Using Equipoise to Determine the Radiographic Characteristics Leading to Agreement on Best Treatment for Thoracolumbar Burst Fractures Without Neurologic Deficits

Charlotte Dandurand, MD, MSc[1], Marcel F. Dvorak, MD, MBA[1], Olesja Hazenbiller, MS[2], Richard J. Bransford, MD[3], Klaus J. Schnake, MD[4,5], Alexander R. Vaccaro, MD, PhD, MBA[6], Lorin M. Benneker, MD[7], Emiliano Vialle, MD[8], Gregory D. Schroeder, MD[6], Shanmuganathan Rajasekaran, MD, PhD[9], Mohammad El-Skarkawi, MD, PhD[10], Rishi M. Kanna, MS, MRCS, FNB[11], Mohamed M. Aly, MD, PhD, MRCS[12,13], Martin Holas, MD[14], Jose A. Canseco, MD, PhD[6], Sander Muijs, MD, PhD[15], Eugen C. Popescu, MD, PhD[16], Jin Wee Tee, BMedSci, MBBS, MD[17], Gaston Camino-Willhuber, MD[18], Andrei Fernandes Joaquim, MD, PhD[19], Ory Keynan, MD[20], Harvinder Singh Chhabra, MBBS, MS[21], Sebastian Bigdon, MD[22], Ulrich Spiegel, MD[23], and Cumhur F. Öner, MD, PhD[15]

[1] Combined Neurosurgical and Orthopedic Spine Program, Vancouver General Hospital, University of British Columbia, Vancouver, BC, Canada
[2] AOSpine International, Davos, Graubunden, Switzerland
[3] Department of Orthopaedics and Sports Medicine, Harborview Medical Center, University of Washington, Seattle, WA, USA
[4] Center for Spinal and Scoliosis Surgery, Malteser Waldkrankenhaus St. Marien, Erlangen, Germany
[5] Department of Orthopedics and Traumatology, Paracelsus Private Medical University Nuremberg, Nuremberg, Germany
[6] Department of Orthopaedic Surgery, Rothman Institute, Thomas Jefferson University Hospital, Philadelphia, PA, USA
[7] Spine Unit, Sonnenhof Spital, University of Bern, Bern, Switzerland
[8] Cajuru Hospital, Catholic University of Paraná, Curitiba, Brazil
[9] Department of Orthopaedics and Spine Surgery, Ganga Hospital, Coimbatore, India
[10] Department of Orthopaedic and Trauma Surgery, Faculty of Medicine, Assiut University, Egypt
[11] Spine Department of Orthopaedics and Spine Surgery, Ganga Hospital, Coimbatore, India
[12] Department of Neurosurgery, Prince Mohammed Bin Abdulaziz Hospital, Riyadh, Saudi Arabia
[13] Department of Neurosurgery, Mansoura University, Mansoura, Egypt
[14] Klinika Úrazovej Chirurgie SZU, FNsP F.D.Roosevelta, Banská Bystrica, Slovakia
[15] University Medical Centers, Utrecht, the Netherlands
[16] Prof. Dr. N. Oblu" Emergency Hospital, Iasi, Romania
[17] Department of Neurosurgery, The Alfred Hospital, National Trauma Research Institute (NTRI), Melbourne, VIC, Australia
[18] Orthopaedic and Traumatology Department, Institute of Orthopedics "Carlos E. Ottolenghi" Hospital Italiano de Buenos Aires, Buenes Aires, Argentina
[19] Neurosurgery Division, Department of Neurology, State University of Campinas, Campinas-Sao Paulo, Brazil
[20] Rambam Health Care Campus, Haifa, Israel
[21] Sri Balaji Action Medical Institute, New Delhi, India
[22] Department of Orthopaedic Surgery and Traumatology, Inselspital, University Hospital, University of Bern, Bern, Switzerland
[23] Department of Orthopaedics, Trauma Surgery and Plastic Surgery, University of Leipzig, Leipzig, Germany

Corresponding Author:
Charlotte Dandurand, MD, MSc, Combined Neurosurgical & Orthopaedic Spine Program, Blusson Spinal Cord Center, University of British Columbia, 6th Floor, 818 West 10th avenue, Vancouver, BC V5Z 1M9, Canada.
Email: charlotte.dandurand@vch.ca

Abstract

Study Design: Retrospective analysis of prospectively collected data.

