Thoracolumbar Injury Classification and Severity Score in children: a reliability study

*Ross L. Dawkins, MD,1 Joseph H. Miller, MD,2 Omar I. Ramadan, MD,1 Michael C. Lysek, MD, MPH,1 Elizabeth N. Kuhn, MD,1 Brandon G. Rocque, MD,1 Michael J. Conklin, MD,3 R. Shane Tubbs, PhD, PA-C,4 Beverly C. Walters, MD, MSc, FRCSC,1 Bonita S. Agee, PhD,1 and Curtis J. Rozzelle, MD1

1Section of Pediatric Neurosurgery, Children’s of Alabama, Department of Neurosurgery, and 2Division of Orthopedic Surgery, Department of Surgery, University of Alabama at Birmingham, Alabama; 3UT Erlanger Neurosurgery, Chattanooga, Tennessee; and 4Seattle Science Foundation, Seattle, Washington

OBJECTIVE There are many classification systems for injuries of the thoracolumbar spine. The recent Thoracolumbar Injury Classification and Severity Score (TLICS) has been shown to be a reliable tool for adult patients. The aim of this study was to assess the reliability of the TLICS system in pediatric patients. The validity of the TLICS system is assessed in a companion paper.

METHODS The medical records of pediatric patients with acute, traumatic thoracolumbar fractures at a single Level 1 trauma center were retrospectively reviewed. A TLICS was calculated for each patient using CT and MRI, along with the neurological examination recorded in the patient’s medical record. TLICSs were compared with the type of treatment received. Five raters scored all patients separately to assess interrater reliability.

RESULTS TLICS calculations were completed for 81 patients. The mean patient age was 10.9 years. Girls represented 51.8% of the study population, and 80% of the study patients were white. The most common mechanisms of injury were motor vehicle accidents (60.5%), falls (17.3%), and all-terrain vehicle accidents (8.6%). The mean TLICS was 3.7 ± 2.8. Surgery was the treatment of choice for 33.3% of patients. The agreement between the TLICS-suggested treatment and the actual treatment received was statistically significant (p < 0.0001). The interrater reliability of the TLICS system ranged from moderate to very good, with a Fleiss’ generalized kappa (κ) value of 0.69 for the TLICS treatment suggestion among all patients; however, interrater reliability decreased when MRI was used to contribute to the TLICS. The κ value decreased from 0.73 to 0.57 for patients with CT only vs patients with CT/MRI or MRI only, respectively (p < 0.0001). Furthermore, the agreement between suggested treatment and actual treatment was worse when MRI was used as part of injury assessment.

CONCLUSIONS The TLICS system demonstrates good interrater reliability among physicians assessing thoracolumbar fracture treatment in pediatric patients. Physicians should be cautious when using MRI to aid in the surgical decision-making process.

https://thejns.org/doi/abs/10.3171/2017.7.PEDS1720

KEY WORDS thoracolumbar; fracture; score; children; reliability; spine; trauma
The TLICS system was developed with the intention of providing a clinically relevant classification system that would be not only descriptive, but also predictive of outcome and helpful for guidance of treatment decisions. Only 1 study considered the reliability of the scoring system in addition to its validity. The purpose of our study was to evaluate the interrater reliability of the TLICS system in a large cohort of pediatric patients. We assess the validity of the TLICS system in a companion paper.

Methods

Study Population

A retrospective review of all medical records from 2002 to 2013 that met study criteria was completed for pediatric patients (age 18 years and younger) with acute, traumatic thoracolumbar fractures at a Level 1 trauma center (Children’s of Alabama). Inclusion in the study required the presence of CT and/or MRI at the time of fracture assessment, in addition to a documented neurological examination. Exclusion criteria included pathological fractures, age-indeterminate fractures with no recent traumatic event at the time of assessment, incomplete clinical records, and evidence of use of the TLICS system during patient evaluation. Patient charts were reviewed for evidence of use of the TLICS system during the initial patient evaluation and subsequent care. Of note, traumatic spine injuries at our institution are typically managed by the neurosurgery service. Prior to the initiation of this study, the TLICS system was unfamiliar to the neurosurgery faculty. Treatment decisions were made using best clinical judgment for each case. Obviously, clinical judgment may have taken into account some of the TLICS parameters, but not as a formalized scoring system. However, it was not confirmed that previous faculty involved in the care of study patients were also unfamiliar with the TLICS system at the time of patient evaluations, although we believe it is safe to assume this is true. A total of 81 patients met the criteria for inclusion. Approval was obtained from the hospital’s IRB prior to review of the medical records and data collection.

