Clinical improvement through nonoperative treatment of adult spinal deformity: who is likely to benefit?

KSENiya SLOBODYANYUK, B.A.,1 CAROLINE E. POORMAN, B.A.,1 JUSTIN S. SMITH, M.D., Ph.D.,2 THEMISTOCLES S. PROTOPSALTSIS, M.D.,1 RICHARD HOSTIN, M.D.,3 SHAY BESS, M.D.,4,5 GREGORY M. MUNDIS JR., M.D.,1 FRANK J. SCHWAB, M.D.,1 and VIRGINIE LAFAGE, Ph.D.,1
ON BEHALF OF THE INTERNATIONAL SPINE STUDY GROUP

1New York University Hospital for Joint Diseases, New York, New York; 2Department of Neurosurgery, University of Virginia Medical Center, Charlottesville, Virginia; 3Department of Orthopedic Surgery, Baylor Scoliosis Center, Plano, Texas; 4Department of Orthopedic Surgery, Rocky Mountain Hospital for Children, Denver, Colorado; and 5San Diego Center for Spinal Disorders, La Jolla, California

Object. The goal of this study was to determine the outcome and risk factors in patients with adult spinal deformity (ASD) who elected to receive nonoperative care.

Methods. In this retrospective study the authors reviewed a nonoperative branch of the International Spine Study Group database, derived from 10 sites across the US. Specific inclusion criteria included nonoperative treatment for ASD and the availability of Scoliosis Research Society (SRS)-22 scores and radiographic data at baseline (BL) and at 1-year (1Y) follow-up. Health-related quality of life measures were assessed using the SRS-22 and radiographic data. Changes in SRS-22 scores were evaluated by domain and expressed in number of minimum clinically important differences (MCIDs) gained or lost; BL and 1Y scores were also compared with age- and sex-matched normative references.

Results. One hundred eighty-nine patients (mean age 53 years, 86% female) met inclusion criteria. Pain was the domain with the largest offset for 43% of patients, followed by the Appearance (23%), Activity (18%), and Mental (15%) domains. On average, patients improved 0.3 MCID in Pain over 1Y, without changes in Activity or Appearance. Baseline scores significantly impacted 1Y outcomes, with up to 85% of patients in the mildest category of deformity being classified as < 1 MCID of normative reference at 1Y, versus 0% of patients with the most severe initial deformity. Baseline radiographic parameters did not correlate with outcome.

Conclusions. Patients who received nonoperative care are significantly more disabled than age- and sex-matched normative references. The likelihood for a patient to reach SRS scores similar to the normative reference at 1Y decreases with increased BL disability. Nonoperative treatment is a viable option for certain patients with ASD, and up to 24% of patients demonstrated significant improvement over 1Y with nonoperative care.

(http://thejns.org/doi/abs/10.3171/2014.3.FOCUS1426)

Key Words • spinal deformity • health-related quality of life • outcomes • nonoperative treatment

Adult spinal deformity (ASD) is a significant source of disability worldwide.8,10,12 In the absence of significant or progressive neurological deficits, initial treatment is usually nonoperative, with conversion to surgery for nonresponders; the idea is that successful nonoperative management can spare the risks and pain of more invasive treatment.9,20 The nonoperative approach generally consists of a combination of treatments includ-
do not always correlate with improvement. However, in spite of their widespread use, the optimal interpretation and application of scores obtained by these measures are subject to debate. Furthermore, statistically significant differences do not always correspond to clinically meaningful differences. Ideal counseling should involve not only use of data to establish treatment ideals, but also a method of establishing realistic treatment thresholds and setting expectations prior to treatment.

