Is the AO spine thoracolumbar injury classification system reliable and practical? a systematic review

Zion Hwang, James Houston, Evangelos M. Fragakis, Cristina Lupu, Jason Bernard, Tim Bishop, Darren F. Lui

From the Department of Orthopaedic and Trauma Surgery, St. George's University Hospital, London, United Kingdom

Controversy surrounding the classification of thoracolumbar injuries has given rise to various classification systems over the years, including the most recent AOSpine Thoracolumbar Injury Classification System (ATLICS). This systematic review aims to provide an up-to-date evaluation of the literature, including assessment of a further three studies not analysed in previous reviews. In doing so, this is the first systematic review to include the reliability among non-spine subspecialty professionals and to document the wide variety between reliability across studies, particularly with regard to sub-type classification. Relevant studies were found via a systematic search of PubMed, EBESCO, Cochrane and Web of Science. Data extraction and quality assessment were conducted in line with Cochrane Collaboration guidelines. Twelve articles assessing the reliability of ATLICS were included in this review. The overall inter-observer reliability varied from fair to substantial, but the three additional studies in this review, compared to previous reviews, presented on average only fair reliability. The greatest variation of results was seen in A1 and B3 subtypes. Least reliably classified on average was A4 subtype. This systematic review concludes that ATLICS is reliable for the majority of injuries, but the variability within subtypes suggests the need for further research in assessing the needs of users in order to increase familiarity with ATLICS or perhaps the necessity to include more subtype-specific criteria into the system. Further research is also recommended on the reliability of modifiers, neurological classification and the application of ATLICS in a paediatric context.

INTRODUCTION

Controversy surrounds the best way to classify thoracolumbar spine injuries. Such injuries commonly present to major trauma units and can cause significant morbidity [1]. Many different management strategies exist, each presenting different risks and benefits for the patients. Thoracolumbar fractures are also highly varied in their presentation, which makes developing a standardised classification system challenging.

■ Zion Hwang,
■ James Houston,
■ Evangelos M. Fragakis,
■ Cristina Lupu,
■ Jason Bernard,
■ Tim Bishop,
■ Darren F. Lui
    Department of Orthopaedic and Trauma Surgery, St. George's University Hospital, London, United Kingdom.
Correspondence : Zion Hwang , St. George's University of London, Cranmer Terrace, Tooting, London, SW17 0RE, United Kingdom.
Email : m1608653@sgul.ac.uk
© 2021, Acta Orthopædica Belgica.

Classification systems of injuries are an important clinical tool, because they can provide efficient communication between clinicians (2). In that respect, it is fundamental that classification systems are comprehensive, accurate and reliable, in order to lead to the most indicated treatment options and desired outcomes (3). Thus, the balance of being precise enough to be practically useful, but simple enough to have good inter-observer reliability is key (4).

Numerous thoracolumbar fracture classification systems have been proposed since Watson-Jones's pioneering classification in 1938 (5). Many of these systems focused on the morphological characteristics of injuries. Nicoll (6) was the first to allow a more practical interpretation by focusing mainly on those injuries with worse functional outcomes. Holdsworth (7) built upon this by correlating fracture patterns with associated neurological deficits and their prognosis. Denis (8) restructured the classification system with a more hierarchical grading, taking into consideration the biomechanical aspects of the fractures, potential instability and the possibility of a neurological involvement. He also introduced the three-column concept at a time where the use of computed tomography (CT) made possible a precise three-dimensional understanding of the morphological aspects of fractures (8). The integration of the morphology, neurological deficit and hierarchical grading of injuries enabled surgeons to identify more accurately the injuries to treat surgically, providing optimised results for the patients.

To address the criticisms that historic classifications were imprecise (2), Magerl proposed the Arbeitsgemeinschaft für Osteosynthesefragen (AO) model (9). Although this model served as a good descriptor of injuries, it was found to be overly complex and unreliable (10). To mitigate the issues with the AO model, the Spine Trauma Study Group (STSG) (11) devised the simpler Thoracolumbar Injury Classification and Severity score (TLICS) based around three important aspects :

1. Injury morphology
2. Posterior Ligamentous Complex (PLC) integrity
3. Patient's neurological status

TLICS combines clinical examination, X-ray, CT and/or magnetic resonance imaging (MRI) to visualise injury morphology and to evaluate PLC integrity.

The TLICS establishes an injury severity pathway to guide the decision-making process but its practicality has been questioned since it fails to consider the vitally important patient-specific modifiers in tailoring the management plan to individual patients (11,12).