Data Collection

A TLICS was calculated for each patient (Table 1). Neurosurgeons and an orthopedic surgeon at our institution completed the scoring. CT and/or MRI were used to assign the morphology and posterior ligamentous complex (PLC) components of the score. The PLC components include the ligamentum flavum, facet joint capsules, interspinous ligament, and supraspinous ligament. Imaging findings on CT that are considered indicative of PLC disruption include diastasis of the facet joints and widening of the interspinous space. For MRI, these same imaging findings are considered pertinent. In addition, high signal intensity on T2-weighted or short tau inversion recovery sequences in the region of the PLC components are considered to be suggestive of PLC disruption, and are not seen on CT. The neurological examination at the time of initial assessment, as recorded in the patient’s medical record, was used for the neurological status component of the score. A total score was obtained by summing the scores of the 3 components. According to the first published description of the TLICS system, the suggested treatment is conservative therapy for a score of 1–3, either surgery or conservative therapy for a score of 4, and surgery for a score of 5 or greater.

The treatment received by each patient was classified as either surgical or conservative for comparison with the calculated TLICS scores and the treatment recommendation. Any surgery for decompression and/or mechanical stabilization involving the fracture site was classified as surgical treatment. Conservative treatments included use of an external orthotic brace with scheduled follow-up, no brace with scheduled follow-up, and no scheduled follow-up. Patients with failure of conservative treatment requiring delayed surgery were kept in the conservative treatment group for analysis of intention to treat. Demographic, mechanism of injury, and outcome data were also collected for each patient.

To assess interrater reliability, 5 physicians independently assigned a TLICS score for each patient. The physicians included 1 junior neurosurgery resident, 2 senior neurosurgery residents, 1 neurosurgery faculty member with 2 years of postfellowship experience, and 1 orthopedic faculty member with 23 years of postfellowship experience, all from the University of Alabama Birmingham/Children’s of Alabama at the time of completion of the ratings. Raters were provided with the medical record number of each patient in addition to the date of the CT and/or MRI to be reviewed. Images were reviewed in our institution’s image viewing system, which can be accessed separately from the other portions of the patient’s medical record. Images obtained at the time of the initial patient assessment were selected for scoring of the fracture morphology and PLC disruption components of the TLICS.
The neurological examination for each patient at the time of initial assessment was provided to the raters for scoring of the neurological status component of the TLICS. No other medical record data, including treatment or outcome information, was provided to the raters during assignment of the TLICS scores. Furthermore, none of the raters were involved in the treatment of any of the study patients.

Statistical Analysis

Differences between patients receiving conservative and surgical treatments were analyzed for categorical demographic, mechanism of injury, and TLICS subcategory score variables using a chi-square test (or Fisher’s exact test when assumptions were not met). Age and TLICS total score comparisons between patients receiving conservative and surgical treatments were tested as continuous variables using an independent t-test and a Mann-Whitney U-test where appropriate. For comparisons of the actual treatment to the TLICS suggested treatment, patients with a TLICS of 4 were excluded. A score of 4 is classified as neutral, so either surgical or conservative treatment is considered reasonable. Odds ratios (ORs) with 95% confidence intervals (95% CI) were calculated for the odds of actual treatment given the TLICS suggested treatment for different scenarios.

For interrater reliability analysis using 5 raters, Fleiss’ generalized kappa (κ), Kendall’s coefficient of concordance (W), and the intraclass correlation coefficient (ICC) were used to assess agreement using the TLICS scoring system. The nominal variable for TLICS treatment suggestion was assessed using only Fleiss’ generalized κ for multiple raters because this parameter is considered appropriate for nominal data.6-7 The TLICS parameters (morphology, PLC, neurological status) were treated as ordinal data; Kendall’s W is an interrater reliability coefficient particularly suited for the ordinal nature of the TLICS parameters.8-10 TLICS total scores were analyzed as continuous data.

