Chapter 06

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


Published in: Acta Orthop. 2015;86(5)523-33

Appendix to Chapter 06 - editorial comment
(after Chapter 06 - page 130-133)

Guest editorial: Spinal disorders, quality-based healthcare and spinal registers
Fairbank JCT
Published in: Acta Orthop. 2015; 86 (5): 521–522
Abstract

Background & Purpose: Outcome measurement has been shown to improve performance in several fields of healthcare. This understanding has driven a growing interest in value-based healthcare, where value is defined as outcomes achieved per money spent. While low back pain (LBP) constitutes an enormous burden of disease, no universal set of metrics has yet been accepted to measure and compare outcomes. Here, we aim to define such a set.

Materials & Methods: An international group of 22 specialists in multiple disciplines of spine care was assembled to review literature and select LBP outcome metrics through a 6-round modified Delphi process. The scope of the outcome set was degenerative lumbar conditions.

Results: Patient-reported metrics include a numeric pain scales, lumbar-related function using the Oswestry Disability Index, health-related quality of life using the EQ-5D-3L questionnaire, and questions assessing work status and analgesic use. Specific common and serious complications were included. Recommended follow-up intervals include 6, 12 and 24 months after initiating treatment, with optional follow-up at 3 months and 5 years. Metrics for risk stratification were selected based on pre-existing tools.

Interpretation: The outcome measures recommended here are structured around specific etiologies of LBP, span a patient’s entire cycle of care, and allow for risk adjustment. Thus, when implemented, this set can be expected to facilitate meaningful comparisons and ultimately to provide a continuous feedback loop, enabling ongoing improvements in quality of care. Much work lies ahead in implementation, revision, and validation of this set, but it represents an essential first step toward establishing a community of LBP providers focused on maximizing the value of care we deliver.
**Introduction**

Measuring outcomes in healthcare has well-documented benefits as well as challenges [1,2]. Simply asking providers to report their outcomes has been shown to improve performance [3]. Additionally, understanding one’s results empowers a provider to continuously learn from and refine the care he or she delivers [4]. 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, allowing patients to make intelligent choices about where to seek care [4]. 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). This concept of ‘value-based healthcare’ has been gaining attention both throughout the medical field [1,5] and specifically within the spine community [6,7]. With evolving reimbursement systems in many countries, it is also conceivable that there will be growing 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 in the population and constitutes a major component of the global burden of disease [8]. Measuring outcomes in the field of low back pain is challenging. Numerous disease states affect the low back, resulting in low back pain, leg pain or both; to compare outcomes, patients must be accurately stratified by both diagnosis and severity. Moreover, existing treatment algorithms are complex and often controversial, including both operative and non-operative options and frequently requiring multidisciplinary provider teams. Additionally, low back pain rarely causes death or other objective end points, so outcomes are best measured by patient-reported metrics, which are inherently subjective and require thorough psychometric testing.

A substantial amount of work has already been done in the field of low back pain and well-validated tools exist for measuring disease-specific outcomes [9]. Similarly, several large registries are already in existence collecting outcomes, along with many other data points [10-13]. Previous consensus-based efforts have been made to define sets of outcome measures or domains for research purposes [14-16]. 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 ‘working 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 [17]. The working group's efforts were coordinated by a core 'project Team' consisting of a working group lead (PF), a project leader (AW), a research fellow (RC), and the ICHOM vice president of research and development (CS).

The project was structured as a modified Delphi process [18] involving 6 teleconferences held between June and November of 2013. The goals of these calls were choosing inclusion and exclusion criteria for the relevant patient population, selecting and defining outcome metrics, and identifying 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 among spine registries, 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 discussion held during the teleconference. Surveys were circulated by email following teleconferences to all working group members 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 vigorous debate, the issue was revisited by the project team and a new proposal was presented to the working group for consideration.

