A proposed set of metrics for standardized outcome reporting in the management of low back pain

Clement, R. Carter; Welander, Adina; Stowell, Caleb; Cha, Thomas D.; Chen, John L.; Davies, Michelle; Fairbank, Jeremy C.; Foley, Kevin T.; Gehrchen, Martin; Hagg, Olle; Jacobs, Wilco C.; Kahler, Richard; Khan, Sajdar N.; Lieberman, Isador H.; Morisson, Beth; Ohnmeiss, Donna D.; Peul, Wilco C.; Shonnard, Neal H.; Smuck, Matthew W.; Solberg, Tore K.; Strömqvist, Björn; Van Hooff, Miranda L.; Wasan, Ajay D.; Willems, Paul C.; Yeo, William; FRitzell, Peter

Published in:
Acta Orthopaedica

DOI:
10.3109/17453674.2015.1036696

2015

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
A proposed set of metrics for standardized outcome reporting in the management of low back pain


To link to this article: http://dx.doi.org/10.3109/17453674.2015.1036696

Copyright: © Nordic Orthopaedic Federation
Published online: 01 Sep 2015.
Submit your article to this journal
Article views: 796
View related articles
View Crossmark data
Citing articles: 2 View citing articles of
A proposed set of metrics for standardized outcome reporting in the management of low back pain

R Carter Clement1, Adina Welander2, Caleb Stowell3, Thomas D Cha4, John L Chen5, Michelle Davies6, Jeremy C Fairbank7, Kevin T Foley8, Martin Gehrchen9, Olle Hagg10, Wilco C Jacobs11, Richard Kahler12, Safdar N Khan13, Isador H Lieberman14, Beth Morisson15, Donna D Ohnmeiss14, Wilco C Peul16, Neal H Shonnard17, Matthew W Smuck18, Tore K Solberg19,20, Bjorn H Stromqvist21, Miranda L Van Hooff22, Ajay D Wasan23, Paul C Willemse24, William Yeo25, and Peter Fritzell26,27

1 Department of Orthopaedic Surgery, University of North Carolina Hospitals, Chapel Hill, NC, USA; 2 Boston Consulting Group, Stockholm, Sweden; 3 International Consortium for Health Outcomes Measurement, Stockholm, Sweden; 4 Department of Orthopaedics, Massachusetts General Hospital, Boston, MA, USA; 5 Department of Orthopaedic Surgery, Singapore General Hospital, Singapore; 6 Patient representative, Brisbane, Australia; 7 Nuffield Orthopaedic Centre, University of Oxford, Oxford, UK; 8 Department of Neurosurgery, University of Tennessee Health Science Center, Memphis, TN, USA; 9 Department of Orthopaedic Surgery, Spine Section, Rigshospitalet, University Hospital of Copenhagen, Copenhagen, Denmark; 10 Spine Center Göteborg, Sweden; 11 Department of Neurosurgery, Leiden University Medical Center, Leiden, the Netherlands; 12 BrizBrain and Spine, Brisbane, Australia; 13 Department of Orthopaedics, Division of Spine, Ohio State University, Wexner Medical Center, Columbus, OH, USA; 14 Texas Back Institute, Plano, TX, USA; 15 Neuro Foundation, Brisbane, Australia; 16 Department of Neurosurgery, Leiden University Medical Center and Medical Centre Haaglanden, Leiden and The Hague, the Netherlands; 17 Director, Spine SCOAP Collaborative, Rainier Orthopedic Institute, Puyallup, WA, US; 18 Department of Orthopaedic Surgery, PM&R Section, Stanford University, Palo Alto, CA, USA; 19 Department of Ophthalmology and Neurosurgery, University Hospital of Northern Norway, Tromsø, Norway; 20 The Norwegian Registry for Spine Surgery (NORspine), Northern Norway Regional Health Authority, Tromsø, Norway; 21 Department of Orthopaedics, Clinical Sciences Lund, Lund University Hospital, Lund, Sweden; 22 Sint Maartenskliniek, Research, Nijmegen, the Netherlands; 23 Department of Anesthesiology, University of Pittsburgh Medical Center Pittsburgh, PA, USA; 24 Maastricht University Medical Center, Maastricht, the Netherlands; 25 Orthopaedic Diagnostic Centre, Singapore General Hospital, Singapore; 26 Department of Orthopaedic Surgery, Ryhov Hospital, Jönköping, Sweden; 27 Registry Manager, Swespine.

Correspondence: carter.clement@gmail.com


Measurement of outcomes in healthcare has well-documented benefits as well as challenges (Porter 2005, Institute of Medicine 2006). Simply asking providers to report their outcomes has been shown to improve performance (Porter et al. 2010). Additionally, understanding one’s results empow-
ers a provider to continuously learn from and refine the care
he or she delivers (Porter and Teisberg 2004). On a broad
scale, outcome reporting also facilitates dissemination of
best practices between physicians and makes it possible to
compare the quality delivered by different providers, allow-
ing patients to make intelligent choices about where to seek
care (Porter and Teisberg 2004). This type of continuous
improvement and informed decision making could be an
important driving force in improving healthcare delivery
by refocusing the system on value (defined as the outcomes
of care divided by the cost). The concept of “value-based
healthcare” has been gaining attention both throughout the
medical field (Porter and Teisberg 2005, Porter 2009) and
specifically within the realm of spine care (McGirt et al.
2014a, 2014b). With evolving reimbursement systems in
many countries, it is also conceivable that there will be grow-
ing interest in “value-based reimbursement” in the future,
with payment levels adjusted based on outcomes. This type
of scheme will only be fair with a broadly-accepted and risk-
adjusted set of outcome metrics.

