Development and Validation of a Multidomain Surgical Complication Classification System for Adult Spinal Deformity

Eric O. Klineberg, MD,a Joseph B. Wick, MD,a Renaud Lafage, MSc,b Virginie Lafage, PhD,b Ferran Pellise, MD, PhD,c Sleiman Haddad, MD, PhD,c Caglar Yilgor, MD,d,e Susana Núñez-Pereira, MD, PhD,f Munish Gupta, MD,g Justin S. Smith, MD, PhD,f,g Christopher Shaffrey, MD,h Frank Schwab, MD,b Christopher Ames, MD,b Shay Bess, MD,h Stephen Lewis, MD,h Lawrence G. Lenke, MD, and Sigurd Berven, MDh.

From the aDepartment of Orthopedic Surgery, University of California Davis, Sacramento, CA; bHospital for Special Surgery, New York, NY; cSurgery Unit, Hospital Universitario Val Hebron, Barcelona, Spain; dAchıadem University School of Medicine, Istanbul, Turkey; eDepartment of Orthopaedic Surgery, Washington University School of Medicine, St. Louis, MO; fDepartment of Neurosurgery, University of Virginia Medical Center, Charlottesville, Virginia; gDepartment of Neurosurgery, University of California, San Francisco, CA; hDenver International Spine Center, Presbyterian St. Luke's/Rocky Mountain Hospital for Children, Denver, CO; iUniversity of Toronto and Toronto Western Hospital, Toronto, ON, Canada; jDepartment of Orthopaedic Surgery, Columbia University Medical Center, New York, NY; and kDepartment of Orthopaedic Surgery, University of California, San Francisco, CA.


The manuscript submitted does not contain information about medical devices/agents.

DePuy Synthes/ISSG Foundation funds were received in support of this work.

Relevant financial activities outside the submitted work: board membership, consultancy, grants, royalties, stocks, expert testimony, travel/accommodations/meeting expenses.

Address correspondence and reprint requests to Eric O. Klineberg, MD, University of California Davis, 4860 Y Street, Suite 3800 Sacramento, CA 95817; E-mail: eoklineberg@ucdavis.edu

DOI: 10.1097/BRS.0000000000003766

Study Design. Prospective analysis of example cases

Objective. The aim of this study was to analyze the accuracy and repeatability of a new comprehensive classification system for capturing complications data in adult spinal deformity.

Summary of Background Data. Complications are common in adult spinal deformity surgery. However, no consensus exists on the definition or classification of complications in adult spinal deformity surgery. The lack of consensus significantly limits understanding of complications’ effects on outcomes in surgery for adult spinal deformity.

Methods. Using a Delphi method, members of the International Spine Study Group, AO Spine, and the European Spine Study Group collaborated to develop an adult spinal deformity classification system. The multidomain classification system accounts for medical complications (cancer, cardiopulmonary, central nervous system, gastrointestinal, infectious, musculoskeletal, renal) and surgical complications (implant complications, radiographic complications, neurologic events, intraoperative events, and wound complications). Seventeen individuals (“event readers”), including spine surgeons, trainees, and research coordinators, used the new classification system two separate times to analyze complications in ten example cases. The accuracy and repeatability of the classification system were subsequently calculated based on the providers’ responses for the example cases.

Results. The 10 example cases included 22 complications. Nearly 95% of complications were captured by ≥95% of the event readers. The system demonstrated good repeatability of 86.9% between the first and second set of responses provided by event readers.

Conclusion. The ISSG-AO Multi-Domain Spinal Deformity Complication Classification System for Adult Spinal Deformity demonstrated good accuracy and repeatability among both surgeons and research coordinators in capturing complications in adult spinal deformity surgery. The ISSG-AO system may be applied to help better understand the impact of complications on outcomes and costs in adult spinal deformity surgery.

Key words: adult spinal deformity, adverse event, complications, costs, length of stay, neurologic deficit, outcomes, postoperative, radiographic, reoperation, revision, thoracolumbar.