Generalized κ estimates between 0.81 and 1.00 are interpreted as almost perfect agreement, κ values between 0.61 and 0.80 as substantial, κ values between 0.41 and 0.60 as moderate, κ values between 0.21 and 0.40 as fair, and κ values 0.20 and lower as poor.11 Extremely strong interrater reliability is inferred for Kendall’s W coefficients between 0.71 and 0.90, strong reliability is inferred for coefficients between 0.51 and 0.70, moderate reliability is inferred for coefficients from 0.31 to 0.50, and less than 0.30 is interpreted as weak interrater reliability.10

Using the Shrout and Fleiss ICC Model 2 (ICC2,1) analyzed for absolute agreement, ICC estimates greater than 0.75 are interpreted as excellent agreement, ICC estimates between 0.40 and 0.75 as fair to good agreement, and ICC estimates less than 0.40 as poor agreement.1,19,31

Results

Demographics and Clinical Characteristics

Of the 81 study patients, 54 (66.7%) were treated conservatively and 27 (33.3%) were treated surgically (Table 2). The mean patient age was 10.9 years. Girls comprised...
51.8% of the patients, and 80.2% of all patients were white. Motor vehicle accident (MVA) was the most common mechanism of injury. Patients who received surgical treatment were significantly older than those who received conservative treatment (12.1 vs 10.2 years, respectively; \( p = 0.04 \)). There were no significant differences in sex, race, or mechanism of injury between the conservative and surgical treatment groups.

The mean number of follow-up visits was 2.9 for the conservative group and 3.9 for the surgical group. The mean length of follow-up was 298 days for the conservative group and 455 days for the surgical group. Of those in the conservative group, 66.7% were treated with bracing. Conservative therapy was unsuccessful in 1 patient (1.2%), who required surgical intervention. This patient had a lumbar Chance fracture and developed progressive focal kyphosis and instability 18 months after the injury, requiring dorsal internal fixation and fusion. Of those in the surgical group, 2 patients (7.4%) had anterior approaches and 25 (92.6%) had posterior approaches. All posterior approaches included internal fixation. Four patients (14.8%) required second operations, and all had undergone posterior approaches. Two of these patients, having suffered significant spinal cord injuries at the time of their accidents, developed surgical site infections requiring reoperation. One patient developed progressive junctional kyphosis and scoliosis, requiring extension of the fusion construct 7 months after the initial surgery. The fourth patient had a postoperative radiculopathy requiring a second operation for screw revision 5 days after the initial surgery.

**TLICS**

Using the TLICSs assigned by one of the authors (R.L.D.), the mean TLICS among all patients was 3.7 ± 2.8. Most patients were scored as having had a fracture with compressive morphology (43.2%), no evidence of injury to the PLC (56.8%), and no associated neurological deficits on examination (88.9%). However, patients in the surgical group had a significantly higher mean TLICS (6.0 vs 2.5, \( p < 0.0001 \)). In addition, they were significantly more likely to have received higher scores for each individual component of the TLICS (Table 3).

For comparisons between actual treatment and TLICS-suggested treatment, 7 patients with a neutral score of 4 were excluded from analyses. The treatment suggested by the TLICS scoring system matched the actual treatment in 79.7% of cases. Patients who received surgical treatment were more likely to have TLICSs suggestive of surgical treatment (OR 17.3, 95% CI 4.9–61.4). The agreement between actual treatment and suggested treatment was statistically significant (\( p < 0.0001 \); Table 4).

**Interrater Reliability**

For the assessment of interrater reliability, all 3 statistical methods were used for each parameter in the TLICS and the total TLICS, and Fleiss’s generalized \( \kappa \) was used for the TLICS treatment suggestion nominal variable. All values ranged from moderate to very good. Consideration was given to the effect of imaging modality (CT and MRI) on agreement. Concerning the type of imaging reviewed for each patient, 42 patients had CT only, 35 patients had CT and MRI, and 4 patients had MRI only. For those patients who had MRI only, a sample size of 4 is small and therefore not very robust for detecting differences in variance. In fact, it was impossible to detect the ICC for some parameters due to no variance across the raters (see Table 4).

