Orthop Clin North Am* 2011;42(4):513–528
- 49 Kao PY, Chan D, Samartzis D, Sham PC, Song YQ. Genetics of lumbar disk degeneration: technology, study designs, and risk factors. *Orthop Clin North Am* 2011;42(4):479–486
- 50 Eskola PJ, Lemmelä S, Kjaer P, et al. Genetic association studies in lumbar disc degeneration: a systematic review. *PLoS ONE* 2012; 7(11):e49995
- 51 Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat Genet* 2012;44(5):491–501
- 52 Copay AG. Commentary: the proliferation of minimum clinically important differences. *Spine J* 2012;12(12):1129–1131
- 53 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(6):968–974
- 54 Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J* 2007;7(5):541–546
- 55 Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 2. *Phys Ther* 1998; 78(11):1197–1207
- 56 Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1. *Phys Ther* 1998;78(11):1186–1196