Low back pain (LBP) is a growing problem and constitutes
a major component of the global burden of disease (Murray
et al. 2012). Measuring outcomes in the field of low back
pain is challenging. Numerous disease states affect the lower
back, resulting in low back pain, leg pain, or both; to com-
pare outcomes, patients must be accurately stratified by both
diagnosis and severity. Moreover, existing treatment algo-
rithms are complex and often controversial, including both
operative and nonoperative options and frequently requiring
multidisciplinary provider teams. Additionally, low back
pain rarely causes death or other objective endpoints, so
outcomes are best measured with patient-reported metrics,
which are inherently subjective and require thorough psy-
chometric testing.

A substantial amount of work on the design of outcome
metrics has already been done in the field of low back pain,
and there are several well-validated tools for measuring dis-
eease-specific outcomes (Longo et al. 2010). Similarly, several
large registries are already in existence, collecting outcomes
along with many other data points (Röder et al. 2005, McGirt
et al. 2013, Strömqvist et al. 2013). Previous consensus-based
efforts have been made to define sets of outcome measures
or domains for research purposes (Deyo et al. 1998, Pincus
et al. 2008, Chiarotto et al. 2014, Deyo et al. 2014). Still, the
field of low back pain care has not yet developed a universal
international set of outcomes to be measured and compared as
a part of standard clinical practice. This type of outcome set
requires availability and validity in many languages, requires
capacity for case-mix adjustment to ensure that comparisons
are made fairly, and should focus on the outcomes that matter
most to patients. The purpose of this study was to define such
a set based on international and interdisciplinary expert and
patient opinion.

Methods

The set of outcomes we present, referred to as the standard
set, was developed by consensus among a 22-member “work-
group” mostly comprised of surgical, rehabilitation, and
medical experts in the field of low back pain, many of
whom are active in spine registries (all members are listed
as authors). The group also included a former spine patient
involved in patient support groups (MD). The working group
was convened and organized by the International Consortium
for Health Outcomes Measurement (ICHOM), a non-profit
organization focused on the development of standard sets of
outcomes and risk factors for multiple medical conditions
(ICHOM Website 2014b). The working group’s efforts were
coordinated by a core “project team” consisting of a working
group leader (PF), a project leader (AW), a research fellow
(RC), and the ICHOM vice president of research and develop-
ment (CS).

The project was structured as a modified Delphi process
(Pill 1971) involving 6 teleconferences held between June
and November of 2013. The goals of these calls were choos-
ing inclusion and exclusion criteria for the relevant patient
population, selecting and defining outcome metrics, and iden-
tifying initial disease conditions and risk factors that would
allow patient stratification and case-mix adjusted comparisons
between providers. Teleconferences were structured around
proposals by the project team regarding how best to meet the
goals of the group. These proposals were based on review of
academic literature, review of existing practices in spine reg-
istries, and in some cases, direct input from working group
members and other experts in the field.

Decisions were made by surveys, which were designed
based on the project team’s proposals and the relevant discus-
sions held during the teleconference. Surveys were circulated
to all working group members by e-mail after teleconferences,
along with detailed minutes. In a small number of cases, live
votes were orchestrated during a call. For surveys and votes
with less than a two-thirds majority or with a particularly vig-
orous debate, the issue was revisited by the project team and
a new proposal was presented to the working group for con-
sideration.

Several recurrent themes emerged throughout this pro-
cess, and developed into guiding principles for the group’s
collaboration. Firstly, we aimed to identify outcome metrics
that are most important to patients, which often resulted in
favoring subjective information reported by patients rather
than objective clinical information traditionally followed by
physicians. Secondly, we sought genuine outcome metrics
to gauge quality, not process metrics—which are often used
as inexact proxies for quality, as they are frequently easier to
track. Third, a consistent effort was made to simplify the set
of outcomes and associated data, especially the information
requested from physicians in order to boost compliance. As
such, we acknowledge that the goal of the standard set should
be to allow comparisons of clinical outcomes and, while it will be sufficient to answer certain research questions, many academic pursuits will require collection of additional data points. Fourth, when possible, existing tools with proven validity and reliability such as the Oswestry disability index (ODI) and EQ-5D were selected in their original format to preserve their proven psychometric properties. Finally, a conscious effort was made to be continually aware of potential bias favoring surgical patients, given the predominance of surgeons in the working group, which reflects the predominant focus on surgical patients in the existing spine registries.

ICHOM had access to all data during the project, but neither ICHOM nor its funders had editorial control over the final publication. The manuscript was drafted by the project team’s research fellow (RC) and subsequently edited based on input from all the experts and co-authors.