**TABLE 3. Breakdown of TLICSs by treatment received**

<table>
<thead>
<tr>
<th>Score Total Cohort</th>
<th>Conservative Treatment</th>
<th>Surgical Treatment</th>
<th>( p ) Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total TLICS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>3.7 ± 2.8</td>
<td>2.5 ± 2.2</td>
<td>6.0 ± 2.3</td>
</tr>
<tr>
<td>Median (range)</td>
<td>2.0 (1.0–9.0)</td>
<td>1.0 (1.0–7.0)</td>
<td>7.0 (2.0–9.0)</td>
</tr>
<tr>
<td>Morphology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>35 (43.2)</td>
<td>34 (63.0)</td>
<td>1 (3.7)</td>
</tr>
<tr>
<td>2</td>
<td>18 (22.2)</td>
<td>10 (18.5)</td>
<td>8 (29.6)</td>
</tr>
<tr>
<td>3</td>
<td>5 (6.2)</td>
<td>2 (3.7)</td>
<td>3 (11.1)</td>
</tr>
<tr>
<td>4</td>
<td>23 (28.4)</td>
<td>8 (14.8)</td>
<td>15 (55.6)</td>
</tr>
<tr>
<td><strong>PLC</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>46 (56.8)</td>
<td>39 (72.2)</td>
<td>7 (25.9)</td>
</tr>
<tr>
<td>2</td>
<td>6 (7.4)</td>
<td>5 (9.3)</td>
<td>1 (3.7)</td>
</tr>
<tr>
<td>3</td>
<td>29 (35.8)</td>
<td>10 (18.5)</td>
<td>19 (70.4)</td>
</tr>
<tr>
<td><strong>Neurological status</strong></td>
<td></td>
<td></td>
<td>(0.0002)</td>
</tr>
<tr>
<td>0</td>
<td>72 (88.9)</td>
<td>53 (98.2)</td>
<td>19 (70.4)</td>
</tr>
<tr>
<td>2</td>
<td>6 (7.4)</td>
<td>0 (0)</td>
<td>6 (22.2)</td>
</tr>
<tr>
<td>3</td>
<td>3 (3.7)</td>
<td>1 (1.8)</td>
<td>2 (7.4)</td>
</tr>
</tbody>
</table>

All data given as number of patients (%) unless otherwise indicated.

* Bivariate analysis; \( p < 0.05 \) is considered statistically significant. For the categorical variables morphology, PLC, and neurological status, statistical significance was assessed using chi-square and Fisher’s exact tests. The continuous variable total TLICS was tested for statistical significance using the Mann-Whitney \( U \)-test. Boldface type indicates statistical significance.
Supplemental Table 1d). For this reason, the patients with MRI only were grouped together with the patients who had both CT and MRI to assess how the use of MRI affected interrater reliability when compared with the use of CT only. Overall, interrater reliability was worse when MRI was employed and used in calculating the TLICS. Using Kendall’s coefficient of concordance (W) and the ICC, the interrater reliability for the total TLICS was extremely strong (W = 0.87) and excellent (ICC = 0.83) for all patients, extremely strong (W = 0.80) and excellent (ICC = 0.88) for patients with CT only, and extremely strong (W = 0.79) and fair to good (ICC = 0.70) for patients with MRI. Using Fleiss’s generalized k the interrater reliability for the TLICS treatment suggestion was good (k = 0.69) for all patients, good (k = 0.73) for patients with CT only, and moderate (k = 0.57) for patients with MRI (Table 5; see Supplemental Tables 1a–c for reliability statistics on each component of the TLICS score, separated by imaging type). When interrater reliability estimates by the fac -ulty and residents were compared, reliability was slightly higher for faculty estimates than for resident estimates (k = 0.77 vs 0.64, respectively, for TLICS treatment suggestion; Supplemental Table 1e).