Several recurrent themes emerged throughout this process, and developed into guiding principles for the group's collaboration. First, 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. Second, 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 predominating 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 of all the experts and co-authors.
Results

Response rates for the 5 working group 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 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 the vast majority of all of lumbar pathology [22]. Other areas of spine care involve different patient populations, treatment approaches, and outcomes - and should be addressed in the future by analogous condition-specific outcome sets. Formal inclusion criteria selected by the working group consist 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 non-specific pain). The relevant corresponding exclusion criteria include spinal infection, tumor, fracture, traumatic dislocation, congenital or idiopathic scoliosis, and age under 18 years.

Outcome domains (Table 6.1)

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 [19]. 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 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’ [19] - were excluded from the set, as historically they have been studied with inconsistent definitions [19] and they are probably reflected in other domains such as HRQOL.
### Table 6.1 Patient-reported outcome measures

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Measurement Tool</th>
<th>Definition/Wording</th>
<th>Answer options</th>
<th>Timeframe 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 Personal Care (washing, dressing, etc.) Lifting Walking Sitting Standing Sleeping Sex Life (if applicable) Social life Traveling</td>
<td>6 options for each domain ranging from no problem to severe impairment (see appendix)</td>
<td>Baseline, index event(s), 6 months, 1 year, 2 years</td>
</tr>
<tr>
<td><strong>Quality of Life</strong></td>
<td>EQ5D-3L</td>
<td>Mobility Self-Care Usual Activities Pain/Discomfort Anxiety/Depression</td>
<td>3 options for each domain ranging from no problem to severe impairment (see appendix)</td>
<td>Baseline, index event(s), 6 months, 1 year, 2 years</td>
</tr>
<tr>
<td></td>
<td>EQ-VAS</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></td>
</tr>
<tr>
<td><strong>Work Status</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><strong>Analgesic Use</strong></td>
<td>Do you take non-narcotic pain relieving medication or tablets for your back problems?</td>
<td>Yes regularly, Yes sometimes, No</td>
<td>Baseline, index event(s), 6 months, 1 year, 2 years</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Do you take narcotic pain relieving medication or tablets for your back problems?</td>
<td>Yes regularly, Yes sometimes, No</td>
<td>Baseline, index event(s), 6 months, 1 year, 2 years</td>
<td></td>
</tr>
</tbody>
</table>
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) [20].

Common and well validated methods for measuring pain include the numeric rating scale (NRS) and the visual analog scale (VAS) [19,21], and the major existing spine registries are divided between those options [10-12]. While there is no gold standard, a VAS allows patients to provide a more specific response while a NRS is typically 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 [21-24]. This option was chosen by the working group (with 21 of 22 members in agreement) for inclusion in the standard set, both for back and leg pain individually.

Numerous tools have been studied for measuring lumbar-related function in patients with low back pathology [9,19]. The ODI is the most commonly used and cited tool for this purpose, followed by the Roland Morris Disability Questionnaire (RMDQ) [19] and the Core Outcome Measures Index (COMI) [25]. 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 [19]. We also felt the ODI to be the most feasible to implement as it is validated in 14 languages [as opposed to 9 for each of the RMDQ [19] and COMI [23]] and is relatively short [10 items as opposed to 24 in the RMDQ [26] and 7 in the COMI [27]]. All are free with online registration being required for use of the ODI [19]. For these 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 exist [28-32], 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 [19]. The SF-36 has been shown to be valid, reliable and responsive in this population, while the NHP and SF-12 have been proven valid and reliable [19]. To our knowledge, these have not been studied for responsiveness and none of the psychometric properties of the EQ-5D have yet been examined among LBP patients. However, the EQ-5D tool has an excellent track record among other demographics as well as the general population [33,34], and has been shown to correlate well with the ODI in LBP patients [35]. 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) [19]. The EuroQol tool is also inexpensive [36] relative to the SF tools [37,38], 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 [19]. 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 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 [10]; the wording was modified slightly by the working group.