**Results**

Response rates to the 5 surveys among the 22 working group members were 21, 20, 21, 20, and 21, respectively. Two original working group members participated in less than half of the teleconferences and surveys and are not included either in these response rates or in the final list of members.

**Scope: Degenerative lumbar conditions**

The standard set targets degenerative lumbar conditions, which comprise by far the greatest part of all lumbar pathology (Andersson 1997). Other areas of spine care involve different patient populations, treatment approaches, and outcomes—and should be addressed in the future with analogous condition-specific outcome sets. Formal inclusion criteria selected by the working group consisted of lumbar spinal stenosis, lumbar spondylolisthesis, degenerative disc disorders including disc herniation, degenerative scoliosis, other degenerative lumbar disorders, and acute and chronic lumbar back pain and back-related leg pain without a clear etiology (often colloquially termed mechanical or nonspecific pain). The relevant corresponding exclusion criteria included spinal infection, tumor, fracture, traumatic dislocation, congenital or idiopathic scoliosis, and age under 18 years.

**Outcome domains (Table 1)**

Traditionally, the 6 domains most commonly used to study outcomes among patients with degenerative lumbar conditions have been function, pain, health-related quality of life (HRQOL), work status, treatment complications, and medication requirements (Chapman et al. 2011). This pattern suggests that historically, spine providers have felt that these domains most accurately reflect success rather than failure in this field. Furthermore, after careful consideration including discussion with the group’s patient representative, the working group considered that these are the factors that matter most to patients. The group also agreed that the combination of these factors provides adequate domain coverage for comprehensive assessment of treatment outcomes in this population. Other metrics that have been used to study LBP care—including psychosocial factors such as depression and “global effect” (Chapman et al. 2011)—were excluded from the set, as historically they have been studied with inconsistent definitions (Chapman et al. 2011) and they are probably reflected in other domains such as HRQOL.

**Patient-reported outcome measures (PROMs)**

The core component of the standard set is a constellation of PROMs covering the 6 domains listed above, collected at the time of enrollment for treatment and then at regular time points. (As detailed below, some information on clinical complications also requires clinician reporting). PROM instruments were chosen by the working group on the basis of clinical interpretability, feasibility of implementation, and psychometric properties (validity, reliability, and responsiveness) (Cleland et al. 2012).

Common and well-validated methods for measuring pain include the numeric rating scale (NRS) and the visual analog scale (VAS) (Jensen et al. 2008, Chapman et al. 2011), and the major existing spine registries are divided between those options (Röder et al. 2005, McGirt et al. 2013, Strömqvist et al. 2013). While there is no gold standard, a VAS allows patients to provide a more specific response while an NRS is usually easier to use as it can be performed verbally and does not require exact size calibration when reprinted or generated on a monitor. The common 0–10 horizontal version asking for average pain over the last week has been shown to be valid, reliable, and to allow adequately specific responses among spine patients (Andersson 1997, Jamison et al. 2006, Jensen et al. 2008, Bolton et al. 2010). This option was chosen by the working group (with 21 of 22 members in agreement) for inclusion in the standard set, for both back and leg pain individually.

Numerous tools have been studied for measuring lumbar-related function in patients with low back pathology (Longo et al. 2010, Chapman et al. 2011). ODI is the most commonly used and cited tool for this purpose, followed by the Roland Morris disability questionnaire (RMDQ) (Chapman et al. 2011) and the core outcome measures index (COMI) (PubMed Search for “COMI, Back Pain, 2001–2011” in 2013). While all of these have been shown to be valid, reliable, and responsive in this population, the ODI is the most heavily studied, providing superior clinical interpretability (Chapman et al. 2011). We also felt that the ODI is the most feasible to implement, as it has been validated in 14 languages (as opposed to 9 for each the RMDQ (Chapman et al. 2011) and COMI (Jamison et al. 2006)) and is relatively short (10 items as opposed to 24 in the RMDQ (Roland and Morris 1983) and 7 in the COMI (Mannion et al. 2009)). All are free, with online registration being required for use of the ODI (Chapman et al. 2011). For these
### Table 1. Patient-reported outcome measures