Objectives: Our goal was to assess radiographic characteristics associated with agreement and disagreement in treatment recommendation in thoracolumbar (TL) burst fractures.

Methods: A panel of 22 AO Spine Knowledge Forum Trauma experts reviewed 183 cases and were asked to: (1) classify the fracture; (2) assess degree of certainty of PLC disruption; (3) assess degree of comminution; and (4) make a treatment recommendation. Equipoise threshold used was 77% (77:23 distribution of uncertainty or 17 vs 5 experts). Two groups were created: consensus vs equipoise.

Results: Of the 183 cases reviewed, the experts reached full consensus in only 8 cases (4.4%). Eighty-one cases (44.3%) were included in the agreement group and 102 cases (55.7%) in the equipoise group. A3/A4 fractures were more common in the equipoise group (92.0% vs 83.7%, $P < .001$). The agreement group had higher degree of certainty of PLC disruption [35.8% (SD 34.2) vs 27.6 (SD 27.3), $P < .001$] and more common use of the M1 modifier (44.3% vs 38.3%, $P < .001$). Overall, the degree of comminution was slightly higher in the equipoise group [47.8 (SD 20.5) vs 45.7 (SD 23.4), $P < .001$].

Conclusions: The agreement group had a higher degree of certainty of PLC injury and more common use of M1 modifier (more type B fractures). The equipoise group had more A3/A4 type fractures. Future studies are required to identify the role of comminution in decision making as degree of comminution was slightly higher in the equipoise group.

Introduction

Thoracolumbar (TL) 'burst type' fractures account for 45% of all major thoracolumbar injuries.[1-4] Despite being a frequent pathology, which is regularly seen and managed by spine surgeons, there is still no consensus on the indications for surgical treatment in cases without neurologic involvement. The current literature offers mixed and inconclusive results with various recommendations and multiple treatment algorithms.[1,5-8] Thus, treatment of TL 'burst type' fractures is an example of 'equipoise' where there is uncertainty within the expert community on the optimal treatment approach.[9] Additionally, thoracolumbar burst fractures are an heterogenous patient population that can hamper the comparability among reported series. To find clarity in this critical clinical dilemma, an alternative method has been proposed: the equipoise methodology. The core principle of the equipoise methodology is that the inclusion of each patient is based on the presence of uncertainty as to the best management among expert reviewers.

Although the concept of 'equipoise' has been frequently utilized in other contexts, the definition of 'equipoise' in terms of the threshold of disagreement has not been clearly defined. Medical ethics researchers suggested a trial to be unethical when agreement among experts is above 70% or 80%.[10,11] Commensurate with using equipoise methodology to determine inclusion for prospective clinical trials, another option is to use the methodology to better understand what leads surgeons to agree or disagree on optimal management of TL burst fractures without neurological deficits. To achieve this, we aimed to explore which fracture characteristics are associated with various thresholds of agreement and disagreement on treatment. In other words, we aimed to assess which features lead surgeons to achieve total agreement, partial agreement or total disagreement on the best management of a particular case. This will be key in understanding what drives surgeon's decision-making and identifying the sources of disagreement in the controversial topic of TL burst fractures.

The goal of this study was to assess the radiographic characteristics associated with various thresholds of agreement in recommending either surgical or non-surgical treatment in TL burst fractures. We aimed to assess the association of agreement and equipoise with fracture classification, the degree of certainty of PLC (Posterior Ligamentous Complex) injury, the use of the AO Spine Thoracolumbar Injury Classification M1 modifier (used to designate fractures with an indeterminate injury to the tension band based on spinal imaging or clinical examination) and the degree of vertebral body comminution.

Methods

The detailed methodology is available in the article of Dandurand et al "Understanding Decision Making as it Influences Treatment in Thoracolumbar Burst Fractures Without Neurological Deficit: Conceptual Framework and Methodology" in this focus issue. The AO Spine Knowledge Forum Trauma completed consent and recruitment for a multicenter prospective observational study of TL Fractures; the Spine A3/A4 study.[12] Each enrolling center obtained local approval from their local institutional review board. The baseline CT scans

and conventional radiographs of 183 patients were available for this study. All patients were neurologically intact and had injuries between T11 and L2.