Level of Evidence: 5


Adult spinal deformity (ASD) is common, with reported prevalence surpassing 60% among elderly patients. Surgical approaches to ASD include a broad spectrum of invasiveness and variability in complication rates, and procedures are often performed in older...
patients with multiple comorbidities.\textsuperscript{2} As a result, complications occur frequently, with rates ranging from 37\% to 71\%.\textsuperscript{3–7} Complications are an important metric for surgical quality and safety, yet significant variability exists in the literature regarding expected complication rates, severity, and impact on outcomes. With increasing prevalence of ASD surgery and health care’s value-based focus,\textsuperscript{3} spine surgeons must be able to accurately characterize complication incidence and impact on surgical outcomes.\textsuperscript{2,7} Understanding complications also has important implications for preoperative counseling, setting appropriate patient expectations, shared decision making, and preoperative planning.

To date, complications in ASD surgery have been broadly categorized as “major” or “minor,”\textsuperscript{9,10} or classified by increasing severity.\textsuperscript{11} However, this is inadequate for understanding complications’ impact on outcomes.\textsuperscript{10} Similarly, inconsistent definitions of complications have likely been responsible for wide variation in reported ASD surgical complication rates.\textsuperscript{2} In response to the need for a better ASD-specific complications classification system, the International Spine Study Group (ISSG), AO Spine Deformity Knowledge Forum, and European Spine Study Group (ESSG) held a consensus meeting utilizing a Delphi method to develop a new ASD complications categorization system. Herein, we describe the development and validation of the ISSG-AO Spine Complications Classification System.

\textbf{MATERIALS AND METHODS}

\textbf{Classification System Development}

A Delphi approach was used to develop the classification system. The Delphi process utilized a consensus conference with a panel comprised of AO Knowledge Forum members, who are spine experts from around the globe. Shared knowledge and experience of the panel was used to determine the relevant domains under which complications should be classified. Open-ended questions were first proposed regarding what comprised a complication classification, and what elements were important to include within the system. Consensus was achieved through in-person voting. Several voting rounds were required to gain consensus. Figure 1 demonstrates the complication categories and subcategories included within the classification system. Intervention severity classification by invasiveness was based on the Clavien-Dindo complication system for general surgery;\textsuperscript{12} similar classifications have been applied to spine surgery.\textsuperscript{11} A high level of granularity was desired to allow for consistent complication type definition, especially as previous systems that broadly categorize complications have been shown to be inadequate.\textsuperscript{10} Neurologic complications were specifically emphasized as they are an especially burdensome risk inherent to spinal surgery. Complications are broadly classified as medical or spine-related adverse events. Medical complications are subclassified as cancer, cardiopulmonary, central nervous system, gastrointestinal, nonspine infection, musculoskeletal not related to the spine procedure, and renal. Spine complications are subclassified as implant-related, radiographic, neurologic, intraoperative, and wound-related. The classification system also records complication timing (intraoperative, in-hospital, or post-discharge), whether the event was related to the spine procedure, intervention details with invasiveness subclassification as none, mild, moderate, or severe, and complication resolution. Care was taken to ensure that complications were distinguished from expected clinical outcomes. For example, preoperative neurologic deficits are recorded to ensure that unresolved preoperative deficits are not counted as complications of the surgery itself. See Supplemental Digital Content 1, http://links.lww.com/BRS/B658 for the ISSG-AO Spine Complications Classification System reporting form.

\textbf{Validation Study With Example Cases}

Ten example cases (Figure 2; See Supplemental Digital Content 2, http://links.lww.com/BRS/B659 for a complete list of example cases) based on actual clinical scenarios were developed and used to assess the classification system’s accuracy and repeatability. Cases included one to three complications with an average of two per case. There were a total of 22 complications, of which 40.9\% were medical and 59.1\% were spine-related. Five spine-related complications included neurologic deficits. Three required no intervention, nine required mild, six moderate, and four severe intervention.