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Measurement tool</th>
<th>Definition/wording</th>
<th>Answer options</th>
<th>Time frame for capturing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pain</strong></td>
<td>Numeric pain rating scale</td>
<td>• How would you rate your average back pain over the last week?</td>
<td>0 (no pain) – 10 (worst pain imaginable)</td>
<td>Baseline, index event(s), 6 months, 1 year, 2 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• How would you rate your average leg pain over the last week?</td>
<td>0 (no pain) – 10 (worst pain imaginable), verbal or visual (horizontal)</td>
<td></td>
</tr>
<tr>
<td><strong>Disability</strong></td>
<td>Oswestry disability index</td>
<td>• Pain intensity</td>
<td>6 options for each domain, ranging from no problem to severe impairment</td>
<td>Baseline, index event(s), 6 months, 1 year, 2 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Personal care (washing, dressing, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Lifting</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Walking</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Sitting</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Standing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Sleeping</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Sex life (if applicable)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Social life</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Traveling</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Quality of life</strong></td>
<td>EQ5D-3L</td>
<td>• Mobility</td>
<td>3 options for each domain, ranging from no problem to severe impairment</td>
<td>Baseline, index event(s), 6 months, 1 year, 2 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Self-care</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Usual activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Pain/discomfort</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Anxiety/depression</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Work status</strong></td>
<td>EQ-VAS 6 items</td>
<td>• Indicate on this scale how good or bad your health is today</td>
<td>Vertical visual analog scale: 0 (worst imaginable health state) – 100 (best imaginable health state)</td>
<td>Baseline, index event(s), 6 months, 1 year, 2 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Analgesic use</strong></td>
<td></td>
<td>• What is your current work status?</td>
<td>Working full time, working part time, seeking employment (I consider myself able to work but can’t find a job), not working by choice (retired, student, homemaker, etc.), unable to work due to problem other than my back and/or leg pain, unable to work due to back and/or leg pain</td>
<td>Baseline, index event(s), 6 months, 1 year, 2 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Are you working at a physically less demanding job now because of your back and/or leg pain?</td>
<td>Yes, no, N/A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• How long after you received treatment for low back pain did you return to work? (if applicable)</td>
<td>&lt; 3 months, 3–6 months, 6–9 months 9–12 months, 1–2 years, &gt; 2 years</td>
<td>6 months, 1 year, 2 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

reasons, the working group unanimously chose the ODI 2.1a for inclusion in the standard set.

There are several tools for measurement of HRQOL in LBP patients (Bergner et al. 1976, Hunt et al. 1980, Ware and Sherbourne 1992, Ware et al. 1996, Brooks 1996, Hung et al. 2014), with the most common and heavily studied being the SF-36 followed by the EQ-5D and accompanying EQ-VAS, Nottingham health profile (NHP), and SF-12 (Chapman et al. 2011). The SF-36 has been shown to be valid, reliable, and responsive in this population, while the NHP and SF-12 have proven to be valid and reliable (Chapman et al. 2011). To our knowledge, these have not been studied for responsiveness and none of the psychometric properties of the EQ-5D have yet been examined in LBP patients. However, the EQ-5D tool has an excellent track record among other demographics as well as in the general population (Hinz et al. 2013, Kim et al. 2013), and it has been shown to correlate well with the ODI in LBP patients (Mueller et al. 2013). Additionally, the volume of recent citations suggests a relatively rapid increase in the use and dissemination of this tool, which is consistent with the anecdotal experience of working group members. The EQ-5D and EQ-VAS also has the advantage of being relatively brief (6 items as opposed to 36 in the SF-36, 38 in the NHP, and 12 in the SF-12) and has proven psychometric properties in over 160 languages (in comparison to 155 for the SF-36, 2 for the NHP, and 134 for the SF-12) (Chapman et al. 2011).
The EuroQol tool is also inexpensive (EuroQol Website 2013) relative to the SF tools (Ware and Sherbourne 1992, SF36 Official Website 2013), while use of the NHP is free. Lastly, the EQ-5D is superior for health economics evaluations as it is a preference-based tool that allows utility calculations and cost-effectiveness analysis (Chapman et al. 2011). For these reasons, the working group chose the EQ-5D for inclusion in the standard set, with 21 of 22 members in favor.

Existing practices used by current registries for questioning of patients about analgesic use and working status were reviewed, and the approach used by the international Spine Tango registry was felt to be the most concise and thorough (Röder et al. 2005); the wording was modified slightly by the working group.

Complications and adverse events
Adverse consequences of treatment, e.g. invasive procedures, make up another category of outcomes. While no objective criteria were used, the working group aimed to include complications and adverse events that are relatively frequent, severe, avoidable, and feasible to capture. Careful attention was paid to the balance between gathering sufficient data to allow comparisons between providers and keeping the collection process simple enough to facilitate a high level of compliance. The decision was made to request that providers report complications/adverse effects recognized at the time of an initial procedure or during the associated hospitalization, which is considered the index period. Subsequently, when completing PROMs questionnaires 6 months after an index period, patients should be asked to report specified complications that occurred after this period. The interventions of interest are operations and injection therapy, and for convenience the same list of complications and time frame for collection should be used for both.

Early provider-reported complications selected for inclusion during the index period include death, nerve injury, dural tear, vascular injury, deep infection, and pulmonary embolus (PE) (Table 2). In practices where reliable administrative data reflecting death records are readily accessible, the working group recommends the use of such data to more accurately track out-of-hospital mortality within the first 30 days. At the time of follow-up PROM questionnaires, patients should be asked if they experienced a deep wound infection or PE, as these can be particularly detrimental complications but may only occur or be recognized after the index period. As providers may not be made aware of unplanned re-hospitalizations within 30 days of the index period, which have become a popular healthcare quality metric, patients should also be asked to report such events (Report to Congress 2007, Axon and Williams 2011). In countries and practices with reliable administrative documentation of re-hospitalization such as electronic medical records or insurance databases, the working group recommends using these administrative data to record such events. Reoperations after an index procedure, and the underlying cause, should be reported by providers (Table 2).