Complications and adverse events
Adverse consequences of treatment, e.g. invasive procedures, comprise 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 surgeries and injection therapy, and for convenience, the same list of complications and timeframe 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 6.2). In regions where reliable administrative death records are readily accessible, the working group recommends the use of such administrative 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 [39,40]. In countries and practices with reliable administrative documentation of re-hospitalization, such as electronic medical records or insurance databases, the working group recommends using this 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 6.3). 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 [41]. 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. [42] were selected, primarily for its applicability to both operative and conservatively-treated patients (Table 6.3). 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 is 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 current registries, the set of operative indications used by the Swespine registry [12] 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). Additionally, the American Society of Anesthesiologists (ASA) Physical Status Classification System has been shown to be prognostic for many surgical procedures [43-45] and the working group felt 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 been the basis for risk adjustment in large patient populations. Patient-reported responses to the Charlson comorbidity Index [46] have been proven predictive of both mortality and various PROMs [47,48]. 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 [49]. This set was augmented with 2 conditions included in the Charlson index that the working group considered particularly prescient in the LBP population: paraplegia/hemiplegia and HIV/AIDS (Table 6.3). Smoking habits [50,51] and BMI [52-54] 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 [55-57] - was discussed at length, and the working group concluded 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 likely 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.

Similarly, a patient's baseline functional status is delineated through initial PROMs 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 6.3 and Figure6.1A), as history of each of these has been shown to be prognostic for subsequent treatment outcomes [58-61]. 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 providers record the types and levels of surgeries and injections performed at the time of intervention to further facilitate risk stratification (Figure 6.1B). 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 timeframe 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 6.1 and Figure 6.2). 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 non-operatively-treated patients but less so for surgical 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 [12], 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 6.3). 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.
Figure 6.2. The recommended timeline for collection of each outcome measure

<table>
<thead>
<tr>
<th>Data points</th>
<th>Baseline</th>
<th>Index period</th>
<th>6 mts</th>
<th>1 yr</th>
<th>2 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROMs for pain, disability and Q.O.L.</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Work status</td>
<td>△</td>
<td></td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Time to return to work</td>
<td>△</td>
<td></td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Continuous oral analgesic use</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Mortality</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Need for reoperation</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Cause of reoperation</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Need for rehospitalisation</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Complications: Nerve root injury, Vascular injury, Dural tear, Other complications</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Complications: Deep wound infection, Pulmonary embolus</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Risk factors</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Descriptors of clinical condition</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Type and level of surgery and surgical indication</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
</tbody>
</table>

- □ Patient-reported
- □ Physician-reported
- □ Administratively-reported (when available)

1. Risk factors and descriptors of clinical condition reported pre-intervention, complications and mortality reported at discharge
2. Collection also recommended at 5 years, but only deemed mandatory as displayed

Figure 6.3. A classification scheme to define interventions as either index events or reoperations

Initiation of non-surgical treatment or 1st lumbar surgery
(Always an index event)

Data captured at this time: Risk factors, descriptors of clinical condition, and when applicable, surgical indication, in-hospital complications, type and level of procedure

(Follow up: 6 mos, 1 yr, and 2 yrs)

2nd 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 mos, 1 yr, and 2 yrs
(Follow up for prior index is discontinued)

Considered reoperation (not an index event) if
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
### Table 6.2 Adverse outcomes of treatment