The need to establish realistic aims is highlighted by the conflicting conclusions of studies investigating the efficacy of nonoperative care. On the whole, there is a lack of compelling evidence supporting the use of any nonoperative deformity treatments. However, there may be a subset of patients who would truly benefit from nonoperative measures and they may account for discrepancies between findings. The goal of the present study is to provide a detailed assessment of clinical outcomes for adults treated nonoperatively for spinal deformity, with an emphasis on what is expected and which factors are most important. The hypothesis is that patients with moderate disability will probably benefit from nonoperative treatment. Past a certain threshold of deformity there would probably be no benefit, and those patients should be referred directly for surgery. Ultimately, in this study our goal was to identify a general subset of patients who would benefit from nonoperative treatment to minimize the expenses and stress of care.

**Methods**

**Patient Population and Inclusion Criteria**

This is a retrospective analysis of a consecutive series of patients enrolled in the multicenter prospective International Spine Study Group (ISSG) database. Patients > 18 years of age with ASD were extracted from the database, which was derived from 10 sites across the US. All patients were enrolled through an institutional review board–approved protocol by each site. Criteria for inclusion in the database were at least one of the following: Cobb angle \( \geq 20^\circ \), sagittal vertical axis (SVA) \( \geq 5 \) cm, pelvic tilt \( \geq 25^\circ \), or thoracic kyphosis \( \geq 60^\circ \). Patients with inflammatory arthritis, tumor, or neuromuscular disease were excluded. Operative and nonoperative management, as well as specific treatments, were selected based on physician and patient preferences.

Specific inclusion criteria were adopted for this study in addition to ISSG criteria. Patients who underwent nonoperative treatment with available baseline (BL) and 1-year (1Y) SRS-22 and radiographic measures were included. Of note, information regarding the specific nonoperative treatments that each patient received was unavailable. Patient demographic data such as body mass index (BMI), comorbidities, and previous spine surgery were collected. For each patient, a Charlson comorbidity score was calculated. Full-length standing anteroposterior and sagittal spine radiographs were analyzed at BL with the aid of validated software (SpineView, Laboratory of Biomechanics ENSAM ParisTech) to obtain the following spinopelvic parameters: coronal Cobb angle and apex location, pelvic tilt, pelvic incidence (PI), lumbar lordosis between L-1 and S-1 (LL), the PI-LL mismatch, and the SVA. Based on these radiographic parameters, each patient was classified based on the SRS-Schwab classification of ASD. This produced 8 distinct groups (Table 1) based on the coronal curves and sagittal modifiers of the classification system.

**Health-Related Quality of Life Normative Data and Clinical Classifications**

At BL and 1Y, the health-related quality of life (HRQOL) was evaluated based on the SRS-22 questionnaire. The HRQOL scores for each patient were expressed in terms of minimum clinically important difference (MCID) to standardize data, such that the scores reflected the MCID deviation of each patient from normative values (see Equation). The HRQOL assessment included the SRS-22 domains of Pain, Activity, Appearance, and Mental, but did not include Satisfaction or Total for these domains.

\[
\text{Diff}_{\text{patient domain}} = \frac{\text{SRS}_{\text{patient domain}} - \text{SRS}_{\text{normative domain}}}{\text{MCID}_{\text{domain}}}
\]

Additionally, patients were categorized into 1 of 4 clinical severity groups for each domain based on MCIDs from normative values. The groups were stratified as follows: 1) Worst, > 4 MCIDs below normative; 2) Severe, 2–4 MCIDs below normative; 3) Poor, 1–2 MCIDs below normative; and 4) Asymptomatic, < 1 MCID below normative, or above normative.

To assess clinical improvement, patients were categorized based on change from BL to 1Y into 3 groups: 1) > 1 MCID gain; 2) > 1 MCID loss; and 3) between −1 and +1 MCID gain or loss between BL and 1Y.