The 22 Spine Trauma experts all with extensive experience in management of spinal trauma were recruited from the AO Spine Knowledge Forum Trauma (KF Trauma). Each member of the expert panel independently reviewed the DICOM images of the 183 TL fracture cases and were asked to classify each injury based on the latest AO TL Injury Classification System, assess the degree of certainty of PLC disruption and the degree of comminution. The reviewers were also asked to indicate the use of the M1 modifier or not. The M1 modifier is used to designate a fracture with an indeterminate injury to the tension band based on spinal imaging. Finally, they were asked to recommend treatment – either surgical or non-operative, which specific type of treatment and finally asked how confident they were in this recommendation. These experts were agnostic to the actual treatment that the patient received within the Spine TL A3/A4 Study and were also agnostic to any results of the TL Spine A3/A4 study.

Subgroups were created based on various agreement thresholds. The subgroup thresholds ranged from total consensus on proposed treatment to total disagreement. Total agreement means that all 22 experts recommended surgery, or all 22 experts recommending non-surgical management. Total disagreement is defined as half of the experts recommending surgery and half recommending non-surgical management (11 vs 11 expert reviewers).

The cases were divided into 2 groups: (1) agreement group and (2) equipoise group. The agreement group was defined as 18 experts or more agreeing on treatment and the equipoise group was defined as 17 experts or less agreeing on treatment.

This corresponds to an equipoise level of 77% (77:23 distribution of uncertainty). This is in accordance with previous literature.[10,11,13,14] A recent randomized clinical trial on surgical approach for cervical spondylotic myelopathy set the level of agreement to achieve equipoise at 80%.[13]

Statistical Analysis

Frequency tables were produced for the distribution of each injury type for each member of the expert panel. Fleiss Multirater Kappa scores were produced for analyzing the agreement of all expert panel raters for both injury classification and treatment. Kappa results were interpreted as follows: values ≤0 indicating no agreement, .01-.20 as none to slight, .21-.40 as fair, .41-.60 as moderate, .61-.80 as substantial, and .81-1.00 as strong.[15] Inter Class correlation coefficients were produced as a measure of reliability whenever data were continuous or ordinal. The intraclass correlation was interpreted as follows: values less than .5 indicating poor reliability, .5-.75 as moderate, .75-.9 as good and .9-1.0 as excellent.[16]

Associations of various imaging characteristics (comminution, PLC status), and treatment recommendations were analyzed through a process of multivariable regression analysis and development of predictive modeling equations. We furthermore employed multivariable logistic regression model (in 2 format of marginal model and mixed effect model) for making predictive models whenever it was necessary.

Results

Of the 183 cases reviewed, the experts reached full consensus in only 8 cases (4.4%) In 6 cases (3.2%), all the experts recommended surgery and in 2 cases (1.1%), there was consensus for non-surgical management.

When applying the defined equipoise threshold of 77%, 81 cases (44.3%) were included in the agreement group and 102 cases (55.7%) were included in the equipoise group.

Fracture Classification

The overall interrater reliability for the classification of the fracture was moderate in the agreement group (k = .441, 95% CI 0.430-.451, $P < .001$) as well as the equipoise group (k = .413, 95% CI 0.403-.424, $P < .001$). The intraclass correlation was good in the equipoise group (ICC = .893, 95% CI 0.859-.922, $P < .001$) and excellent in the agreement group (ICC = .948, 95% CI 0.929-.963, $P < .001$).

The distribution of fracture type by equipoise thresholds is represented in Table 1. Overall, A3/A4 fractures were more common in the equipoise group compared to the agreement group (92.0% vs 83.6%, respectively, $P < .001$). In comparison, the proportion of fractures classified as non-A (B1/B2/C) was 13.3% in the consensus group and 6.7% in the disagreement group.

The Posterior Ligamentous Complex

The degree of certainty of PLC disruption showed excellent reliability for both the equipoise group (ICC = .879, 95% CI 0.901-.938) and the agreement group (ICC = .958, 95% CI 0.943-.970, $P < .001$). Overall, the degree of certainty of PLC

Table 1. Fracture Type Distribution by Equipoise Thresholds.