Baseline characteristics and risk factors for case-mix adjustment
In order to statistically adjust analyses for fair and meaningful calculations, relevant data on patients’ risk factors and initial conditions must be collected. The working group tried to balance the time and financial cost of collecting data with the need for accurate comparisons, while seeking internationally comparable data points. This information was addressed in 4 categories: demographics, baseline clinical status, baseline functional status, and previous treatments (Table 3). Common demographics currently in use in international registries were reviewed and age, sex, and socioeconomic status were chosen, with education level being used as an internationally acceptable proxy for the latter. Specifically, the United Nations Educational, Scientific, and Cultural Organization (UNESCO) definitions of education levels, which allow for international and cross-cultural comparisons, were selected for use (United Nations Educational, Scientific and Cultural Organization 2013). Race and ethnicity were discussed but they were ultimately felt to be of limited value as risk adjusters.

To define a patient’s baseline clinical status, the lumbar pathology criteria defined and studied by Glassman et al. (2011) were selected, primarily for their applicability to both operative and conservatively treated patients (Table 3). To our knowledge, no single tool has been validated to define the diagnoses of patients across the entire realm of degenerative lumbar pathology, and the Glassman criteria are the only such tool that has been shown to be reliable between providers. Additionally, our review suggests that providers will rapidly be able to learn and use these criteria. In addition to these clinical data, indications for surgery should be recorded to facilitate risk stratification. After review of the literature and the current registries, the set of operative indications used by the Swespine Registry (Strömqvist et al. 2013) was felt to be the most complete yet concise example of such a list, and was chosen for inclusion in the standard set, to be completed by providers at the time of surgery (Table 3). Also, the American Society of Anesthesiologists (ASA) Physical Status Classification System has been shown to be prognostic for many surgical procedures (Bo et al. 2007, Schoenfeld et al. 2013, Tabouret et al. 2013) and the working group felt that it should be reported before surgery.

In addition to data related directly to the lumbar spine and surgical risk, a patient’s baseline clinical status also encompasses other comorbidities, which have historically served as a basis for risk adjustment in large patient populations. Patient-reported responses to the Charlson comorbidity index have been proven to be predictive of both mortality and various PROMs (Bayliss et al. 2005, Chaudhry et al. 2005). To our knowledge, no comorbidity list has been validated for risk adjustment in LBP patients. For this purpose, we chose the collection of 13 conditions used by the UK National Health Service for risk stratification in total hip replacement (Department of Health 2012). This list was augmented with 2 condi-
sections included in the Charlson index that the working group considered particularly prescient in the LBP population: paraplegia/hemiplegia and HIV/AIDS (Table 3). Smoking habits (Jenkins et al. 1994, Shimia et al. 2013) and BMI (Papavero et al. 2009, Rihn et al. 2012, 2013) have been shown to provide prognostic value in lumbar patients and were therefore also designated for collection at baseline. It should be noted that depression—which is included among the patient-reported comorbidities described above and which has been shown to be predictive of outcomes among spine patients (Trief et al. 2006, Celestin et al. 2009, Daubs et al. 2011)—was discussed at length, and the working group concluded that this information should be collected by patient report rather than formal depression screening or physician report, both for the sake of efficiency and because depression is probably reflected in other PROMs such as HRQOL. Lastly, some PROMs collected at baseline provide relevant information about a patient’s baseline clinical status and should be used for risk adjustment analyses—namely pain level, duration of symptoms, and current analgesic use.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Definition/wording</th>
<th>Answer options</th>
<th>Time frame for capture</th>
<th>Reported by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality</td>
<td>Death in hospital (all-cause mortality)</td>
<td>Yes/no</td>
<td>While in-house for procedure</td>
<td>Provider</td>
</tr>
<tr>
<td>Nerve root injury (including cauda equina syndrome)</td>
<td>Iatrogenic nerve root damage</td>
<td>Yes/no</td>
<td>— “ —</td>
<td>— “ —</td>
</tr>
<tr>
<td>Vascular injury</td>
<td>Clinically significant iatrogenic damage to a vessel</td>
<td>Yes/no</td>
<td>— “ —</td>
<td>— “ —</td>
</tr>
<tr>
<td>Dural tear</td>
<td>Iatrogenic damage of the dura with liquor emission (e.g. hematoma, malpositioned implant, DVT without PE, device failure, persistent donor-site pain, other)</td>
<td>Yes/no</td>
<td>— “ —</td>
<td>— “ —</td>
</tr>
<tr>
<td>Other b</td>
<td>Post-intervention deep/subfascial wound infection</td>
<td>Yes/no</td>
<td>While in-house for procedure and again on next patient follow-up questionnaire</td>
<td>Provider reports if occurring in-house, otherwise patient reports at next follow-up</td>
</tr>
<tr>
<td>Deep wound infection b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulmonary embolus b</td>
<td>PE diagnosed by radiological study after the intervention</td>
<td>Yes/no</td>
<td>— “ —</td>
<td>— “ —</td>
</tr>
<tr>
<td>Re-hospitalization</td>
<td>Were you admitted to an acute care facility as an inpatient within 30 days from the date of your intervention for ANY reason? (Do not include admissions to rehabilitation hospital or nursing home)</td>
<td>Yes/no, date(s)</td>
<td>Next patient follow-up questionnaire</td>
<td>Patient</td>
</tr>
<tr>
<td>Need for reoperation (if yes, specify cause) c</td>
<td>Second or multiple performed interventions caused by complications after index surgery, not planned in advance</td>
<td></td>
<td>At time of reoperation</td>
<td>Provider</td>
</tr>
<tr>
<td>Hardware removal</td>
<td>Removal of implants, e.g. screws, rods</td>
<td>Yes/no</td>
<td>— “ —</td>
<td>— “ —</td>
</tr>
<tr>
<td>Symptomatic non-union</td>
<td>Pain related to failure of bony consolidation of bridge/union at least 12 months after surgery</td>
<td>Yes/no</td>
<td>— “ —</td>
<td>— “ —</td>
</tr>
<tr>
<td>Neuro-compression</td>
<td>Compression of neural structures with or without neurological deficits</td>
<td>Yes/no</td>
<td>— “ —</td>
<td>— “ —</td>
</tr>
<tr>
<td>Postoperative infection</td>
<td>Superficial or deep (subfascial) wound/ tissue infection after surgery</td>
<td>Yes/no</td>
<td>— “ —</td>
<td>— “ —</td>
</tr>
<tr>
<td>Implant malposition</td>
<td>Incorrect position of the implant</td>
<td>Yes/no</td>
<td>— “ —</td>
<td>— “ —</td>
</tr>
<tr>
<td>Implant failure</td>
<td>Problem due to an implant, e.g. loosening, breakage</td>
<td>Yes/no</td>
<td>— “ —</td>
<td>— “ —</td>
</tr>
<tr>
<td>Wrong site surgery</td>
<td>Unintentional intervention at the wrong level/site, not at level of main pathology</td>
<td>Yes/no</td>
<td>— “ —</td>
<td>— “ —</td>
</tr>
<tr>
<td>Sagittal imbalance</td>
<td>Sagittal malalignment of the spine</td>
<td>Yes/no</td>
<td>— “ —</td>
<td>— “ —</td>
</tr>
<tr>
<td>CSF leakage</td>
<td>Including CSF fistula, pseudomeningocele, etc.</td>
<td>Yes/no</td>
<td>— “ —</td>
<td>— “ —</td>
</tr>
<tr>
<td>Epidural hematoma</td>
<td>Bleeding hematoma outside dural sac but inside bony spinal canal (with or without neuro-compression)</td>
<td>Yes/no</td>
<td>— “ —</td>
<td>— “ —</td>
</tr>
<tr>
<td>Other</td>
<td>State reason for reoperation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Complication definitions modified from the Spine Tango Registry.