Effect of Imaging Modality

The effect of imaging modality on agreement between the actual treatments received and the treatments suggested by the TLICS system was also analyzed. As noted previously, the agreement was statistically significant when all patients with nonneutral scores were considered (Table 4). The statistical significance was maintained for patients who had only CT scans (OR 46.5, 95% CI 5.4–397.6, p = 0.0002), but the agreement between actual treatment and suggested treatment was not as strong when an MRI scan was used to assess injury (OR 6.5, 95% CI 1.1–37.5, p = 0.03; Table 6). When only patients who were treated surgically were considered, the presence of an MRI did not affect the suggested treatment (p = 0.59; Table 7). However, when only patients who were treated conservatively were considered, MRI made it more likely for TLICS to suggest surgical treatment (p = 0.0003; Table 7).

Discussion

Among spine surgeons, no single classification system has been adopted as the optimal scale for evaluating thoracolumbar fractures. The Denis and Magerl-AO classifications are 2 commonly used scales. Each classifies fractures on the basis of morphology and mechanism of injury.4,18 However, neither scale emphasizes the importance of neurological status or provides a reliable definition of mechanical stability. Furthermore, studies have demonstrated poor to moderate inter- and intrarater reliability for both systems.3,4,16,18,30,32

The recently introduced TLICS classification system was developed to provide a clinically relevant scale that could stratify injury severity and guide treatment decisions.8,18 Because the majority of spine trauma occurs in the adult population,9 it stands to reason that the TLICS system, like its predecessors, was developed for use in adult thoracolumbar fractures. Since its introduction in 2005, numerous studies have demonstrated good reliability of this system in samples of adult patients.12,17,23,24 However, the reliability of the TLICS in adults cannot be extrapolated to children, for multiple reasons. First, the pediatric spinal column is anatomically and biomechanically quite different from the adult spinal column, most notably in prepubescent children; its increased water and cartilage contents provide for greater ligamentous flexibility and elasticity.2,12,25,32 The facet joints are shallower and more horizontal, making the pediatric spinal column tolerant of translational and flexion/extension forces.1,32 Thus, flexion/distraction and rotational injuries are more often seen in adolescents, whereas younger children present with compression injuries.32 Furthermore, the pediatric spinal column is exposed to a different array of traumatic events. While mechanisms of injury vary among pediatric age groups, MVAs are generally the most common. Other common mechanisms include falls in younger children and higher-risk activities such as all-terrain vehicles, mo-

### TABLE 4. Comparison of TLICS-suggested treatment to actual treatment received

<table>
<thead>
<tr>
<th>Actual Treatment</th>
<th>OR (95% CI)</th>
<th>p</th>
<th>Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>TLICS suggested</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservative</td>
<td>17.3 (4.9–61.4)</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>Surgical</td>
<td>2.8 (1.2–6.3)</td>
<td>0.024</td>
<td></td>
</tr>
</tbody>
</table>

Patients with a neutral TLICS of 4 excluded (n = 74 in this comparison). All data given as number of patients (%) unless otherwise indicated.

* p < 0.05 is considered statistically significant; statistical significance for differences in proportions were tested using chi-square and Fisher’s exact tests. Boldface type indicates statistical significance.