	A0/A1/A2	A3/A4	B1/B2/C
100% agreement (n, %)	2 (1.1)	131 (74.9)	42 (24.0)
91% agreement (n, %)	20 (5.7)	253 (72.1)	78 (22.2)
86% agreement (n, %)	14 (4.0)	303 (86.1)	35 (9.9)
82% agreement (n, %)	11 (2.1)	466 (88.3)	51 (9.7)
77% agreement (n, %)	8 (2.1)	336 (89.8)	30 (8.0)
73% agreement (n, %)	3 (.7)	373 (89.7)	40 (9.6)
68% agreement (n, %)	9 (2.2)	363 (87.1)	45 (10.8)
64% agreement (n, %)	5 (1.2)	392 (93.8)	21 (5.0)
59% agreement (n, %)	5 (1.2)	386 (92.8)	25 (6.0)
55% agreement (n, %)	5 (1.1)	420 (95.7)	14 (3.2)
Total disagreement (n, %)	2 (1.6)	121 (95.3)	4 (3.1)

disruption was higher in the agreement group compared to the equipoise group [35.8% (SD 34.2) vs 27.6 (SD 27.3), $P <$ .001]. The degree of certainty of PLC injury distribution by equipoise thresholds is represented in Table 2. The mean degree of certainty of PLC disruption decreased from 46.62% in the complete consensus group to 22.48% in the total disagreement group.

The Use of M1 Modifier

The use of the M1 modifier showed slight reliability in the equipoise group (k = .091, 95% CI 0.078-.104, $P <$ .001) and fair reliability in the agreement group (k = .229, 95% CI 0.214-.243, $P <$ .001). The use of the M1 modifier was more common in the agreement group than the equipoise group (44.3% vs 38.3%, $P <$ .001). The use of M1 modifier distribution by equipoise thresholds is represented in Table 2. The M1 modifier was used in 51.7% in the consensus group and 27.3% in the disagreement group.

The Degree of Comminution

The degree of comminution showed excellent reliability for both the equipoise group (ICC = .947, 95% CI 0.930-.961) and the agreement group (ICC = .979, 95%CI 0.971-.985, $P <$ .001). Overall, the degree of comminution was slightly higher in the equipoise group compared to the agreement group [47.8 (SD 20.5) vs 45.7 (SD 23.4), $P <$ .001]. The degree of comminution distribution by equipoise thresholds is represented in Table 2. The degree of comminution showed very small variation from 49.69% in the consensus group to 48.46% in the disagreement group.

Discussion

This part of the study explored the radiographic characteristics of TL burst fractures associated with various thresholds of agreement in treatment recommendations among expert reviewers. First, we found that the expert panel group had moderate interrater reliability and at least good correlation when it came to determining fracture classification as well as excellent reliability in the probability of PLC injury and the degree of vertebral body comminution. However, the agreement on the use of M1 modifier was slight and fair. This means that the disagreement on best treatment was not caused by differences in PLC integrity evaluation or degree comminution. It is also less likely to be due to fracture classification as the correlation was good, but possible given that the agreement was moderate.

For reasons explained in previous articles in this issue, we have included the cases with suspected or probable Type B injuries. The great majority of the cases were, however, A3 and A4 fractures. These types were also more common in the equipoise group than the agreement group. In the agreement group, there was a higher degree of certainty of PLC injury, more common use of M1 modifier and slightly lower degree of comminution compared to the equipoise group. This means that possible non-A types are more easily identified and lead to more agreement on the treatment choice as illustrated by the agreement having more type B injuries.

The vast majority of fractures in the disagreement subgroup were classified as A3/A4 type (95.3%). In comparison, B type fractures consisted of only 3.1% of the disagreement group. This illustrates the controversy in best treatment surrounding the typical A-type burst fracture. Interestingly, the reliability in classifying fractures as either A3/A4 was comparable in both the agreement and equipoise groups. This supports the idea that the disagreement in treatment recommendation is likely not related to disagreement on how to classify a particular fracture. When the expert surgical community collectively has equipoise around treatment decisions, the influence of surgeon training, local experience, and resource availability likely all factor into the decision to recommend specific treatment much more than variables related to fracture morphology or classification. This will be further assessed in subsequent articles in this focus issue.