\textbf{Case 1}

\begin{itemize}
  \item 76 year old female underwent revision T4-L4 for sagittal plane deformity on 9/3/10
  \item Post-op was noted to have anemia requiring transfusion.
  \item She developed an ileus on POD #2 which required 2 days of bowel rest, but then resolved. The surgical team felt this increased her LOS by 1 day.
  \item At 1-year post-op she was noted to have L4-S and L5-S1 degenerative changes with mild central stenosis, severe foraminal stenosis with radiculopathy related to on set of adjacent segment degeneration in lumbar spine, no weakness. Due to patient’s age, will attempt non-operative management
\end{itemize}

\begin{figure}
\centering
\includegraphics[width=0.8\textwidth]{figure1.png}
\caption{Complication categories and subcategories within the ISSG-AO Spine Complications Classification System.}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=0.8\textwidth]{figure2.png}
\caption{Example case #1 as given to event readers. Cases two to 10 were presented in a similar format.}
\end{figure}
required severe interventions. At final follow-up, 15 complications were resolved, three were partially resolved, three were unresolved, and one resulted in death.

Twenty-two spine care providers (“event readers”) were invited to apply the classification system to the ten cases. Seventeen event readers, including ten fellowship trained spine surgeons, five trainees (residents/fellows), and two research coordinators, completed data collection forms for all cases. To assess intrarater reliability, event readers repeated their evaluations of the same 10 cases 2 weeks after their initial evaluation. “Round one” and “round two” refer to event readers’ responses provided in the first and second evaluations of cases, respectively. Responses were compared to the answer key and considered to be “at least” correct if event readers chose correctly under event category and were considered “exactly” correct if the correct response under the event specifics category was chosen. For example, if the patient sustained a medical-related complication event, an event reader’s response was considered at least correct if their answer matched the answer key response under the “medical event” category, and was exactly correct if the response matched the answer key response for “medical event specifics.” An intervention answer was at least correct if the reader chose correctly for intervention severity, and was exactly correct if the reader chose correctly for intervention details (ie, the patient required a nonroutine ICU admission under the moderate severity intervention category). In sub-analysis by event type, “mechanical” complications were a combination of implant-related and radiographic complications, including proximal junctional kyphosis.

Statistical Analysis
Descriptive statistics were utilized for validation study outcomes. The classification system’s ability to accurately capture complications data to an increasing degree of granularity was assessed by sequentially combining and comparing the percent of correct responses from each category; these are reported as the M1, M2, and M3 data points (Figure 3). The M1 data point combines percent of correct responses from timing, type, medical/spine event, and medical/spine event specifics categories. M2 combines M1 with percent of correct responses from the neurologic complications category. M3 combines M2 with the correct responses from the intervention category. M1, M2, and M3 were calculated for all 10 example cases, then subcategorized to evaluate the ability of the system to accurately capture complication events based on whether the event was medical versus spine-related, an isolated neurologic event, an event with complications including a neurologic component, intraoperative, radiographic or implant-related, or a central nervous system event. M1, M2, and M3 were further subcategorized as “at least” or “exactly” correct, as described above. After calculating M1, M2, and M3, the Pearson correlation coefficient between case number and percent of correct responses was calculated to assess whether accuracy improved as event readers gained experience with the system. Analyses were performed using SPSS version 22.0 (IBM Corp, Armonk, NY).

RESULTS
Overall Accuracy of Capturing Events
In round one, the 17 event readers captured a total of 94.4% of complication events, with 16 of 22 events being captured by >95% of the event readers. Type of event (medical versus spine) was captured correctly in 87.8% of events, event category within medical or spine complications was captured correctly in 92.4% of events, and event specifics were captured correctly for 88.6% of events. Timing, intervention grading (mild, moderate, severe), and presence of
neurologic complications were accurately captured for 89.1%, 81.3%, and 76.5% of events, respectively. In round two, type of event was captured correctly in 93.8% of events, category was captured correctly in 93.2% of events, and event specifics were captured correctly in 89.2% of events. Timing, grading of intervention, and presence of neurologic complications were accurately captured in 96.3%, 88.4%, and 77.9% of events, respectively.