<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><strong>Mortality</strong></td>
<td>Death in-hospital (all-cause mortality)</td>
<td>Yes/No</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nerve root injury</strong> (including cauda equina syndrome)</td>
<td>Iatrogenic nerve root damage</td>
<td>Yes/No</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vascular injury</strong></td>
<td>Clinically significant iatrogenic damage to a vessel</td>
<td>Yes/No</td>
<td>While in-house for procedure</td>
<td></td>
</tr>
<tr>
<td><strong>Dural tear</strong></td>
<td>Iatrogenic damage of the dura with liquor emission</td>
<td>Yes/No</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>(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><strong>Deep wound infection</strong> 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><strong>Pulmonary embolus</strong> c</td>
<td>PE diagnosed by radiologic study after the intervention</td>
<td>Yes/No</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rehospitalization</strong></td>
<td>Were you admitted to an acute care facility as an in-patient 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><strong>Need for reoperation</strong> (if yes, specify cause)</td>
<td>Second or multiple performed interventions caused by complications after index surgery, not planned in advance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hardware removal</strong></td>
<td>Removal of implants: e.g. screws, rods</td>
<td>Yes/No</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Symptomatic non-union</strong></td>
<td>Pain related to failure of bony consolidation of bridge/union at minimum 12 months after surgery</td>
<td>Yes/No</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Neuro-compression</strong></td>
<td>Compression of neural structures with or without neurological deficits</td>
<td>Yes/No</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Postoperative infection</strong></td>
<td>Superficial or deep (subfascial) wound/tissue infection after surgery</td>
<td>Yes/No</td>
<td>At time of reoperation</td>
<td>Provider</td>
</tr>
<tr>
<td><strong>Implant malposition</strong></td>
<td>Incorrect position of the implant</td>
<td>Yes/No</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Implant failure</strong></td>
<td>Problem due to an implant e.g., loosening, breakage</td>
<td>Yes/No</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Wrong site surgery</strong></td>
<td>Unintentional intervention on the wrong level/site, not on level of main pathology</td>
<td>Yes/No</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sagittal imbalance</strong></td>
<td>Sagittal malalignment of the spine</td>
<td>Yes/No</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CSF leakage</strong></td>
<td>Including CSF fistula, pseudomeningocele, etc.</td>
<td>Yes/No</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Epidural hematoma</strong></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><strong>Other</strong></td>
<td>State reason for reoperation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

*a* Complication definitions modified from 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 Spine Tango registry
### Table 6.3 Risk factors and Initial conditions