### TABLE 5. TLICS interrater reliability estimates by imaging modality

<table>
<thead>
<tr>
<th>Parameter</th>
<th>All Patients</th>
<th>CT Only, n = 42</th>
<th>CT/MRI or MRI Only, n = 39</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>κ</td>
<td>W</td>
<td>ICC</td>
</tr>
<tr>
<td>Morphology</td>
<td>0.60</td>
<td>0.73</td>
<td>0.67</td>
</tr>
<tr>
<td>PLC</td>
<td>0.60</td>
<td>0.82</td>
<td>0.76</td>
</tr>
<tr>
<td>Neurological status</td>
<td>0.87</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>Total TLICS</td>
<td>0.50</td>
<td>0.87</td>
<td>0.83</td>
</tr>
<tr>
<td>TLICS treatment</td>
<td>0.69</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Three separate statistical estimates of interrater reliability are presented for the application of the TLICS scoring system to patients at Children’s of Alabama by 5 raters. All p values presented for κ, W, and ICC probabilities are < 0.0001.
The distribution of mechanisms of injury in our study was similar to distributions reported in previously published works, with most injuries resulting from MVAs. Clearly, pediatric thoracolumbar fractures are a distinct pathology and their evaluation using the TLICS system must be assessed independently. Two studies to date have applied the TLICS system to a pediatric patient population. Savage et al. included 20 spine surgeons who each scored 20 pediatric thoracolumbar fractures. They found moderate interrater reliability for the total TLICS score, with greater agreement for operative versus nonoperative treatment recommendations based on the TLICS.27 However, the unweighted Cohen coefficient used in the study by Savage et al. is not as robust a statistic as Fleiss’s generalized $\kappa$ for multiple raters, Kendall’s $W$ coefficient for ordinal data, or the ICC for ordered, Likert-type, or continuous data. Furthermore, a potential conflict of interest stems from the inclusion of the original authors of the TLICS system, which is the case in numerous studies assessing TLICS in adults.8,15,21–24,26 Sellin et al. published the second study on TLICS in children,26 but they did not discuss or assess the reliability of the TLICS in pediatrics, and made no reference to the only reported reliability statistics on TLICS in pediatrics as found in the study by Savage et al. This is of consequence because reliability is a necessary condition for validity.11 Furthermore, the raters in this study were not blinded to treatment.

Our objective was to assess the reliability of the TLICS scoring system in a large cohort of pediatric patients evaluated at a Level I trauma center for traumatic thoracolumbar fractures. The interrater reliability of the TLICS system was assessed as good to excellent for the total TLICS, and moderate to very good for the TLICS treatment suggestion, depending on the choice of statistical method.

We hypothesized that MRI would be useful for clarifying the presence of disruption of the PLC when this was not clear on CT, thereby increasing the interrater reliability. It must be noted that the TLICS system was developed for use with CT. By including MRI, we applied the system in a manner that was not originally intended. However, MRI was used in nearly half of our patients, and excluding those patients would have significantly decreased the size of our study population. Furthermore, by including MRI, we discovered an interesting association between this imaging technique and the reliability of the TLICS system in pediatrics. The interrater reliability was actually worse among patients who underwent MRI. Thus, our hypothesis was false. While we are not sure how to explain this finding, it is likely that the raters assessed imaging parameters differently on the MRI of a patient who had a negative or equivocal PLC score on CT, so some raters increased the PLC score and some kept it the same. Also, it is likely that MRI was more often obtained in cases in which the surgical decision-making process was not clear from CT alone. This would result in an association between the decreased reliability of the TLICS system and the use of MRI. The negative effect of MRI on the reliability of PLC assess-

### TABLE 6. Actual treatment versus TLICS recommendation by imaging modality

<table>
<thead>
<tr>
<th>Treatment</th>
<th>CT Only, n = 41</th>
<th>CT/MRI or MRI Only, n = 33</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conservative Treatment</td>
<td>Surgical Treatment</td>
</tr>
<tr>
<td>TLICS suggested</td>
<td>46.5 (5.4–397.6)</td>
<td>0.0002</td>
</tr>
<tr>
<td>Conservative</td>
<td>31 (93.9)</td>
<td>2 (25.0)</td>
</tr>
<tr>
<td>Surgical</td>
<td>2 (6.1)</td>
<td>6 (75.0)</td>
</tr>
</tbody>
</table>

Patients with a neutral TLICS score of 4 excluded. All data given as number of patients (%) unless otherwise indicated.

*p < 0.05 is considered statistically significant; statistical significance for differences in proportions were tested using chi-square and Fisher’s exact tests. Boldface type indicates statistical significance.