Table 1 and Figure 4 show the system’s ability to accurately categorize events as level of complication detail increases (M1, M2, M3). Overall, the M1 data point percentage for at least correct answers was 86.7%, and 82.4% for exactly correct answers. For M2, at least and exactly correct were 71.9% and 67.1%, respectively. For M3, at least correct was 57.9% and exactly correct was 47.6%.

Among complication events without a neurologic component, there were a total of 255 potentially correct answers that could be provided by the event readers. For example, for events without a neurologic component, readers could correctly answer “no” to the question “Was there a neurologic component to this adverse event” (see Supplemental Digital Content 1, http://links.lww.com/BRS/B658, slide 3). Of the 255 potentially correct selections, readers provided 197 correct answers (77%) and 58 incorrect answers (23%). However, answers were considered incorrect if readers provided no answer (neither “no” or “yes” was selected for whether a neurologic component existed in the adverse event). In the neurologic category for events without a neurologic component, there were a total of 53 nonselected answers that were counted as incorrect. Overall, nonselected answers accounted for 91% of the total number of incorrect answers and 21% of answers overall in the neurologic category for events without a neurologic component.

### Accuracy Subanalysis by Event Type

Table 1 shows at least and exact accuracy with complication events subdivided by type of complication, as well as M1, M2, and M3 data points for each event type. The largest gap between at least and exactly correct answers was noted for medical complications, at 88.3% versus 57.9%, respectively. In contrast, accuracy for both at least and exactly correct answers in spine complication events was >80%, and remained >70% for events when neurologic data were included (M2), but dropped to 58.2% if intervention data were included (M3).

For intraoperative, mechanical, and central nervous system (CNS) complications, exactly correct accuracy for category and event details ranged from 92.3% to 100%. Neurologic complications were also captured with a high level of accuracy; neurologic data from both isolated

<table>
<thead>
<tr>
<th>TABLE 1. “At Least” and “Exact” Accuracy With Complication Events Subdivided by Type of Complication (Medical, Surgical, Isolated Neurologic Events, Events With a Neurologic Component, Mechanical, and Central Nervous System), as well as M1, M2, and M3 Analysis for Each Event Type. Mechanical Complications Combine Implant-related and Radiographic Complications</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Timing</strong></td>
</tr>
<tr>
<td>Overall complications</td>
</tr>
<tr>
<td>At least</td>
</tr>
<tr>
<td>Exact</td>
</tr>
<tr>
<td>Medical-related complications</td>
</tr>
<tr>
<td>At least</td>
</tr>
<tr>
<td>Exact</td>
</tr>
<tr>
<td>Spine surgery-related complications</td>
</tr>
<tr>
<td>At least</td>
</tr>
<tr>
<td>Exact</td>
</tr>
<tr>
<td>Intraoperative complications</td>
</tr>
<tr>
<td>At least</td>
</tr>
<tr>
<td>Exact</td>
</tr>
<tr>
<td>Mechanical complications (implant-related and radiographic)</td>
</tr>
<tr>
<td>At least</td>
</tr>
<tr>
<td>Exact</td>
</tr>
<tr>
<td>Isolated neurologic complications</td>
</tr>
<tr>
<td>At least</td>
</tr>
<tr>
<td>Exact</td>
</tr>
<tr>
<td>Complications with neurologic component</td>
</tr>
<tr>
<td>At least</td>
</tr>
<tr>
<td>Exact</td>
</tr>
<tr>
<td>Central nervous system complications</td>
</tr>
<tr>
<td>At least</td>
</tr>
<tr>
<td>Exact</td>
</tr>
</tbody>
</table>
neurologic complications and complications with a neurologic component were captured with at least accuracy in >90% of events. However, accuracy decreased substantially when intervention data were combined with neurologic complication data. For example, exactly correct M3 for medical events was 36.3%, spine events 58.2%, isolated neurologic events 45.1%, events with a neurologic component 69.4%, mechanical events 56.9%, and CNS events 44.9%.