<table>
<thead>
<tr>
<th>Categories and metrics</th>
<th>Definition/wording</th>
<th>Answer options</th>
<th>Timeframe for capture</th>
<th>Reported by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Date of birth</td>
<td>dd/mm/yyyy</td>
<td>Baseline</td>
<td>Patient</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td>Male/female</td>
<td>Baseline</td>
<td>Patient</td>
</tr>
<tr>
<td>Education level</td>
<td>Please indicate your highest level of schooling completed*</td>
<td>None, primary, secondary, tertiary</td>
<td>Baseline</td>
<td>Patient</td>
</tr>
<tr>
<td>Baseline clinical status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glassman criteria</td>
<td>Symptoms</td>
<td>Back pain dominant (acute), leg pain dominant (acute) back pain = leg pain (acute), back pain dominant (chronic), 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>Provider</td>
</tr>
<tr>
<td>Structural pathology</td>
<td>No study or interpretation available, age appropriate, disc pathology with normal height, disc space collapse, spondylysis/ spondylolisthesis, scoliosis/kyphosis, facet pathology, non-union</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compressive pathology</td>
<td>No study or interpretation available, no clinically relevant compression, central compression, lateral compression, combined central &amp; lateral compression, recurrent compression following surgery at the same level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain</td>
<td>How would you rate your average back pain over the last week?</td>
<td>1-10, as per Table 1</td>
<td>Baseline</td>
<td>Patient</td>
</tr>
<tr>
<td></td>
<td>How would you rate your average leg pain over the last week?</td>
<td>1-10, as per Table 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of symptoms</td>
<td>How long have you had your current back pain?</td>
<td>I don’t have back pain, &lt; 3 months, 3-12 months, 1-2 years, &gt; 2 years</td>
<td>Baseline</td>
<td>Patient</td>
</tr>
<tr>
<td></td>
<td>How long have you had pain radiating to your leg(s)?</td>
<td>I don’t have pain radiating to my legs, &lt; 3 months, 3-12 months, 1-2 years, &gt; 2 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking habits</td>
<td>Do you smoke?</td>
<td>Yes/no</td>
<td>Baseline</td>
<td>Patient</td>
</tr>
<tr>
<td>BMI</td>
<td>Indicate the patient’s height</td>
<td>Measured in cm</td>
<td>Baseline</td>
<td>Provider</td>
</tr>
<tr>
<td></td>
<td>Indicate the patient’s mass</td>
<td>Measured in kg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comorbidities</td>
<td>Indicate if you have been diagnosed with each of the following conditions</td>
<td>Heart disease, hypertension, poor circulation, lung disease, diabetes, kidney disease, liver disease, nervous system disease, cancer, depression, arthritis, peptic ulcer disease, hemiplegia/paraplegia, AIDS</td>
<td>Baseline</td>
<td>Patient</td>
</tr>
<tr>
<td>ASA score (surgical patients only)</td>
<td>Physical Status Classification System</td>
<td>1. healthy, 2. mild/moderate, 3. severe, 4. life threatening, 5. moribund, unknown</td>
<td>At time of operation</td>
<td>Provider</td>
</tr>
<tr>
<td>Surgical indication</td>
<td>Paramedian disc herniation, central disc herniation, central spinal stenosis with degenerative llisthesis, central spinal stenosis without degenerative llisthesis, lateral spinal stenosis, isthmic spondylysis/spondylolisthesis, segmental pain (with or without degenerative llisthesis, degenerative scoliosis, other</td>
<td>At time of operation</td>
<td>Provider</td>
<td></td>
</tr>
</tbody>
</table>
**Baseline functional status**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Description</th>
<th>Baseline</th>
<th>Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disability</td>
<td>ODI</td>
<td>As per Table 1</td>
<td>Baseline</td>
</tr>
<tr>
<td>Quality of Life</td>
<td>EQ-5D-3L</td>
<td>As per Table 1</td>
<td>Baseline</td>
</tr>
<tr>
<td>Work status</td>
<td>What is your current work status?</td>
<td>As per Table 1</td>
<td>Baseline</td>
</tr>
<tr>
<td></td>
<td>Are you working at a physically less demanding job now because of your back and/or leg pain?</td>
<td>As per Table 1</td>
<td>Baseline</td>
</tr>
<tr>
<td>Duration of sick leave (if applicable)</td>
<td>Are you currently on sick leave from work?</td>
<td>Yes, full time for my back problems, Yes part-time for my back problems, Yes due to another disease, No</td>
<td>Baseline</td>
</tr>
<tr>
<td></td>
<td>If yes, for how long?</td>
<td>1 week or less, 1-4 weeks, 1-3 months, 3-6 months, 6-9 months, 9-12 months, 1-2 years, &gt; 2 years</td>
<td>Baseline</td>
</tr>
</tbody>
</table>

**Prior treatment**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Description</th>
<th>Baseline</th>
<th>Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgery</td>
<td>Please specify any prior procedure(s) and the level(s) by ticking one or several of the boxes below</td>
<td>See Figure 3</td>
<td>Baseline</td>
</tr>
<tr>
<td>Injection therapy</td>
<td>Have you previously received spinal injections for your current symptoms? (e.g. epidurals, specific nerve root injections, facet injections or discograms)</td>
<td>Yes/no</td>
<td>Baseline</td>
</tr>
</tbody>
</table>

*Level of schooling using culture-specific definitions per ISCED (International Standard of Schooling Classification, UNESCO) [40]*

**Discussion**

We present a standard set of clinical outcome metrics for the 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 [10-12]. While these undertakings have been beneficial in many regards, including providing descriptive information about spine care at a 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 most, but not all [62], spine registries do not follow LBP patients who are managed non-operatively. 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 [14,15,63]. Most recently, a research task force chartered by the National Institute 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 [63]. 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 adequately 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 PROMs tools should be used. Furthermore, while the PROMIS instrument offers great potential efficiency through computer adaptive testing and may eventually become favorable to the legacy measures recommended here, it is not yet broadly translated and validated beyond English [64] and is therefore not 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’ [16]. 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 and variables chosen by our working group [15]. While reasonable for research efforts, the MMICS outcome set and especially the associated timeline for collecting data would be overly burdensome for ongoing use in non-research settings.