**Statistical Analysis**

The means and SDs of HRQOL scores were com-

---

**TABLE 1: Group definition based on the SRS-Schwab classification**

<table>
<thead>
<tr>
<th>Group</th>
<th>Coronal Criteria</th>
<th>Sagittal Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>unclassified</td>
<td>Type N</td>
<td>all modifiers at Grade 0</td>
</tr>
<tr>
<td>thoracic only</td>
<td>Type T</td>
<td>all modifiers at Grade 0</td>
</tr>
<tr>
<td>lumbar only</td>
<td>Type L</td>
<td>all modifiers at Grade 0</td>
</tr>
<tr>
<td>double curve</td>
<td>Type D</td>
<td>at least 1 modifier at Grade + or ++</td>
</tr>
<tr>
<td>sagittal only</td>
<td>Type N</td>
<td>at least 1 modifier at Grade + or ++</td>
</tr>
<tr>
<td>sagittal-thoracic</td>
<td>Type T</td>
<td>at least 1 modifier at Grade + or ++</td>
</tr>
<tr>
<td>sagittal-lumbar</td>
<td>Type L</td>
<td>at least 1 modifier at Grade + or ++</td>
</tr>
<tr>
<td>double sagittal</td>
<td>Type D</td>
<td>all modifiers at Grade 0</td>
</tr>
</tbody>
</table>

* Modifier grades are determined by comparing measurements on frontal and sagittal radiographs against validated criteria that have been shown to correlate with quality of life measures. For further information see Schwab et al., 2012.
Nonoperative treatment of adult spinal deformity: who benefits?

pared with matched normative data, with change determined based on the difference between BL and 1Y. Domains in which patients deviated the least (best) and most (worst) from normative data were determined for BL and 1Y. Next, the patients’ change category (gain > 1 MCID, loss > 1 MCID, or no change) for each of the 4 domains was compared with BL data by using paired-samples t-tests to determine if the BL score was predictive of clinical change. The distribution of patients in each severity category was assessed at BL and 1Y, and comparisons were made across groups based on age, sex, comorbidities, curve type, and previous spine surgery by using ANOVA and chi-square tests, with an adopted significance of 0.05.

Results

Demographic Data

Overall, 189 patients with a mean age of 53 ± 16 years and a BMI of 25 ± 5 kg/m² were included. Female patients comprised 86% of the cohort, and 18% had a history of previous spine surgery. The mean BL Charlson score was 0.97 ± 1.3. At BL, the patients’ radiographic Schwab classifications, based on coronal curve and sagittal modifiers, were as follows: 7% no major deformity (unclassified), 9% thoracic only, 12% lumbar only, 16% double curve, 14% sagittal only, 4% sagittal-thoracic, 20% sagittal-lumbar, and 18% double sagittal (Table 1).

Clinical Classification Groups

Patients differed significantly from matched normative data at both BL and 1Y in all SRS categories. At BL, Pain was the domain of greatest concern (largest deviation from normative) for 43%, followed by the Appearance (23%), Activity (18%), and Mental (16%) categories. This is also reflected in the mean scores for each domain at BL (Table 2). Pain continued to be the domain of most concern at 1Y (37%), followed by Appearance (24%), Activity (22%), and Mental (17%). Thus, the most pertinent domains to evaluate for nonoperative patients are Pain, Activity, and Appearance.

Analysis of demographic data by Pain classification revealed no difference between groups in sex or previous spine surgery, although age, BMI, and Charlson score differed by BL Pain classification (Table 3). Differences in BL demographic data were also seen in the Activity and Appearance domains.

Clinical Improvement and Who Does Best?

Overall analysis of HRQOL scores demonstrated that, on average, patients gained < 0.5 MCID at 1Y (Table 2). However, analysis of the percentage of patients who showed significant change (defined as > 1 MCID gain or loss) within each domain revealed that 24% gained and 11% lost > 1 MCID in Pain; 21% gained and 22% lost > 1 MCID in Activity; 13% gained and 7% lost > 1 MCID in Appearance; and 17% gained and 14% lost > 1 MCID in Mental (Fig. 1).

