Clinically important deterioration in patients undergoing lumbar spine surgery: a choice of evaluation methods using the Oswestry Disability Index, 36-Item Short Form Health Survey, and pain scales

Clinical article

JEFFREY L. GUM, M.D.,1 STEVEN D. GLASSMAN, M.D.,1,2 AND LEAH Y. CARREON, M.D., M.Sc.2

1Department of Orthopaedic Surgery, University of Louisville School of Medicine; 2Norton Leatherman Spine Center, Louisville, Kentucky

Object. Health-related quality of life (HRQOL) measures have become the mainstay for outcome appraisal in spine surgery. Clinically meaningful interpretation of HRQOL improvement has centered on the minimum clinically important difference (MCID). The purpose of this study was to calculate clinically important deterioration (CIDET) thresholds and determine a CIDET value for each HRQOL measure for patients undergoing lumbar fusion.

Methods. Seven hundred twenty-two patients (248 males, 127 smokers, mean age 60.8 years) were identified with complete preoperative and 1-year postoperative HRQOLs including the Oswestry Disability Index (ODI), 36-Item Short Form Health Survey (SF-36), and numeric rating scales (0–10) for back and leg pain following primary, instrumented, posterior lumbar fusion. Anchor-based and distribution-based methods were used to calculate CIDET for each HRQOL. Anchor-based methods included change score, change difference, and receiver operating characteristic curve analysis. The Health Transition Item, an independent item of the SF-36, was used as the external anchor. Patients who responded “somewhat worse” and “much worse” were combined and compared with patients responding “about the same.” Distribution-based methods were minimum detectable change and effect size.

Results. Diagnoses included spondylolisthesis (n = 332), scoliosis (n = 54), instability (n = 37), disc pathology (n = 146), and stenosis (n = 153). There was a statistically significant change (p < 0.0001) for each HRQOL measure from preoperatively to 1-year postoperatively. Only 107 patients (15%) reported being “somewhat worse” (n = 81) or “much worse” (n = 26). Calculation methods yielded a range of CIDET values for ODI (0.17–9.06), SF-36 physical component summary (−0.32 to 4.43), back pain (0.02–1.50), and leg pain (0.02–1.50).

Conclusions. A threshold for clinical deterioration was difficult to identify. This may be due to the small number of patients reporting being worse after surgery and the variability across methods to determine CIDET thresholds. Overall, it appears that patients may interpret the absence of change as deterioration.

Key Words • clinically important deterioration • lumbar fusion • outcome measures • health-related quality of life • Oswestry Disability Index • minimum clinically important difference • SF-36

Health-related quality of life (HRQOL) measures have become the mainstay for outcome appraisal in spine surgery over the past two decades. Commonly used, validated, and reliable HRQOL measures include the Medical Outcomes Study 36-Item Short Form Health Survey (SF-36),12,26 Oswestry Disability Index (ODI),7,9 and visual analog scales or numeric rating scales for back and leg pain.17 The definition of improvement has evolved concepts such as minimal clinically important difference (MCID)6 and substantial clinical benefit,11 which represent patient-perceived and clinically meaningful improvement.

In contrast to accepted definitions of improvement, the concept of being worse after lumbar spinal fusion is not well defined. Current descriptions include the occurrence of perioperative or postoperative complications, radiographic nonunion, need for revision surgery, or im-

Abbreviations used in this paper: CIDET = clinically important deterioration; HRQOL = health-related quality of life; HTI = Health Transition Item; MCID = minimal clinically important difference; MCS = mental component summary; ODI = Oswestry Disability Index; PCS = physical component summary; ROC = receiver operating characteristic; SF-36 = 36-Item Short Form Health Survey.
Clinically important deterioration in lumbar spine surgery

With a more significant deformity were tracked in a separate adult deformity outcomes protocol. The instability subgroup included patients with abnormal motion (≥3 mm) on flexion and extension radiographs, substantial retrolisthesis, or lateral listhesis. Patients in the stenosis subgroup included patients with central or foraminal stenosis without any abnormalities listed above, or patients undergoing fusion due to the risk for iatrogenic instability following decompression alone.