### Subanalysis by Event Recorder Type

Table 2 shows surgeon and research coordinator accuracy in capturing event details. For the entire data set, M3 at least accuracy for research coordinators was 67.3% compared to 63.3% among surgeons; M3 exact accuracy was 46.7% for research coordinators and 48.4% for surgeons.

### Accuracy Over Time

Figure 5 shows a trend toward increased accuracy with each subsequent case analyzed during round one, with $r = 0.747$ between case number and M3 at least accuracy.

### Repeatability

Event readers submitted 3984 answers between rounds one and two. Of these, 3060 were similar between rounds (76.8%), and the accuracy of identifying complications within each case remained similar, with most cases having >75% accuracy. Similar consistency between rounds was noted between surgeons and research coordinators, at 77.6% and 76.1%, respectively. Repeatability was found to be higher at 86.9% when answers were analyzed categorically (ie, whether answers between the first and second rounds remained “at least” correct, remained wrong, or remained the same). See Supplemental Digital Content 3, http://links.lww.com/BRS/B660 for a visual representation of answer repeatability between rounds one and two when analyzed by response categories.

### DISCUSSION

This study analyzed the ISSG-AO Spine Complications Classification System’s accuracy and repeatability in capturing ASD surgical complication data. In 10 example cases including 22 complications analyzed by 17 event readers, nearly 95% of complications were captured by >95% of the readers. Event-specific details were captured in 88.6% of complications. Accuracy remained high in capturing event details, ranging from 92.3% to 100% when complication events were subdivided by complication type. The system accurately captured complication type, timing, and category as reflected by the M1 analysis; however, accuracy decreased with inclusion of neurologic complication and intervention details between the M1 and M3 analyses. Diminishing accuracy between M1-3 was influenced by a large number of missing responses for neurologic data; this

### TABLE 2. Comparison of Analysis Accuracy Between Surgeons and Research Coordinators

<table>
<thead>
<tr>
<th>Event Reporter Type</th>
<th>Timing</th>
<th>Type</th>
<th>Category</th>
<th>Subcategory</th>
<th>Neurologic</th>
<th>Intervention</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgeon</td>
<td>99.4%</td>
<td>95.2%</td>
<td>95.8%</td>
<td>92.1%</td>
<td>83.0%</td>
<td>86.1%</td>
<td>88.5%</td>
<td>75.8%</td>
<td>67.3%</td>
</tr>
<tr>
<td>Research coordinator</td>
<td>97.9%</td>
<td>92.6%</td>
<td>93.6%</td>
<td>92.6%</td>
<td>79.8%</td>
<td>90.4%</td>
<td>85.1%</td>
<td>68.6%</td>
<td>63.3%</td>
</tr>
<tr>
<td>Exact accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgeon</td>
<td>97.0%</td>
<td>95.2%</td>
<td>93.3%</td>
<td>87.3%</td>
<td>78.2%</td>
<td>66.1%</td>
<td>83.0%</td>
<td>69.1%</td>
<td>46.7%</td>
</tr>
<tr>
<td>Research coordinator</td>
<td>95.7%</td>
<td>92.6%</td>
<td>93.1%</td>
<td>91.0%</td>
<td>77.7%</td>
<td>73.9%</td>
<td>81.9%</td>
<td>65.4%</td>
<td>48.4%</td>
</tr>
</tbody>
</table>
accounted for 91% of incorrect answers and 21% of overall answers in the neurologic category for events without a neurologic component. Surgeons and research coordinators had similar accuracy in capturing event details, as demonstrated by similar M3 accuracy levels. The system demonstrated good repeatability of 86.9% between the first and second rounds, and accuracy tended to improve over time as event readers’ experience increased. The granularity and resultant complexity of our data precluded calculation of kappa values, preventing direct comparison to prior classification systems’ inter- and intrarater repeatability. However, the ISSG-AO Spine Complications Classification system demonstrated a high degree of repeatability, similar to other classification systems, including the Spine Adverse Events Severity System (SAVES)11 and adaptations of the Clavien-Dindo classification system to orthopedic surgery.13