### TABLE 7. TLICS score recommendation by imaging modality for surgically and conservatively treated patients

<table>
<thead>
<tr>
<th>Treatment</th>
<th>CT Only</th>
<th>CT/MRI or MRI Only</th>
<th>OR (95% CI)</th>
<th>p Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgical</td>
<td>No. of patients</td>
<td>8</td>
<td>15</td>
<td>2.2 (0.2–19.3)</td>
</tr>
<tr>
<td>TLICS suggested</td>
<td>Conservative</td>
<td>2 (25.0)</td>
<td>2 (13.3)</td>
<td>6 (75.0)</td>
</tr>
<tr>
<td></td>
<td>Surgical</td>
<td>15.5 (2.8–85.1)</td>
<td>0.0003</td>
<td></td>
</tr>
<tr>
<td>Conservative</td>
<td>No. of patients</td>
<td>33</td>
<td>18</td>
<td>31 (93.9)</td>
</tr>
<tr>
<td>TLICS suggested</td>
<td>Conservative</td>
<td>2 (6.1)</td>
<td>9 (50.0)</td>
<td></td>
</tr>
</tbody>
</table>

Patients with a neutral TLICS score of 4 excluded. All data given as number of patients (%) unless otherwise indicated.

*p < 0.05 is considered statistically significant; statistical significance for differences in proportions were tested using Fisher’s exact test. Boldface type indicates statistical significance.
ment is actually supported in the literature, which demonstrates that MRI has only poor to moderate reliability when the integrity of the PLC is assessed in adults.14,28

Given the interesting effect of imaging modality on inter-rater reliability, we surmised that MRI might also affect the significance of the agreement between the actual treatment received and the treatment suggested by the TLICS system. As noted previously, we found that the agreement remained significant when patients underwent CT only, but was not as significant when they had MRI. We gained insight into this relationship by looking at the association of imaging with conservatively treated versus surgically treated patients. Here we found that patients who were treated conservatively were more likely to have a TLICS suggesting surgical treatment when MRI was used to calculate the score. This tells us that MRI negatively affects the agreement between actual and suggested treatment by making conservatively treated patients more likely to receive a suggested treatment of surgery.

One study in adults demonstrated that MRI has poor specificity for identifying PLC disruption in thoracolumbar injuries.34 Assuming this holds true for our study population, the most likely explanation for our findings is that MRI made it more likely for a rater to give a false-positive rating for disruption of the PLC in a patient who went on to be treated conservatively. This caused a decrease in the agreement between actual treatment and suggested treatment when MRI was used. Taking this into account, along with the knowledge that MRI worsens inter-rater reliability of the TLICS system, surgeons must exercise caution when MRI is used to assist in the treatment decision, as it could sway the decision in the direction of surgical treatment in a patient who would be likely to tolerate conservative treatment.

Limitations
Our study is naturally limited by its retrospective nature. The only objective measurement of poor outcome was the requirement for repeat surgery in the surgical group, or a first-time surgery for those who were treated unsuccessfully with conservative therapy. There were no objective quality of life outcomes for analysis. The retrospective chart review could not prove that each patient was treated without use of the TLICS system. However, the concept of the TLICS system was unfamiliar to the current neurosurgery and orthopedic faculty prior to initiation of this study. Thus, although a potential limitation, we believe it is safe to say that a TLICS was not used for treatment decisions in the study patients. Another study limitation is that we were unable to have a study population with a similar imaging evaluation. There were 4 patients without CT scans, which is the gold standard for evaluation of fracture morphology. Also, scoring systems with neutral scores present an analytical dilemma, given that the statistical results can be swayed by how these scores are considered. Therefore, we chose to exclude TLICSs of 4 for comparisons of actual treatment to suggested treatment.

Conclusions
The TLICS has proven to be a reliable tool for assessing adult traumatic thoracolumbar fractures. However, its reliability has not been adequately assessed in a pediatric patient population. This study demonstrates that the TLICS system has good inter-rater reliability when applied to children. However, surgeons should not overestimate the value of MRI when making surgical decisions, as it decreases the reliability of the TLICS system.

Acknowledgments
Joseph H. Miller, MD, completed work on this study as a Stephens Scholar. Elizabeth N. Kuhn, MD, completed work on this study as a Kaul Foundation Clinical Research Scholar.