AO Spine Thoracolumbar Injury Classification System (ATLICS) is a newer classification system built upon the original AO system by Magerl (3,9). It strikes a balance between AO's descriptive detail

Table I. — Comparison of the pros and cons of AO, TLICS, and ATLICS classifications

	AO	TLICS	ATLICS
Pros	- Detailed descriptions. - Highly accurate.	- Simple and easy to use. - Accounts for neurological injury and PLC integrity. - Severity ranked and treatment suggestions indicated using TLICS. - Validated for paediatric use.	- Balance : details vs simplicity. - 9 subcategories. - Accounts for neurological injury, PLC integrity and patient-specific modifiers. - Severity ranked and more specific treatment options indicated, using TL AOSIS.
Cons	- Too complex, 53 subcategories. - Low intra-observer reliability. - Mechanism of injury based only on morphological classification. - PLC integrity not fully assessed. - No indication of severity or treatment suggestions.	- Only 4 categories. - Does not account for patient-specific modifiers.	- Lack of awareness/training. - Not validated for paediatric use.

Table II. —AOSpine Thoracolumbar Injury Classification System

Classification	Type	Description
A0	Minor/no injury *or* process fracture	Clinically insignificant fracture of the spinous/transverse process
A1	Wedge compression *or* impaction	Single end-plate fracture, without posterior vertebral wall involvement
A2	Split *or* pincer-type	Fracture line involves both endplates, but not the posterior vertebral wall
A3	Incomplete burst	Fracture involves a single endplate with any involvement of posterior vertebral wall
A4	Complete burst	Fracture involves posterior vertebral wall and both endplates
B1	Posterior transosseous disruption *or* 'chance' fracture	Osseous failure of posterior tension band, extending into the vertebral body. Only affects a segment of motion
B2	Posterior tension band or posterior ligamentous disruption	Disruption of the posterior tension band with/without osseous involvement. Can affect multiple vertebrae
B3	Hyperextension injury *or* anterior ligamentous disruption	Disruption of the anterior longitudinal ligament (anterior tension band), extending through the vertebral body/ intervertebral disc
C	Displacement *or* Translation injury	Displacement beyond physiological range of the cranial and caudal parts of the spinal column in any plane: hyperextension, translation, separation

Source : Vaccaro, A.R., et al. (*10*)

Table III. — Neurological status and Case-Specific Modifiers according to ATLICS

Neurological status	Description	Detail
N0	Intact	
N1	Transient	Deficit no longer present
N2	Radiculopathy	
N3	Incomplete *or* cauda equina	
N4	Complete	
Case-specific modifiers		
M1	Indeterminate injury to tension band	Seems stable from bony standpoint but operative stabilisation may be considered depending on PLC integrity
M2	Patients with a co-morbidity affecting the surgical decision	E.g. ankylosing spondylitis, rheumatologic conditions, osteo-porosis

Source : Vaccaro, A.R., et al. (*10*)

and TLICS' simpler objective focus on management plans [Table I]. ATLICS achieves this by dividing fractures into 3 types : A – compression injuries (subtypes : A0-4) ; B – tension band injuries (subtypes : B1-3) ; C – displacement injuries (no subtypes) [Table II]. ATLICS also grades fractures by their neurological status (subtypes : N0-4) and formally accounts for patient-specific modifying factors (subtypes : M1-2) [Table III]. This allows treatment algorithms to be more patient-tailored.

Since most thoracolumbar fractures present as an emergency, it is important to use a precise but reliable classification system that is designed around the resources available in emergency departments. ATLICS uses CT as its choice investigative modality, which has good accessibility and sensitivity even for less experienced surgeons (*13,14*). It has the benefit over MRI for not over-diagnosing PLC injury (*15*). Additionally, ATLICS has better standardised use of additional variables, such as dislocation and increased interspinous distance, as highlighted by Barcelos et al. (*13*), which helps particularly to differentiate between Types A and B injuries.

Spine surgeons are yet to universally agree on the use of a single classification system. Many studies have supported ATLICS as the most complete and

reliable, compared to previous classification systems [16-19]. Two previous systematic reviews have been performed on the topic but failed to include reliability data from all available and relevant studies in their assessment [20,21]. Previous reviews were also limited exclusively to reliability among spine surgeons. This still leaves open the question about whether their findings are generalisable to clinicians from other disciplines [20].