b Definition provided is designed for providers; a modified definition will be included on patient questionnaires.

c Reoperation definition and definitions for causes of reoperation modified from the Spine Tango Registry.
Table 3. Risk factors and initial conditions

<table>
<thead>
<tr>
<th>Categories and metrics</th>
<th>Definition/wording</th>
<th>Answer options for capture</th>
<th>Time frame for capture</th>
<th>Reported by</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Date of birth</td>
<td>Baseline</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Male/female</td>
<td>Baseline</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td>Education level</td>
<td>Please indicate your highest level of schooling completed a</td>
<td>Baseline</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td><strong>Baseline clinical status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glassman criteria</td>
<td>Symptoms</td>
<td>Baseline and at time of any intervention</td>
<td>Provider</td>
<td></td>
</tr>
<tr>
<td>Structural pathology</td>
<td>Back pain dominant (acute), leg pain dominant (acute), back pain = leg pain (acute), leg pain dominant (chronic), back pain = leg pain (chronic), neurogenic claudication, cauda equina syndrome</td>
<td>Baseline and at time of any intervention</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td><strong>Compressive pathology</strong></td>
<td>No study or interpretation available, no clinically relevant compression, central compression, lateral compression, combined central and lateral compression, recurrent compression following surgery at the same level</td>
<td>Baseline and at time of any intervention</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td>Pain</td>
<td>How would you rate your average back pain over the last week?</td>
<td>Baseline</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td>Duration of symptoms</td>
<td>How long have you had your current back pain?</td>
<td>Baseline</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td>Smoking habits</td>
<td>Do you smoke?</td>
<td>Baseline</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>Indicate the patient's height</td>
<td>Baseline</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td>Comorbidities</td>
<td>Indicate whether you have been diagnosed with each of the following conditions</td>
<td>Baseline</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td>ASA score (surgical patients only)</td>
<td>Physical Status Classification System 1. healthy, 2. mild/moderate, 3. severe, 4. life threatening, 5. moribund, unknown Paramedian disc herniation, central disc herniation, central spinal stenosis with degenerative listhesis, central spinal stenosis without degenerative listhesis, lateral spinal stenosis, isthmic spondylolisthesis/spondyloïdysthesis, segmental pain (with or without degenerative listhesis, degenerative scoliosis, other</td>
<td>At time of operation</td>
<td>Provider</td>
<td></td>
</tr>
<tr>
<td>Surgical indication</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline functional status</td>
<td>ODI</td>
<td>Baseline</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td>Quality of life</td>
<td>EQ-5D-3L</td>
<td>Baseline</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td>Work status</td>
<td>What is your current work status?</td>
<td>Baseline</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td>Duration of sick leave (if applicable)</td>
<td>Are you working at a physically less demanding job now because of your back and/or leg pain?</td>
<td>Baseline</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td>Previous treatment</td>
<td>Specify any previous procedure(s) and the level(s) by ticking one or several of the boxes below</td>
<td>Baseline</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Injection therapy</td>
<td>Have you previously received spinal injections for your current symptoms?</td>
<td>Baseline</td>
<td>Patient</td>
<td></td>
</tr>
<tr>
<td>Report</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a Level of schooling using culture-specific definitions per ISCED (International Standard of Schooling Classification, UNESCO).*
Similarly, a patient’s baseline functional status is delineated through initial PROM collection, i.e. by measuring disability, HRQOL, work status, and (when applicable) duration of sick leave.