For all domains, patients who gained > 1 MCID at 1Y had significantly more severe BL scores than patients who lost > 1 MCID (p < 0.001). Patients who gained > 1 MCID had a mean −2.78, −2.50, and −1.67 BL MCID offset from the normative value in Pain, Activity, and Appearance, respectively, whereas patients who lost > 1 MCID had a mean −1.23, −0.48, and −0.74 BL offset from the normative values in the aforementioned domains, in that order. Figure 2 shows the proportion of patients in each 1Y MCID category by BL score for each domain.

Figure 3 summarizes BL and 1Y clinical groups in each pertinent domain. For patients who had the best 1Y outcome (< 1 MCID of normative), BL scores had a significant impact on final outcomes. For Pain, Activity, and Appearance, 77%–85% of patients who were classified Asymptomatic at BL remained so at 1Y. However, only 28%–37% of patients who were in the Poor, 11%–16% who were in the Severe, and 0% who were in the Worst category at BL were classified Asymptomatic at 1Y.

Radiographic Predictors for Outcome

Baseline radiographic parameters, as reflected by Schwab classification, did not significantly impact 1Y clinical outcome. Moreover, no individual radiographic parameter significantly impacted outcome.

Discussion

These findings suggest that there is a subset of patients who may benefit more than others from nonoperative care. Most notably, there was a correlation between baseline HRQOL evaluation and nonoperative HRQOL outcomes at 1 year. Such a correlation has not previously been described. However, the importance of baseline clinical evaluation in predicting treatment results has been highlighted previously. Carreon et al.4 studied the effects of baseline mental and physical status, as reflected by Physical and Mental Component scores of the 36-Item

**TABLE 2: Differences in relation to matched normative reference in 189 patients with ASD**

<table>
<thead>
<tr>
<th>MCID</th>
<th>Activity</th>
<th>Pain</th>
<th>Appearance</th>
<th>Mental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffs from normative ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL</td>
<td>−1.00 ± 2.19</td>
<td>−1.75 ± 1.66</td>
<td>−1.15 ± 0.94</td>
<td>−0.64 ± 1.92</td>
</tr>
<tr>
<td>1Y</td>
<td>−0.96 ± 2.11</td>
<td>−1.48 ± 1.58</td>
<td>−1.08 ± 0.95</td>
<td>−0.48 ± 1.86</td>
</tr>
<tr>
<td>1Y improvement</td>
<td>0.036 ± 0.27</td>
<td>0.27 ± 1.01</td>
<td>0.069 ± 0.72</td>
<td>0.16 ± 1.5</td>
</tr>
</tbody>
</table>

* Expressed as the mean ± SD. Diffs = differences; ref = reference.
Short Form Health Survey (SF-36) and the ODI, on postoperative clinical and functional outcomes, finding that patients with superior preoperative Physical Component and ODI scores also had better postoperative scores.

Although patients with superior preoperative scores demonstrated better outcome, those patients who improved > 1 MCID on follow-up actually had more severe baseline scores than those who deteriorated > 1 MCID, for all domains. A study by Moal et al. evaluating predictors of surgical outcome noted that only 17% of patients with the worst baseline scores did not improve or deteriorated, as opposed to 44% of patients in the least-impaired baseline category. Schwab et al. found that those patients who were most likely to improve substantially had significant preoperative disability, suggesting that patients with superior baseline characteristics have less potential to improve and would be at risk for surgical morbidity comparable to the severity of their spine deformity.

Taken together, these findings suggest fundamental differences between mildly and more severely affected patients with deformity, not only in their initial clinical picture, but also in the extent to which they are affected by treatment. Although it is unlikely that the patients with the worst scores will approximate normative values after any treatment, upward of 40% have been shown to improve significantly postoperatively. On the other hand, patients in the mildest deformity category often retain baseline status following either surgical or nonoperative treatment. In these patients, the relative merits and risks of available treatments warrant further evaluation prior to management planning.