Table IV. — Summary of sample articles

First author (year)	Number of cases	Patient characteristics	Number of observers	Observer characteristics	Consensus
Azimi (2015)[24]	56	Type A: 41.9% Type B: 28.4% Type C: 29.7%	2	Spine surgeons	ATLICS is reliable
Barcelos (2016)[13]	43	Type A: 32.5% Type B: 16.3% Type C: 51.2%	3	Spine surgeons	When CT was used as the only diagnostic tool, PLC injury was identified in most cases and demonstrated satisfactory reliability.

Cheng (2017)[25]	109	Type A: $\bar{x}$=75.2 (75-75.4)
		Type B: $\bar{x}$=21.9 (21.6-22.2)
		Type C: $\bar{x}$=2.9 (2.8-3)
			6	Orthopaedic surgeons
				Subgroups (years of experience in spine trauma):
				   Subgroup A (2 years)
				   Subgroup B (1 year) in China
					ATLICS is applicable to everyday clinical practice

Kaul (2017)[16]	50	Type A: 39.45% Type B: 24% Type C: 36.55%	11	Spine surgeons from 4 counties (n=10 orthopaedic surgeons)	ATLICS has better reliability for identifying fracture morphology than TLICS
Kepler (2016)[17]	25	Type A: 53%. Type B: 32.4%. Type C: 13.4%	100	Surgeons (Africa, Europe, North America, South America)	ATLICS has moderate inter-observer reliability, substantial intra-observer reliability and can be applied to trauma patients more reliably than previous classification systems
Pishnamaz (2018)[18]	91	-	7	Spine surgeons	ATLICS reliability is superior to TLICS
Lopes (2019)[19]	25	-	24	In Brazil Specialist (n=6) Resident (n=18)	ATLICS reliability is superior to AO Magerl
Morgonskold (2019)[22]	50	Age range 9-96yrs	5	Orthopaedics/Spine consultant (n=2) Orthopaedics resident (n=1) Emergency medicine resident (n=1) Medical student (n=1) in Sweden	Adaptation of ATLICS shows acceptable reliability
Sadiqi (2015)[26]	25	Type A: A0: 4% Type B: 12% Type C: 12%	100	International group of spine surgeons new to the classification. Subgroups (years of experience): Subgroup 1 (≤10 years): n = 30 Subgroup 2 (11-20 years): n = 44 Subgroup 3 (>20 years): n = 26	The surgeons' level of experience does not significantly influence the intra-observer reliability of ATLICS
Urrutia (2015)[27]	70	Type A: 49.8% Type B: 30% Type C: 20.2%	6	Spine surgeons: n=3 trauma and orthopaedic residents n=3 spine fellowship	ATLICS has adequate intra-observer agreement
Vaccaro (2013)[10]	40	Type A: 54% Type B: 24% Type C: 22%	9	Spine surgeons (n=9 fellowship-trained)	ATLICS is clinically relevant and has reasonable reliability and accuracy.
Yacoub (2017)[28]	54		2	Spine surgeon (n=1) Neurosurgery resident (n=1)	ATLICS has good reliability as groups. Subgroups demonstrated worse and varying reliability. Group A can be treated conservatively or surgically. Groups B and C are treated surgically.
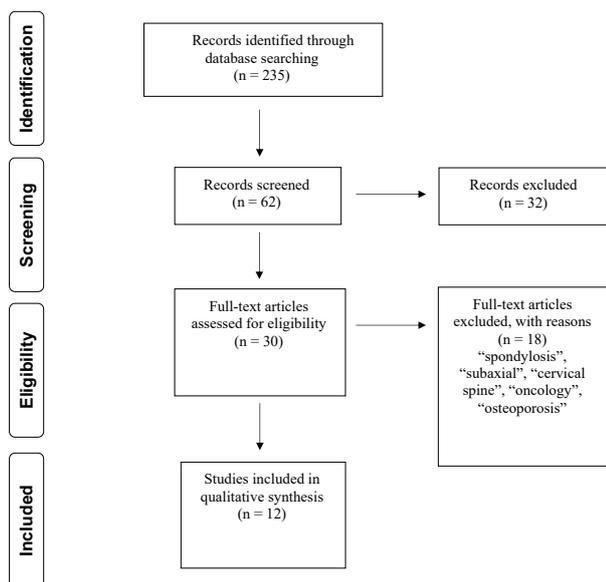
Figure. 1. Literature search based on
The PRISMA Statement *(23)*

Therefore, this systematic review aims to provide an up-to-date evaluation of the literature by assessing three further studies *(18,19,22)* not analysed in previous reviews. In doing so, this is the first systematic review to include the reliability among non-spine subspecialty professionals ; something that is important in validating the classification as an inter-disciplinary communication tool. Also detailed in this review is the wide variety between reliability across studies, particularly with regard to sub-type classification, documenting findings that have not been previously published.