Finally, the working group felt strongly that information on previous treatments is essential for accurate risk adjustment, and selected previous surgery and injection therapy for collection at baseline (Table 3 and Figure 1A), as history of each of these has been shown to be prognostic of subsequent treatment outcomes (Herno 1995, Lee et al. 2010, MacVicar et al. 2013, Mandel et al. 2013). Stratification of previous operations as either discectomy, decompression, or fusion was deemed to be adequately simple for data collection purposes while being sufficiently detailed for risk adjustment. Additionally, while technically a process metric, the working group recommends that providers record the types and levels of surgeries and injections performed at the time of intervention to further facilitate risk stratification (Figure 1B). Again, this level of detail is intended to be as brief as possible in order to streamline data collection while simultaneously allowing meaningful risk adjustment.

**“Index events” and time frame of follow-up**

Regarding the timing of data collection, we elected to establish follow-up at 6 months, 1 year, and 2 years after initiating treatment (Table 1 and Figure 2). Additional follow-up points at 3 months and 5 years were recommended, though not mandatory, as the former is probably meaningful in the management of nonoperatively-treated patients but less so for surgically-affected patients; and for the latter, the contrary is usually true. To simplify data collection and improve compliance, we decided to record complications only following index operations and not after reoperations, which would complicate the follow-up process substantially.

Index events, a term adopted from the SweSpine Registry (Strömqvist et al. 2013), are points in the course of care that should trigger the follow-up schedule to be reset. The initiation of treatment for any new condition, whether managed surgically or not, clearly constitutes an index event. Reoperation for management of a complication or failure to attain the therapeutic goals of an initial surgery is not an index event. However, surgery for a new diagnosis or at a new vertebral level is considered to be a new index event and should cause follow-up, including all measurement of PROMs, to be reset (Figure 3). At that point, the follow-up schedule started after the initial index event is discontinued, as it is not practical to simultaneously conduct 2 follow-up schedules for a single patient.

**Discussion**

We present a standard set of outcome metrics for use in clinical practice for assessing the management of degenerative low back conditions based on the existing literature and on international expert opinion. The set includes patient-reported information on physical function, HRQOL, pain, and work status as well as complications of treatment and baseline characteristics to facilitate risk adjustment.

Several registries of spine patients already exist, tracking tens of thousands of patients in numerous countries (Röder et al. 2005, McGirt et al. 2013, Strömqvist et al. 2013). While these undertakings have been beneficial in many regards, including providing descriptive information about spine care at the population level and answering research questions involving comparisons of various interventions, broader international comparisons have been limited because each existing registry has developed its own metrics for gauging outcomes and definitions for categorizing specific diseases and associated risk factors. Furthermore, registries often do not capture the complete patient population for various diseases because...
used. Furthermore, while the PROMIS instrument offers great
such as the timeline of patient follow-up, the specific adverse
decisions to the discretion of future researchers—
comprehensive to launch into clinical practice, as it leaves
by the NIH task force, the work of the latter is not sufficiently
the domains chosen by our working group and those selected
outcome measurement information system (PROMIS) instru-
reporting outcomes and largely relying on the patient-reported
come set for use in chronic LBP research centered on patient-
Institutes of Health (NIH) Pain Consortium described an out-
for standardized outcome measurement in low back pain
research, but not specifically for use in everyday clinical prac-
time: Risk factors, descriptors of clinical condition, and when applicable, surgical indica-
tion, in-hospital complications, type and level of procedure
Follow-up: 6 months, 1 year, and 2 years
Second and following lumbar surgeries
An index event if
1. Operation is on a different level than index surgery, regardless of diagnosis
2. Operation is on the same level as prior surgery, but for a different diagnosis
Follow-up: 6 months, 1 year, and 2 years
Follow-up for prior index is discontinued

Considered reoperation (not an index event)
1. Operation is on the same level for the same diagnosis as index event
2. Operation is on the same level as index event due to a complication
3. Operation is on another level but due to complication from index surgery
Data captured at time of reoperation: Risk factors, descriptors of clinical condition, cause of reoperation, in-hospital complications, type and level of procedure
Follow-up: Continues as planned from index surgery

Figure 3. A classification scheme to define interventions as either index events or reoperations.

most, but not all (Kessler et al. 2011), spine registries do not follow LBP patients who are managed nonoperatively. This limitation precludes complete comparisons of all available treatment options. Moreover, existing registries often do not capture the entire cycle of care but instead tend to focus on the course of surgical care.