Several of the present findings were inconsistent with previous results. First, patients demonstrated substantial improvement in the Pain category. However, previous studies conducted using a numeric rating scale questionnaire to evaluate pain outcomes found that nonoperative patients did not experience significant improvement in back or leg pain at 2 years. Similarly, Molinari and

<table>
<thead>
<tr>
<th>Pain Category</th>
<th>No. (%)</th>
<th>Age in Yrs</th>
<th>BMI in kg/m²</th>
<th>Charlson Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>189</td>
<td>53 ± 16</td>
<td>25 ± 5</td>
<td>0.97 ± 1.3</td>
</tr>
<tr>
<td>Asymptomatic</td>
<td>69 (36)</td>
<td>49 ± 15†</td>
<td>23 ± 4†</td>
<td>0.48 ± 0.7†</td>
</tr>
<tr>
<td>Poor</td>
<td>42 (22)</td>
<td>51 ± 15†</td>
<td>24 ± 5†</td>
<td>0.95 ± 1.4‡</td>
</tr>
<tr>
<td>Severe</td>
<td>60 (32)</td>
<td>57 ± 17</td>
<td>28 ± 6</td>
<td>1.6 ± 1.5</td>
</tr>
<tr>
<td>Worst</td>
<td>18 (10)</td>
<td>55 ± 12</td>
<td>26 ± 4§</td>
<td>0.94 ± 1.2</td>
</tr>
</tbody>
</table>

* BMI values were missing for 8 patients (3 Asymptomatic, 2 Poor, 2 Severe, and 1 Worst). Age, BMI, and Charlson score are expressed as the mean ± SD.
† Baseline differences in mean age, BMI, and Charlson score were statistically significant compared to Severe classification.
‡ The mean Charlson score for the Poor classification at BL was significantly different from both Severe and Asymptomatic classifications.
§ Baseline difference in mean BMI was statistically significant compared to Asymptomatic classification.
Gerlinger\textsuperscript{18} found that active-duty military personnel who were treated nonoperatively did not experience substantial improvements in pain or return to duty.

Second, in the present study there was no correlation between radiographic parameters and clinical outcome. A 2013 prospective cohort of patients with ASD undergoing surgery evaluated whether the SRS-Schwab classification can predict HRQOL changes as a treatment response.\textsuperscript{23} One-year changes in the SVA category were associated with changes in ODI, SF-36 Physical Component score, and SRS-22 in total and all subscores. One-year changes in the PI-LL parameter were associated with changes in SF-36 and SRS-22 total score and subscores. Overall, SVA and PI-LL changes were correlated with significant changes in ODI and SRS subscores.

One possible explanation for deviations in results relates to the present study design. Patients were evaluated using the SRS-22 survey, a disease-specific questionnaire encompassing several domains of subjective function and well-being. This survey was selected for the opportunity to use age- and sex-stratified normative data. In addition to establishing discriminant validity, normative values provide a reference by which to interpret and compare results. Baldus et al.\textsuperscript{1} distributed an SRS-22 survey to 1346 healthy subjects, finding that declining values were seen in females and older populations. Thus, it is difficult to establish a reference for HRQOL scores without accounting for demographic factors through the use of age- and sex-matched controls.

The use of MCIDs with the SRS-22 instrument offers additional potential advantages. The MCID allows for the measurement of change from baseline that is important to the patient and clinically meaningful. The MCID is defined as the smallest change in a score that is detected with sufficient precision to rule out chance and is important to the patient. The MCID is not a fixed value but varies with the population and the specific outcome measure. The MCID is calculated as the standard deviation of the measurement times a constant, typically 0.8 or 1.0. The use of the MCID allows for the assessment of meaningful change in the context of the patient's baseline status.

Fig. 2. Bar graphs showing frequencies of 1Y MCID change by BL classification for the SRS-22 instrument in the Pain, Activity, and Appearance categories.