Outcome Assessments

Health-related quality of life measures were administered at baseline prior to surgery and 1-year postoperatively and included the ODI,7 SF-36,12,26 numerical rating scale for back pain, and numerical rating scale for leg pain (0–10 rating scales).17 For the ODI and the back and leg pain scores, higher scores denote increasing disability or higher pain levels. For the SF-36 physical component summary (PCS), higher scores imply better functioning. Thus, decreasing scores for the ODI and the back and leg pain scores denote improvement, while decreasing scores on the SF-36 PCS denote deterioration. The 1-year change in scores for the ODI and the back and leg pain scores were defined as baseline scores subtracted from 1-year postoperative scores. The 1-year change in scores for the SF-36 PCS was defined as 1-year postoperative scores subtracted from baseline scores. Thus, in all the HRQOL measures, positive values denote improvement.

The Health Transition Item (HTI), an independent item of the SF-36, was used as an external ad hoc anchor. The HTI is not included in the algorithm to determine domain and composite summary scores of the SF-36. The HTI asks patients to compare their current health to their health 1-year previously. Patients’ responses range from “much worse,” “somewhat worse,” “about the same,” “somewhat better,” and “much better.” The patients who responded “much worse” and “somewhat worse” were grouped into a “worse” cohort and compared with patients responding “about the same.” The change difference value was calculated as the difference in the mean change scores between the “worse” and “about the same” patients.4

A receiver operating characteristic (ROC) curve was created for each of the 4 HRQOL measures. The sensitivity and specificity of the change scores for the “worse” and “about the same” cohorts were plotted. The ideal cut-off value for an ROC curve corresponds to the point of equal trade-off between greatest sensitivity and specificity to distinguish the “worse” from the “about the same” patients.

The SEM was calculated as SD/√1-ś, where SD is the standard deviation of the baseline scores and r is the test-retest reliability coefficient.4,16,20,30 A reliability of 0.90 was used for ODI,19 0.92 for PCS,28 and 0.95 for back and leg pain scales.29 Minimum detectable change was calculated as 1.96√2 × SEM, for a 95% CI.5,8,13,27 A change that corresponds to a small effect is typically considered the minimal change.3,8,14,18,24,25 Effect size–derived CIDET was determined by multiplying the SD of the baseline scores by 0.2 (a small effect size).4,24

**Methods**

**Study Population**

A database of 1154 patients who underwent lumbar spinal fusion from a single spine specialty center (the Norton Leatherman Spine Center) with prospectively collected patient-reported outcomes data was evaluated. The data were retrospectively reviewed with Institutional Review Board approval. A total of 722 patients with complete baseline and 1-year outcome measures following primary, instrumented, posterior lumbar fusion were identified. Patients who had fusion extending into the thoracic spine or that included pelvic fixation were excluded. The patients included had an average age of 60.8 ± 12.1 years with an average body mass index of 30.3 ± 12.1 kg/m². Only 17.6% (n = 127) were smokers and 34.3% (n = 248) were males.

Primary diagnosis included spondylolisthesis (n = 332, 46%), stenosis (n = 153, 21.2%), disc pathology (n = 146, 20.2%), scoliosis (n = 54, 7.5%), and instability (n = 37, 5.1%). Preoperative diagnostic etiology and indication for fusion were classified based on a previously published study by Glassman et al.10 Primary surgical cases were classified as disc pathology, spondylolisthesis, instability, stenosis, or scoliosis. Disc pathology included patients with internal disc desiccation and/or high intensity zones,1 disc space collapse, Modic changes,2,21 or prior intradiscal procedures such as intradiscal electrothermal therapy. The spondylolisthesis group included both isthmic and degenerative listhesis. Patients with spondylolisthesis had an associated coronal plane deformity from 10° to 30° with involvement over a few segments in the lumbar spine. Patients with a more significant deformity were tracked in a separate adult deformity outcomes protocol. The instability subgroup included patients with abnormal motion (≥3 mm) on flexion and extension radiographs, substantial retrolisthesis, or lateral listhesis. Patients in the stenosis subgroup included patients with central or foraminal stenosis without any abnormalities listed above, or patients undergoing fusion due to the risk for iatrogenic instability following decompression alone.
Results