The proposal we present aims to overcome these shortcomings by establishing a standard terminology for measuring outcomes in LBP patients, largely based on well-validated tools that are available in numerous languages. This specific outcome set is particularly well suited to facilitating meaningful comparisons between providers, because it stratifies patients by disease and includes the entire patient population associated with a given diagnosis throughout the full course of their care.

Several recommendations have previously been published for standardized outcome measurement in low back pain research, but not specifically for use in everyday clinical practice (Deyo et al. 1998, Pincus et al. 2008, Deyo et al. 2014). Most recently, a research task force chartered by the National Institutes of Health (NIH) Pain Consortium described an outcome set for use in chronic LBP research centered on patient-reported outcomes and largely relying on the patient-reported outcome measurement information system (PROMIS) instrument (Deyo et al. 2014). While there is substantial overlap in the domains chosen by our working group and those selected by the NIH task force, the work of the latter is not sufficiently comprehensive to launch into clinical practice, as it leaves several decisions to the discretion of future researchers—such as the timeline of patient follow-up, the specific adverse events to be recorded, and even which PROM tools should be used. Furthermore, while the PROMIS instrument offers great and variables that we have chosen (Pincus et al. 2008). While reasonable for research purposes, the MMICS outcome set and especially the associated timeline for collecting data would be overly burdensome for use in clinical practice. Another commendable effort was previously described as part of the Multinational Musculoskeletal Inception Cohort Study Collaboration (MMICS), which again showed substantial overlap with the domains

potential efficiency through computerized adaptive testing and may eventually become favorable to the PROM tools recommended here, it is not yet broadly translated and validated beyond English (NIH PROMIS Website 2014) and is not therefore ready for international use.

A similar effort is currently being conducted by the “International Steering Committee for the Core Outcome Set for Low Back Pain” (Chiarotto et al. 2014). Initial findings presented at the “Core Outcome Measures in Effectiveness Trials” (COMET) meeting in November, 2014 prioritized 3 domains identical to those chosen by our working group: physical function, pain intensity, and HRQOL, with work ability ranked fourth. While useful for guiding researchers who are developing LBP outcome measures, these recommendations are not detailed enough for use in clinical practice. Another commendable effort was previously described as part of the Multinational Musculoskeletal Inception Cohort Study Collaboration (MMICS), which again showed substantial overlap with the domains

Looking forward, revisions to this outcome set will be needed. For example, computerized adaptive testing may provide efficiency gains in PROM collection as software development progresses (Fries et al. 2005). ICHOM and representatives of the working group plan to actively monitor use of this set through a steering committee comprised of representatives from several existing outcome measurement efforts. Their work will involve communication with users, including collection of direct feedback. The frustrations and innovations of providers using this set will be crucial for its improvement, and structured revisions to the set will be reviewed on an annual basis. The steering committee will also be available to communicate with relevant third parties. For example, some instruments recommended in this outcome set, such as the
EQ-5D, are proprietary and continued inclusion in the set may depend on future negotiated agreements.

Our work has a number of limitations that should be mentioned. Firstly, the proposed outcome set remains untested, and while it is largely based on existing tools and familiar data points, the specifics of survey circulation and the associated time frames for data collection and reporting will inevitably lead to bumps in the road. Secondly, despite our best efforts to generate diverse international consensus, our outcome set is surely not equally applicable to all cultures. Much work in linguistic and cross-cultural validation remains. Still, we feel that our work represents a sufficient and important starting point. Thirdly, much work remains to be done on the practical issues of compiling and analyzing data, ultimately building robust risk-adjustment models with appropriate quality assurance to produce reports that accurately reflect provider performance while simultaneously protecting patient privacy. This will be particularly important if value-based reimbursement does indeed come to fruition. For example, in Sweden there are ongoing efforts in conjunction with the SweSpine Registry to link reimbursement levels to postoperative patient-reported outcomes. ICHOM intends to continue its facilitative role to guide development of such models and their inclusion in quality reporting initiatives.

In summary, the members of the working group feel that the introduction of this set of outcomes for the treatment of degenerative low back pathology is an essential step toward an international spine community that routinely measures and reports its performance in common and meaningful ways. We invite all providers caring for patients with low back pain to join us in measuring this set; the full list of metrics, contact information, and other resources to facilitate implementation are available on the ICHOM website (ICHOM Website 2014a).

RCC: literature review, study design, data collection, data analysis, interpretation of data, writing, preparation of figures and tables, revisions, and approval of final work. AW: literature review, study design, data collection, data analysis, interpretation of data, preparation of figures and tables, revisions, and approval of final work. PF: literature review, study design, data collection, data analysis, interpretation of data, revisions, and approval of final work. RCC: literature review, study design, data collection, data analysis, interpretation of data, revisions, and approval of final work. PF: literature review, study design, data collection, data analysis, interpretation of data, revisions, and approval of final work.