Fig. 3. Bar graphs showing frequencies of 1Y outcomes by BL classification for the SRS-22 instrument in the Pain, Activity, and Appearance categories.
systematic and objective measures of clinical significance required by research. Furthermore, although past experience is an accepted basis for significance assessment in clinical practice, researchers may not always have the necessary background to subjectively assess HRQOL measures.

Although the MCID scale offers certain advantages, it also has limitations. Julious and Walters recently concluded that there is a weak relationship between the MCID and the proportion of patients who benefit from treatment, reflecting the common variability in treatment response and the power of the placebo effect. The authors suggested that measurements of effect size are more indicative of treatment benefit than is the MCID. Rennard also proposed that, because select individuals are inherently better able to discriminate subtle variations in human activity, the MCID is a more valuable tool for groups than for individuals.

Beyond the controversy about the utility of the MCID scale, another limitation of this study was the generalization of nonoperative care without indication of individual treatment types. Evaluating nonoperative management overall limits information on which specific treatments are of most value. On the other hand, because nonoperative management often involves a combination of treatments, grouping it as a single category allows for a more realistic assessment of expected outcomes with nonoperative care as it is used in the community. A third limitation of this study is that outcome was evaluated based on SRS classification, without elaborating on other specific characteristics that differed between the groups. Future studies should aim to compare specific medical and psychiatric comorbidities to establish more elaborate treatment criteria. However, this study provides the initial step by supporting the notion that some patients will benefit more than others, increasing treatment efficiency and potentially improving the perceived utility of nonoperative care.

Conclusions

This study provides evidence that the odds of improvement with nonoperative treatment are optimal for patients in the mildest baseline disease category, with the odds approaching zero as baseline status declines. Thus, the treatment implications of these findings most benefit those in the mildest category, for whom surgery may be avoided, as well as the patients with the worst baseline scores, who may be advised to proceed directly to surgery and avoid the unnecessary costs and prolonged discomfort of a futile nonoperative trial. However, for those patients who are between the best and the worst baseline classifications, the optimal approach is still unclear. Furthermore, the findings of this study can help set patient expectations. More severely disabled patients can be informed that although there is a high chance of substantial improvement following nonoperative care, it is unlikely that they would approach a normative state without surgery. On the other hand, patients with milder symptoms can be informed that they may show slight improvement or remain in their baseline mild category of impairment with nonoperative care, at least over a short-term follow-up period.

Disclosure

Dr. Smith is a consultant for Biomet, Globus, Medtronic, and DePuy. He received clinical or research support for this study from DePuy, and also for a non-study-related clinical or research effort that he oversaw. Dr. Bess is a consultant for DePuy Spine, Medtronic, and K2M. He received clinical or research support for this study from DePuy Spine, and he also received support from Medtronic for a non-study-related clinical or research effort that he oversaw. Dr. Mundis is a consultant for NuVasive and K2M. He received support from NuVasive for a non-study-related clinical or research effort that he oversaw, and he also received fellowship support from NuVasive, Pioneer, and the Orthopaedic Research and Education Foundation (OREF). Dr. Protopsaltis is a consultant for Globus, and he is a member of the speaker’s bureau for Alphatec and K2M. Dr. Hostin is a consultant for DePuy. He also received clinical or research support for this study from DJO Global and DePuy Spine.

Author contributions to the study and manuscript preparation include the following. Conception and design: Lafage, Slobodyanyuk, Poorman, Schwab. Acquisition of data: Lafage, Smith, Protopsaltis, Hostin, Bess, Mundis, Schwab. Analysis and interpretation of data: Lafage, Slobodyanyuk, Poorman. Drafting the article: all authors. Critically revising the article: Lafage, Slobodyanyuk. Reviewed and approved the final version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Lafage. Statistical analysis: Lafage, Slobodyanyuk, Poorman. Administrative/technical/material support: Lafage, Smith, Protopsaltis, Hostin, Bess, Mundis, Schwab. Study supervision: Lafage, Schwab.

References

Nonoperative treatment of adult spinal deformity: who benefits?