Overall, results suggest that the ISSG-AO Spine Complications Classification System allows for better classification of ASD complications than current practices, which are inadequate for evaluating complications’ impact on clinical outcomes.10 Although the system may be overly inclusive or “double count” complication events, this is unavoidable when designing a system to thoroughly capture and account for complications. The risk of being overly inclusive is preferable to missing complications, as the intention is to develop a repeatable, accurate system for capturing complications and subsequently characterizing their impact on outcomes.

The wide range of reported ASD complication rates, from 37% to 71%, reflects the need for a thorough, accurate, and repeatable ASD-specific complication classification system.2–6 Although attempts have been made to adapt other complication systems, such as the Clavien-Dindo system for general surgery, to orthopedic procedures,13 and spine complication systems have been described,11 there is no comprehensive, detailed, and repeatable system specifically intended for classifying complications in ASD surgery.

Therefore, there is significant potential for the ISSG-AO Spine Complications Classification system to better allow surgeons to study and understand the clinical implications of complications. The potential impact of this system is especially important given the current focus on value-based care and the increasing incidence of ASD surgery.8

Although this study was not designed to detect differences in ability of research coordinators and surgeons to capture event data, our results suggest that research coordinators can utilize the complications system with accuracy comparable to that of surgeons. This is significant, as the ISSG-AO Spine Complications Classification System may allow nonclinicians to accurately compile complication data, ultimately increasing the proportion of patients for whom complications data are recorded. This can facilitate further research into ASD complications, including investigating the impact of specific complication categories and intervention severity on costs and outcomes, while also assisting with value analyses, preoperative planning, and patient counseling. Data quality and entry can be further improved by applying the system as a logic-based entry form, limiting the data fields to simplify data entry and minimize missing data. This is especially significant as missing data accounted for a large proportion of incorrect answers in the neurologic category.

The validation study had additional useful findings. As accuracy improved with reader experience, cases used in this study may serve as a training set before real-world application of the system. Data obtained from this study may also be used as an audit tool to assess fidelity of data in future applications of the system.

This study has important limitations. Results were non-empirical, system performance was not directly compared to previous classification systems, and we did not attempt to use the system to predict outcomes. Furthermore, cases were simple with an average of two complications per case, and were analyzed under idealized conditions with descriptions...
provided to event readers, rather than requiring readers to search for details in the electronic medical record. However, use of such cases for initial validation of the system minimized potential confounding effects, such as transcriptional errors in real clinical documentation. Use of these cases also ensured that event readers were assessing a wide variety of complication types. Additional limitations include a small overall number of event readers, and inadequate power to detect differences between reader types. Accuracy may have been biased by readers’ training backgrounds, as most were spine care providers. Despite potential limitations, initial results from the present study are necessary to justify future studies of the system in real clinical populations. Additional strengths include the large number of data points analyzed, high response rates of event readers, and assessment of response repeatability.

In conclusion, the system demonstrated good accuracy and repeatability among surgeons and research coordinators in capturing complication event type, timing, and details. Future studies should assess real-world application of the system in understanding complication impacts on ASD surgical outcomes and costs.

Key Points

- Complications are common in adult spinal deformity surgery.
- Previously, no Adult Spinal Deformity-specific complication classification system existed, limiting understanding of complications’ impact on costs and outcomes following adult spinal deformity surgery.
- The International Spine Study Group, AO Spine, and European Spine Study group utilized a Delphi method to develop a comprehensive classification system for adult spinal deformity surgery complications.
- The new complication classification system was validated, demonstrating both accuracy and repeatability in capturing complications.

Supplemental digital content is available for this article. Direct URL citations appearing in the printed text are provided in the HTML and PDF version of this article on the journal’s Web site (www.spinejournal.com).

References