The biomechanical importance of the PLC is clear among the expert community. Our results illustrate that a higher

Table 2. Degree of Certainty of PLC Injury, Use of M1 Modifier and Degree of Comminution Distributions by Equipoise Thresholds.

	Degree of Certainty of PLC Injury (Mean, SD)	Use of M1 Modifier (n, %)	Degree of Comminution (Mean, SD)
100% agreement	46.6 (38.3)	91 (51.7)	49.7 (23.3)
91% agreement	45.2 (37.7)	190 (54.0)	51.1 (25.8)
86% agreement	33.3 (33.6)	150 (42.6)	43.1 (21.9)
82% agreement	31.5 (31.7)	210 (39.8)	42.2 (22.5)
77% agreement	30.2 (29.8)	148 (39.6)	46.4 (22.5)
73% agreement	32.3 (29.8)	188 (45.0)	48.2 (20.5)
68% agreement	32.2 (30.8)	177 (42.3)	49.9 (21.6)
64% agreement	26.5 (26.7)	153 (36.6)	49.1 (21.3)
59% agreement	27.0 (25.2)	171 (40.9)	43.7 (19.1)
55% agreement	21.9 (23.2)	134 (30.5)	47.7 (19.4)
Total disagreement	22.5 (24.7)	36 (27.3)	48.5 (20.4)

degree of certainty in PLC injury was associated with higher rate of agreement on treatment. Schroeder et al. previously showed that the overall interobserver reliability in determining the integrity of the PLC was slight (kappa = .11).[17] In their study, surgeon reviewers were asked: «Do you see any significant injury to the posterior ligamentous complex (PLC)? ». The answer could be binary (yes or no). However, evaluation of the PLC integrity relies on the evaluation of many radiographic factors by surgeons, which creates a spectrum of certainty of PLC injury. To account for this, in our study, we asked surgeons: «Based on these CT images how confident are you that the posterior ligamentous complex is injured? (0%-100%, with 0% no PLC injury and 100% absolutely certain the PLC is disrupted) ». When treating PLC injury as a continuous variable on an uncertainty spectrum, the intraclass correlation showed good reliability in the equipoise group and excellent reliability in the agreement group (in which type-B injuries were more common). This can be interpreted as the direction of the surgeon's evaluation of the PLC is very similar whether there is agreement on treatment or not. Our current study furthermore shows that with lower certainty of PLC injury, more disagreement on treatment strategy occurred. It is noteworthy that for most fractures, upright radiographs and magnetic resonance imaging were not available. It is likely that in the absence of convincing radiographic evidence of instability on CT, other individual factors related to practice environment or training influenced the expert surgeon's decision making during the review of images. This individual variation created a wide spectrum of opinions on proposed management.

In the latest AO Spine Thoracolumbar Injury classification,[18] a definite ligamentous injury is defined as type B and if uncertainty about ligamentous injury occurs based on available imaging, a patient-specific modifier (M1) is assigned. In our study, the use of the M1 was more common in the agreement group. This correlates with the higher degree of certainty of PLC injury in the agreement group. However, the values of uncertainty that should be awarded the M1 modifier have not been well defined. Evaluating the PLC as a continuous variable on a spectrum of certainty may help define the threshold of the M1 modifier. This will be explored in a subsequent article of this focus issue.

The degree of vertebral body comminution has been included in previous classifications to optimally define fractures as well as the treatment options. The Load Sharing Classification of Spine Fractures included the degree of vertebral body comminution in the score to evaluate anterior support and assist surgeons in making the choice between short segment pedicle construct and long segment constructs.[19] However, the score did not account for ligamentous integrity making this score incomplete for influencing surgical decision making. Highly comminuted fractures can evolve towards local kyphosis as a result of losing anterior support especially in the thoracic or thoracolumbar spine compared to the lumbar spine.[20,21] The absence of comminution has been identified as a drawback or limitation of the TLICS system although the inclusion of A4 fractures in the updated AO Spine Thoracolumbar Injury classification may provide a surrogate measure for comminution.