Baseline, 1-year follow-up, and change scores for the ODI, SF-36 PCS, SF-36 mental composite summary (MCS), and pain scales are summarized in Tables 1 and 2. Only 26 patients (3.6%) reported “much worse,” with 81 patients (11.2%) reporting “somewhat worse,” 166 patients (23.0%) “about the same”, 225 patients (31.2%) “somewhat better,” and 224 patients (31.0%) reporting “much better.” Because of the small number of patients reporting “much worse,” and ANOVA post hoc comparisons showing that the outcome scores of the “much worse” and “somewhat worse” were not statistically significantly different from each other, the “much worse” patients were combined with the “somewhat worse” patients to create a “worse” cohort consisting of 107 patients (14.8%). The change scores broken down by the patient responses to HTI are reported in Table 2. The association between the HTI responses and change scores was significant ($p < 0.0001$) for each outcome measure: ODI ($r = 0.540$), SF-36 PCS ($r = 0.537$), back pain ($r = 0.519$), and leg pain ($r = 0.427$).

Each calculation method yielded a different CIDET threshold value for each HRQOL measure (Table 3). These CIDET threshold values ranged from 0.17 to 9.06 for ODI, −0.32 to 4.43 for SF-36 PCS, 0.02−1.50 for back pain, and 0.02−1.50 for leg pain. The SEM and effect size averages, it appears that patient-perceived deterioration may actually be the lack of improvement. It makes sense that patients treated surgically would expect to improve and in the absence of improvement believe they have deteriorated. This is important for clinical decision-making and should be addressed in the preoperative counseling informed consent process. It is critical to discuss with patients the possibility of being worse, and within our cohort of primary, instrumented lumbar fusions, 14.8% of patients feel they are “worse” 1 year after surgery.

Intuitively, the HTI of the SF-36 may be seen as an adequate qualitative measure of a patient’s change in health status, without the need to quantify this change. However, the HTI alone is a measure of overall perception of a health change, and lacks the depth and range of disease specific measures, such as the ODI. Also, the HTI is an ordinal scale with a range of 1 to 5, while the ODI is a longer ordinal scale with a range of 0 to 100. Helping

<table>
<thead>
<tr>
<th>Response</th>
<th>Baseline</th>
<th>1 Year</th>
<th>Baseline</th>
<th>1 Year</th>
<th>Baseline</th>
<th>1 Year</th>
<th>Baseline</th>
<th>1 Year</th>
<th>Baseline</th>
<th>1 Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worse</td>
<td>107</td>
<td>8.16</td>
<td>7.28</td>
<td>7.77</td>
<td>7.01</td>
<td>56.742</td>
<td>55.08</td>
<td>26.598</td>
<td>26.28</td>
<td>33.360</td>
</tr>
<tr>
<td>About the same</td>
<td>166</td>
<td>7.74</td>
<td>5.40</td>
<td>7.21</td>
<td>4.97</td>
<td>50.08</td>
<td>39.97</td>
<td>27.735</td>
<td>31.84</td>
<td>38.590</td>
</tr>
<tr>
<td>Somewhat better</td>
<td>225</td>
<td>7.69</td>
<td>4.61</td>
<td>7.37</td>
<td>4.54</td>
<td>51.240</td>
<td>35.44</td>
<td>27.379</td>
<td>34.43</td>
<td>38.518</td>
</tr>
<tr>
<td>Much better</td>
<td>224</td>
<td>7.28</td>
<td>1.95</td>
<td>7.36</td>
<td>2.38</td>
<td>47.210</td>
<td>18.20</td>
<td>29.604</td>
<td>44.06</td>
<td>41.164</td>
</tr>
<tr>
<td>p value</td>
<td>0.001</td>
<td>0.000</td>
<td>0.229</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Clinically important deterioration in lumbar spine surgery

TABLE 2: Mean change in HRQOL scores from baseline to 1-year stratified by response to the HTI*