With the working group’s goal fulfilled, including a complete set of outcomes and associated data defined; the focus can be shifted to implementation. For providers, group practices, and registries that ascribe to the benefits of outcome measurement, this set will be available for voluntary adoption. Practices with existing data collection processes may be able to incorporate the standard set into ongoing workflows in ways that will minimize additional work after an initial learning curve. Other organizations may need to begin by establishing an infrastructure for prospective data collection. ICHOM is committed to facilitating broad adoption of this set and has made the full recommendations of this group freely available on its website, along with a reference guide to assist with technical aspects of implementation [65]. Looking forward, revisions to this outcome set will be needed. For example, computerized adaptive testing may provide efficiency gains in PROM collection as software progresses [66]. ICHOM and representatives from 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 employing this set will be
crucial to 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 be dependent 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 timeframes 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 surely is 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 [65].
References

18. Pili J. The Delphi Method: Substance, context, a critique and an annotated bibliography. 2nd ed. 1971; p.57-71
40. Axon RN, Williams MV. Hospital readmission as an accountability measure. JAMA 2011;305:504-505

Proposed set of metrics for LBP
Guest editorial: Spinal disorders, quality-based healthcare and spinal registers

Jeremy Fairbank
Professor of Spine Surgery, University of Oxford, UK

In this issue of Acta Orthopaedica 2 articles [1,2] represent waymarks on a road towards a conception of universal value-based healthcare for spine. This goal is expected to be a convergence of the interests of patients, payers, politicians and clinicians. The articles are focused on the painful lumbar spine, which represent the top ranking chronic healthcare complaint [3]. Clinicians’ interest in quality dates back to Florence Nightingale in the Crimea. For patients it goes back to time immemorial. For economists it has become an issue in the last 2 decades. All politicians should be interested in this topic, and this is kindled by public dissatisfaction and the rising costs of healthcare.

ICHOM (International Consortium of Health Outcomes Measurement, www.ichom.org/) is an organisation recently founded by the Institute for Strategy and Competitiveness at Harvard Business School, the Boston Consulting Group, and the Karolinska Institute to enable the shift towards value-based health care. The concept of value-based health care has been described by Michael Porter, a co-founder of ICHOM, as the only strategy that will fix health care [4]. The central point of his work is the need to look at outcomes and cost together in driving clinical improvement and policy.

Health registers play a central role in the measurement of outcomes and thereby enable the shift towards a value-based system: the Swedish Knee Arthroplasty Register (start 1975) (www.knee.se) and the Swedish Hip Arthroplasty Register (start 1979) (www.shpr.se), being the first and most notable in the musculoskeletal world. In 2012 ICHOM identified 4 areas of healthcare to define international standard sets of outcome measures. One of these was low back pain, and one output of this process is the paper by Clement et al. [1]. There have been earlier attempts to identify optimum outcome measures for research [5,6], but this is the first to search for an international consensus on quality measures for use in daily clinical practice. ICHOM is in the process of extending this exercise, which involves a defined methodology, to many other areas of healthcare: it has now completed 12 ‘standard sets’ (of outcome measures), and plans to have 50 completed by 2017.