For TL burst fractures without neurological deficits, the total TLICS score is usually 4 points with an uncertain PLC injury. Comminution has been recognized as a cause of difficult decision-making in those cases.[20,22] Our results showed a slightly higher degree of comminution in the equipoise group. The variation in mean degree of comminution between the consensus subgroup and disagreement subgroup was also rather small (1.2%). Compared to fracture type and PLC injury, comminution does not seem to differentiate well whether surgeons agree or not on surgical management. Future studies are required to interpret the meaning of comminution in surgeon's decision making before incorporating this variable in future treatment algorithms. This would be to account for the likelihood of failing conservative treatment.

Considering that the experts agreed on most of the radiological parameters but disagree on the management, this prospective multicenter cohort can be considered as 'randomized' in a natural fashion. These patients would get different treatments in different centers although the surgeons agree on what kind of injuries they have. The threshold set for equipoise was based on current literature, which is limited. Appropriate threshold may vary depending on the treatment or pathologies studied. It is possible that the panel of experts does not represent all clinical environments around the world. However, the widespread geography of the AO Spine Knowledge Forum Trauma increases the overall generalizability of these results. It is possible that experts made a treatment recommendation based on their local available resources and their own relative expertise. It is possible that the expert's treatment recommendation may differ if their available resources or capabilities to offer a certain treatment were different.

Conclusion

Our study showed that the best treatment for TL burst fractures (type A3/A4) remains a controversial topic in modern spine surgery. These fracture types were more common in the equipoise group compared to the agreement group. The agreement group showed a higher degree of certainty of PLC injury and more common use of the M1 modifier. Future studies are required to identify the role of comminution in decision making as the degree of comminution was noted to be surprisingly slightly higher in the equipoise group than in the full agreement group.

Declaration of Conflicting Interests

Funding

ORCID iDs

Charlotte Dandurand https://orcid.org/0000-0002-4520-6454
Lorin M. Benneker https://orcid.org/0000-0002-5320-5638
Emiliano Vialle https://orcid.org/0000-0003-1157-4889
Shanmuganathan Rajasekaran https://orcid.org/0000-0001-6043-006X
Mohammad El-Skarkawi https://orcid.org/0000-0001-6177-7145
Rishi M. Kanna https://orcid.org/0000-0001-5817-4909
Mohamed Aly https://orcid.org/0000-0001-5221-8288
Martin Holas https://orcid.org/0000-0002-3763-8767
Jose A. Canseco https://orcid.org/0000-0002-2152-5725
Eugen C. Popescu https://orcid.org/0000-0001-5732-5402
Gaston Camino-Willhuber https://orcid.org/0000-0002-5684-7679
Andrei Fernandes Joaquim https://orcid.org/0000-0003-2645-0483
Sebastian Bigdon https://orcid.org/0000-0002-4649-0610
Ulrich Spiegel https://orcid.org/0000-0002-5179-4192

References

1. Rometsch E, Spruit M, Härtl R, et al. Does operative or non-operative treatment achieve better results in A3 and A4 spinal fractures without neurological deficit? systematic literature review with meta-analysis. *Global Spine J*. 2017;7(4):350-372. doi:10.1177/2192568217699202

2. Pehlivanoglu T, Akgul T, Bayram S, et al. Conservative versus operative treatment of stable thoracolumbar burst fractures in neurologically intact patients: is there any difference regarding the clinical and radiographic outcomes? *Spine*. 2020;45(7):452-458. doi:10.1097/BRS.0000000000003295

3. Tanasansomboon T, Kittipibul T, Limthongkul W, Yingsakmongkol W, Kotheeranurak V, Singhatanadgige W. Thoracolumbar burst fracture without neurological deficit: review of controversies and current evidence of treatment. *World Neurosurg*. 2022;162:29-35. doi:10.1016/j.wneu.2022.03.061

4. Bigdon SF, Saldarriaga Y, Oswald KAC, et al. Epidemiologic analysis of 8000 acute vertebral fractures: evolution of treatment and complications at 10-year follow-up. *J Orthop Surg Res*. 2022;17(1):270. doi:10.1186/s13018-022-03147-9