The final goal of this up-to-date and accurate analysis of the ATLICS is to comment on its characteristics and whether it should be adopted by the spinal community, taking into account aspects that could be further improved.

MATERIALS AND METHODS

This systematic review was conducted using PubMed, EBESCO, Cochrane, and Web of Science databases between October 2018 and December 2019 following the PRISMA guidelines *(23)* [Figure 1]. The literature search strategy involved combining keywords and their expansions with the Boolean and Proximity search strategies : "AO", "AOSpine"

"ATLICS", "thoracolumbar", "thoracic", "lumbar", "spine", "spinal", "vertebra*", "fracture", "break*", "damage", "injury", "trauma", "classification", "score", "scoring", "system", "reliability".

Inclusion criteria

- English language
- Peer reviewed
- Publications dating 2013 to 2019
- Thoracolumbar spine injuries only

Exclusion criteria

- Non-english language
- Grey literature
- Self-reported outcomes
- Letters/Editorials
- Publication dating prior to 2013 (ATLCS was published in 2013)
- Cervical spine injuries
- Sacral spine injuries
- Non-traumatic fractures (e.g. pathological, osteoporotic)

The literature search yielded 235 unique articles. Initial abstract review identified 62 articles, with 30 eligible for full-text review. From these, 12 articles were selected for inclusion in the systematic review [Table IV] using the Critical Appraisal Skills Programme (CASP) quality assessment tool *(29)* [Figure 1]. There was only one study on neurological classification *(16)*. The search yielded no literature on the reliability of patient-specific modifiers (M1/M2).

RESULTS

The reliability of a classification assesses how consistently similar injuries are classified into a given severity group by the same surgeon (intra-observer) or different surgeons (inter-observer).

This study reviewed the data from twelve publications [Table V], where overall intra- and/ or inter-observer reliability was yielded by eight studies *(10,16-19,25-27)*, overall reliability without subtypes by eight studies *(10,13,16-19,26,27)*, reliability according to type by ten studies *(10,13,16-18,22,24,25,27,28)*, and reliability of subtypes by five studies *(10,17,25,27,28)*.

Table V. — Summary of reliability data

	Azimi (2015)[24]	Barcelos (2016)[13]	Cheng (2017)[25]	Kaul (2016)[16]	Kepler (2016)[17]	Lopes (2018)[18]	Morgonskold (2019)[22]	Pishnamaz (2018)[19]	Sadiqi (2015)[26]	Urrutia (2015)[27]	Vaccaro (2013)[10]	Yacoub (2017)[28]
Overall	-	-	0.362/-	0.45/0.61	0.56/x̄=0.68	0.59/x̄=0.56	-	0.41/0.57	Group 1 -/x̄=0.69 Group 2 -/x̄=0.60 Group 3 -/x̄=0.67	0.55/0.71	0.64/x̄=0.77	-
Overall without subtypes	-	x̄=0.59/-	-	0.59/0.68	0.74/x̄=0.81	0.85/x̄=0.89	-	0.61/0.71	Group 1 -/x̄=0.83 Group 2 -/x̄=0.81 Group 3 -/x̄=0.79	0.62/0.77	0.72/x̄=0.85	-
Types												
A	0.88/0.84	x̄=0.65/-	0.385/0.442	0.64/-	0.80/0.57	0.88/x̄=0.91	x̄=0.72/x̄=0.63	-	-	0.61/-	0.72/0.72	0.75/0.75
B	0.86/ 0.83	x̄=0.22/-	0.292/0.485	0.40/-	0.68/0.43	0.76/x̄=0.89	x̄=0.6/x̄=0.56	-	-	0.57/-	0.58/0.43	0.70/0.70
C	0.89/ 0.86	x̄=0.67/-	0.552/0.412	0.71/-	0.72/-	0.8/x̄=0.84	x̄=0.7/x̄=0.68	-	-	0.69/-	0.7/-	0.85/0.85
Subtypes												
A0	-	-	1.000/0.620	-	0.96/-	-	-	-	-	0.94/-	1.00/-	0.60/-
A1	-	-	0.338/0.548	-	0.60/-	-	-	-	-	0.63/-	0.59/-	0.00/-
A2	-	-	0.412/0.561	-	0.68/-	-	-	-	-	0.45/-	0.50/-	0.33/-
A3	-	-	0.278/0.332	-	0.42/-	-	-	-	-	0.55/-	0.45/-	0.59/-
A4	-	-	0.312/0.482	-	0.19/-	-	-	-	-	0.56/-	0.58/-	0.16/-
B1	-	-	0.312/0.413	-	0.53/-	-	-	-	-	0.18/-	0.65/-	0.48/-
B2	-	-	0.284/0.152	-	0.41/-	-	-	-	-	0.60/-	0.34/-	0.77/-
B3	-	-	0.120/0.323	-	0.66/-	-	-	-	-	0.80/-	0.41/-	0.85/-

Each study is represented by the First Author (year). For each study reliability data is presented as kappa values in the format : interobserver reliability/intraobserver reliability.