Since the establishment of the Swedish Spine Register (SweSpine, www.swespine.se) in 1992, spine registers have sprung up in other countries. The systematic review by van Hooff et al. [2] is a first attempt to see if these registers are influencing quality. Whilst the case remains unproven, they are able to cite a number of examples where it would appear that a register has altered behaviour and improved quality. I do not doubt that, as this movement evolves, so we shall see better evidence of impact. This can only be good from the patient perspective, but Registers do not yet to have the capacity to answer the universal question as to which doctor is likely to deliver an individual the optimum health care. The register should be able to confirm whether a given provider is an outlier nationally (and if ICHOM is successful) internationally. The ambitions of ICHOM are much grander: it is their intention that health care systems should reimburse providers on the basis of quality rather than quantity. This is beginning to happen
in Sweden, but I believe it will be some time before such systems evolve in other healthcare economies. There are early signs of this movement developing in my country.

There is devil in the detail. Table 5.3 in van Hooff et al. gives a detailed list of recommendations for enhancing spine registers. Register models can only work with patient involvement and particularly the completion of follow-up questionnaires has to be high. Van Hooff et al. recommend 60 to 80% compliance. The reported completion rate from the National British Spine Register is only 20%. Subtle ‘sticks and carrots’ need to be developed to encourage patient completion. It means the forms have to be a short and clear. More qualitative research is needed to explore these issues. Consideration may be needed to offering adequate reimbursement for the time and trouble of completing outcome measures [7].

The outcome measures themselves require care and attention. Most have evolved since the 1970s. Some like EQ-5D, are managed by committees. Others like ODI, are managed by their original authors. Others, such as NRS pain, are essentially orphans since their original description by Huskisson in the 1970’s [8-10]. ICHOM has a duty to ensure that the measures they recommended are looked after and not subjected to alterations. It is vital that they are translated accurately by a standard protocol. ODI, for example, recommended as part of the ICHOM low back pain outcome set, is now licenced to the MAPI Trust in Lyon, which has pioneered good practice in this regard (www.proqolid.org/instruments/oswestry_disability_index_odi). ICHOM should consider sponsoring orphan instruments (perhaps at MAPI) such as NRS to ensure that they are delivered in as consistent a way as possible.

It must be noted that there are problems with registers and the interpretation of their data. Most are owned and managed by clinicians. So long as this model is transparent, it probably keeps the data safer from meddling than it would be in the hands of governments, who also serve as purchasers. Registers need funding, which also generates problems of long term viability and conflicts of interest [2]. Properly defining criteria of success and failure is an important challenge. The Swedish Knee and Hip Registers and its various international descendants have used revision rates as an important criterion of failure. When this is applied to the knee register, unicompartmental knee replacements are shown to have higher revision rates than total knee replacement, arguably not because they actually fail more frequently, but because surgeons are more ready and willing to revise a unicompartmental knee replacement than a total knee replacement [11]. Unsurprisingly this view is disputed by the directors of knee registers [12], but some complex issues of quality and cost are involved here. I see a parallel situation in the use of interspinous spacers to treat spinal stenosis. 2 RCT’s have shown similar outcomes between spacers and conventional decompressive surgery but with higher revision rates in the spacer group [13,14]. Both these examples are significant because the initial implant cost is high, but they do reflect the important point that defining outcome metrics is potentially treacherous. This matters in many ways, but particularly when reimbursement depends on it. Rigorous metrics defined based on established methodologies, arguably such as those presented in the ICHOM standard set, should present an answer to this challenge.

Those registers that use mortality as an outcome, such as cardiac surgery, may make cardiac surgeons more risk averse so that they avoid high-risk patients [15,16]. Statistical variation becomes a major problem when these results are distilled down to individual surgeon data.
This has provoked problems previously in the US and currently in the UK where SSMD (Surgeon Specific Mortality Data) has recently become mandated. Discussion on this continues [15-19]. The quality movement in healthcare needs the strong support of everyone involved. We need the education and involvement of our public and politicians to make this happen. As van Hooff et al. spell out, registers are important, and need good methodology and design, with care and attention to deliver quality data. How these data are interpreted and presented will need continual scrutiny and innovation as their importance in the health economy increases.

Conflict of interest
Jeremy Fairbank is coauthor of the Clement et al. paper, and is a copyright holder of the Oswestry Disability Index
References