References
17. Lewkonia P, Paulucci EO, Thomas K: Reliability of the thoracolumbar injury classification and severity score and com-
18. Magerl F, Aebi M, Gertzbein SD, Harms J, Nazarian S: A
19. McGraw KO, Wong SP: Forming inferences about some in-
traclass correlation coefficients. Psychol Methods 1:30–46,
1996
20. Parent S, Dimar J, Dekutoski M, Roy-Beaudry M: Unique
features of pediatric spinal cord injury. Spine (Phila Pa
PG, et al: Thoracolumbar spine trauma classification: the
Thoracolumbar Injury Classification and Severity Score
system and case examples. J Neurosurg Spine 10:201–206,
2009
22. Patel AA, Vaccaro AR: Thoracolumbar spine trauma classifi-
23. Patel AA, Whang PG, Brodke DS, Agarwal A, Hong J,
trauma classification systems. Indian J Orthop 41:322–326,
2007
Bono CM, et al: Agreement between orthopedic surgeons
and neurosurgeons regarding a new algorithm for the treat-
ment of thoracolumbar injuries: a multicenter reliability
89 (Suppl 1):98–107, 2007
26. Rihn JA, Anderson DT, Harris E, Lawrence J, Jonsson H,
Wilsey J, et al: A review of the TLICS system: a novel, user-
friendly thoracolumbar trauma classification system. Acta
27. Savage JW, Moore TA, Arnold PM, Thakur N, Hsu WK,
Patel AA, et al: The reliability and validity of the thoraco-
lumbar injury classification system in pediatric spine trauma.
JA, Rechting GR, et al: Interrater reliability of identifying
indicators of posterior ligamentous complex disruption when
plain films are indeterminate in thoracolumbar injuries. J
29. Sellin JN, Steele WJ III, Simpson L, Huff WX, Lane BC,
Chern JJ, et al: Multicenter retrospective evaluation of the
validity of the Thoracolumbar Injury Classification and
Severity Score system in children. J Neurosurg Pediatr
18:164–170, 2016
30. Sethi MK, Schoenfeld AJ, Bono CM, Harris MB: The evolu-
tion of thoracolumbar injury classification systems. Spine J
9:780–788, 2009
31. Shrout PE, Fleiss JL: Intraclass correlations: uses in assess-
ing rater reliability. Psychol Bull 86:420–428, 1979
32. Slotkin JR, Lu Y, Wood KB: Thoracolumbar spinal trauma in
33. Vaccaro AR, Lehman RA Jr, Hurlbert RJ, Anderson PA, Har-
is M, Hedlund R, et al: A new classification of thoracolum-
bar injuries: the importance of injury morphology, the in-
tegrity of the posterior ligamentous complex, and neurologic
34. Vaccaro AR, Rihn JA, Saravanja D, Anderson DG, Hilibrand
complex of the thoracolumbar spine: a prospective evaluation
of the diagnostic accuracy of magnetic resonance imaging.
35. Wood KB, Khanna G, Vaccaro AR, Arnold PM, Harris MB,
Mebbod AA: Assessment of two thoracolumbar fracture clas-
sification systems as used by multiple surgeons. J Bone Joint
Surg Am 87:1423–1429, 2005

Disclosures
The authors report no conflict of interest concerning the materi-
als or methods used in this study or the findings specified in this
paper.

Author Contributions
Conception and design: Dawkins, Miller, Tubbs, Walters, Roz-
zelle. Acquisition of data: Dawkins, Miller, Ramadan, Lysek,
Kuhn, Rocque, Conklin. Analysis and interpretation of data:
Dawkins, Miller, Ramadan, Kuhn, Rocque, Conklin, Walters,
Agee, Rozzelle. Drafting the article: Dawkins, Miller, Agee.
Critically revising the article: Dawkins, Miller, Ramadan, Kuhn,
Rocque, Conklin, Walters, Agee, Rozzelle. Reviewed submitted
version of manuscript: all authors. Approved the final version
of the manuscript on behalf of all authors: Dawkins. Statistical
analysis: Dawkins, Miller, Agee. Administrative/technical/mate-
rial support: Dawkins, Miller, Walters, Rozzelle. Study
supervision: Dawkins, Miller, Walters, Rozzelle.

Supplemental Information
Online-Only Content
Supplemental material is available with the online version of the
article.

Supplemental Table 1a–e. https://thejns.org/doi/suppl/10.3171/
2017.7.PEDS1720.

Correspondence
Ross L. Dawkins: University of Alabama at Birmingham, AL.
ross.dawkins.md@gmail.com.