5. Gnanenthiran SR, Adie S, Harris IA. Nonoperative versus operative treatment for thoracolumbar burst fractures without neurologic deficit: a meta-analysis. *Clin Orthop Relat Res*. 2012; 470(2):567-577. doi:10.1007/s11999-011-2157-7

6. Abudou M, Chen X, Kong X, Wu T. Surgical versus non-surgical treatment for thoracolumbar burst fractures without neurological deficit. *Cochrane Database Syst Rev*. 2013;6: CD005079. doi:10.1002/14651858.CD005079.pub3

7. Ghobrial GM, Jallo J. Thoracolumbar spine trauma: review of the evidence. *J Neurosurg Sci*. 2013;57(2):115-122.

8. Nataraj A, Jack AS, Ihsanullah I, Nomani S, Kortbeek F, Fox R. Outcomes in thoracolumbar burst fractures with a thoracolumbar injury classification score (TLICS) of 4 treated with surgery versus initial conservative management. *Clin Spine Surg*. 2018; 31(6):E317-E321. doi:10.1097/BSD.0000000000000656

9. Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med*. 1987;317(3):141-145. doi:10.1056/NEJM198707163170304

10. Mhaskar R, B Bercu B, Djulbegovic B. At what level of collective equipoise does a randomized clinical trial become ethical for the members of institutional review board/ethical committees? *Acta Inf Med*. 2013;21(3):156-159. doi:10.5455/aim.2013.21.156-159

11. Johnson N, Lilford RJ, Brazier W. At what level of collective equipoise does a clinical trial become ethical? *J Med Ethics*. 1991;17(1):30-34. doi:10.1136/jme.17.1.30

12. AOCenter Innovation Translation. Thoracolumbar Burst Fractures (AOSpine A3, A4) in Neurologically Intact Patients: An Observational, Multicenter Cohort Study Comparing Surgical versus Non-surgical Treatment. clinicaltrials.gov; 2021. https://clinicaltrials.gov/ct2/show/NCT02827214. Accessed January 20, 2022.

13. Ghogawala Z, Terrin N, Dunbar MR, et al. Effect of ventral vs dorsal spinal surgery on patient-reported physical functioning in patients with cervical spondylotic myelopathy: a randomized clinical trial. *JAMA*. 2021;325(10):942-951. doi:10.1001/jama.2021.1233

14. Ghogawala Z, Barker FG, Carter BS. Clinical equipoise and the surgical randomized controlled trial. *Neurosurgery*. 2008;62(6): N9-N10. doi:10.1227/01.neu.0000333288.21621

15. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med*. 2012;22(3):276-282.

16. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

17. Schroeder GD, Kepler CK, Koerner JD, et al. A worldwide analysis of the reliability and perceived importance of an injury to the posterior ligamentous complex in AO type A fractures. *Global Spine J*. 2015;5(5):378-382. doi:10.1055/s-0035-1549034

18. Vaccaro AR, Oner C, Kepler CK, et al. AOSpine thoracolumbar spine injury classification system: fracture description, neurological status, and key modifiers. *Spine*. 2013;38(23):2028-2037. doi:10.1097/BRS.0b013e3182a8a381

19. McCormack T, Karaikovic E, Gaines RW. The load sharing classification of spine fractures. *Spine*. 1994;19(15):1741-1744. doi:10.1097/00007632-199408000-00014

20. Costachescu B, Popescu CE, Iliescu BF. Analysis of the classification systems for thoracolumbar fractures in adults and their evolution and impact on clinical management. *J Clin Med*. 2022; 11(9):2498. doi:10.3390/jcm11092498

21. Schnake KJ, von Scotti F, Haas NP, Kandziora F. [Type B injuries of the thoracolumbar spine: misinterpretations of the integrity of the posterior ligament complex using radiologic diagnostics]. *Unfallchirurg*. 2008;111(12):977-984. doi:10.1007/s00113-008-1503-z

22. Park CJ, Kim SK, Lee TM, Park ET. Clinical relevance and validity of TLICS system for thoracolumbar spine injury. *Sci Rep*. 2020;10(1):19494. doi:10.1038/s41598-020-76473-9