<table>
<thead>
<tr>
<th>Response</th>
<th>No. (%)</th>
<th>Back Pain</th>
<th>Leg Pain</th>
<th>ODI</th>
<th>SF-36 PCS</th>
<th>SF-36 MCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>much worse</td>
<td>26 (3.6)</td>
<td>1.13</td>
<td>0.92</td>
<td>-1.11</td>
<td>-1.61</td>
<td>-2.18</td>
</tr>
<tr>
<td>somewhat worse</td>
<td>81 (11.2)</td>
<td>0.80</td>
<td>0.71</td>
<td>2.51</td>
<td>0.11</td>
<td>-2.31</td>
</tr>
<tr>
<td>about the same</td>
<td>166 (23)</td>
<td>2.35</td>
<td>2.24</td>
<td>10.72</td>
<td>4.11</td>
<td>3.10</td>
</tr>
<tr>
<td>somewhat better</td>
<td>225 (31.2)</td>
<td>3.08</td>
<td>2.83</td>
<td>15.80</td>
<td>7.05</td>
<td>4.36</td>
</tr>
<tr>
<td>much better</td>
<td>224 (31)</td>
<td>5.32</td>
<td>4.98</td>
<td>29.02</td>
<td>14.46</td>
<td>10.84</td>
</tr>
</tbody>
</table>

* The association between the HTI responses and change scores was significant (p < 0.0001) for each outcome measure.

define a quantitative measurement to describe a qualitative opinion is necessary to study outcomes in a statistically meaningful method.

Limitations to the study include the small sample size. We attempted to address this by combining the “much worse” and “somewhat worse” cohorts into a “worse” group. Although a study using a single identifiable lumbar degenerative cohort to investigate deterioration is ideal, the reported rate of deterioration of the entire cohort as well as within the different diagnostic pathologies is relatively small, such that a valid subanalysis cannot be performed. The use of a general health assessment (HTI) as an external anchor may be influenced by other health conditions not under study. That is, patients may report that they are “much worse” than 1 year previously because of health problems unrelated to their lumbar fusion. However, this may apply to all patients in the entire cohort as well, not just the “worse” group. Also, recall bias must be considered, as patients are required to retrospectively recall and compare their health status from 1 year prior. Future studies with larger cohorts could more clearly define a CIDET threshold.

Clinically important deterioration is a quantitative assessment of clinically relevant change (deterioration) in HRQOL measures. Unlike MCID, a threshold for clinical deterioration was difficult to identify. This may be due to the relatively small number of patients who reported being worse after surgery and the variability across anchor-based and distribution-based calculation methods to determine CIDET thresholds. Future studies with larger cohorts could more clearly define a CIDET threshold.

Conclusions

It was difficult to define a clinically important threshold of deterioration for the commonly used HRQOLs in spine surgery. Approximately 15% of patients report being worse after lumbar fusion surgery, with some patients interpreting the absence of improvement in HRQOLs as deterioration.

Disclosure

Dr. Glassman is an employee of Norton Healthcare, has received support of non–study-related clinical or research effort from Norton Healthcare, is a patent holder with Medtronic, served as a consultant to Medtronic until July of 2011, served on a board at Medtronic until July of 2011, and is President-Elect of the Scoliosis Research Society (unpaid). Dr. Carreon is an employee of Norton Healthcare; has received research grants from Norton Healthcare and AO Spine paid directly to the Scoliosis Research Society, and from the Orthopedic Research and Educational Fund (OREF); previously received funds for travel expenses and accommodations from the OREF, Department of Defense, Association for Collaborative Spine Research, and NIH; has received honoraria for participation in Review Panels for Medtronic and the Children’s Tumor Fund; is a member of the Editorial Advisory Board for Spine and Spine Journal; and is a Board Member on the University of Louisville Institutional Review Board. Medtronic provided funds directly to the database company; no funds were paid directly to an individual or an individual’s institution (January 2002 to September 2009). NuVasive provides funds directly to the database company; no funds are paid directly to an individual or an individual’s institution. Author contributions to the study and manuscript preparation include the following. Conception and design: all authors. Acquisition of data: Carreon. Analysis and interpretation of data: all authors. Drafting the article: Gum. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Carreon. Statistical analysis: Carreon. Administrative/technical/material support: Carreon.

References

4. Copay AG, Glassman SD, Sabuch BR, Berven S, Schuler TC,
Carreon LY: Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. Spine J 8:968–974, 2008

Accepted August 1, 2013. Please include this information when citing this paper: published online September 6, 2013; DOI: 10.3171/2013.8.SPINE12804.
Address correspondence to: Leah Y. Carreon, M.D., M.Sc., Norton Leatherman Spine Center, 210 E. Gray St., Ste. 900, Louisville, KY 40202. email: leah.carreon@nortonhealthcare.org.