Table VI. — Landis and Koch [30] interpretation of kappa values

Kappa value range, k	Agreement
0.0 - 0.2	Slight
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Substantial
0.81 - 1.00	Almost perfect

All studies reported reliability using Kappa coefficient (k) and interpreted them according to Landis and Koch [30] [Table VI].

The three additional studies included in this systematic review are, Lopes et al. [18], Pishnamaz et al. [19], and Morgonskold et al. [22]. Importantly, the latter [22] uses emergency medicine clinicians and medical students as observers and a mixed adult and paediatric (age range : 9-96) subject cohort in their dataset.

Overall inter-observer reliability was reported in seven studies [10,20,16-19,27], of which six [10,16-19,27] showed moderate to substantial reliability and one [20] fair reliability. Seven studies [10,13,16-19,27] found inter-observer reliability when subtypes were excluded and out of those, five [10,17-19,27] found that this improved reliability in all studies to substantial or better.

Ten studies [10,13,16-18,22,24,25,27,28] reported inter-observer reliability for Types A, B and C. Type A 'compression' injuries had substantial or better reliability in nine of them [10,13,16-18,22,24,27,28]. Type B 'tension band injuries' had varied reliability throughout all studies with an average k-value of 0.56 (range : 0.22-0.86). Type C "displacement injuries" had substantial or better reliability in nine [10,13,16-18,22,24,27,28] out of ten studies with the other reported study[25] still moderate (k=0.552).

Five studies [10,17,25,27,28] reported inter-observer reliability by subtypes. Results showed almost perfect reliability of A0 injuries except Yacoub et al. [28] (k=0.60) but varied considerably from slight to substantial between studies for A1 and A4 compression injuries. Results for A2 and A3 were comparatively more consistent, averaging moderate reliability. Significant variation in reliability was seen in B subtypes [Table VII] – particularly B3 ranging from slight to almost perfect across studies. Type C is not divided into subgroups in ATLICS ; therefore, no data was collected.

Overall intra-observer reliability was reported in seven studies *(10,16-19,26,27)*, of which six studies *(10,16-18,26,27)* presented substantial reliability. Excluding subtypes, reliability improved in all studies, with four studies *(10,17,18,26)* achieving almost perfect agreement. Seven studies *(10,17,18,22,24,25,28)* reported intra-observer reliability for Types A and B, showing moderate to almost perfect agreement. Five studies *(18,22,24,25,28)* reported intra-observer reliability for Type C, of which three were almost perfect. [Table VII]. Only Cheng et al. *(25)* studied intra-observer reliability by subtypes. A0 was the only subtype to yield substantial reliability. Moderate reliability was found for A1, A2, A4 compression and B1 tension band injuries ; A3 and B3 had fair reliability ; and for B2 the reliability was only slight.

Only Kaul et al. *(16)* assessed for reliability of the neurological classification, reporting almost perfect agreement.

DISCUSSION

This study explores the reliability of ATLICS, demonstrating overall substantial reliability, particularly when subtypes were disregarded. Through considering morphology of fracture, PLC integrity, neurological injury and case modifiers [Table I, II], ATLICS is undoubtedly the product of many evolutionary steps since the early classification systems. Nevertheless, its assessment in the literature varies, with some experts *(16)* claiming ATLICS as the most reliable classification that currently exists, while others *(21)* suggest that TLICS demonstrates comparatively better reliability based on kappa values alone.

This study finds that ATLICS has much variation between studies concerning the inter-observer reliability when classifying Type B thoracolumbar injuries and subtypes A1 and B3. It is therefore difficult to conclude the level of reliability of ATLICS for those groups. Furthermore, ATLICS is found to be least reliable on average in classifying A4 injuries (k≈0.3604). Nevertheless, this study confirms substantial overall inter-observer reliability in classifying Types A and C injuries. In addition, as A2 and A3 subtypes mostly score

Table VII. — Points allocated to subgroups

Subgroup	Points
Morphological subgroup	
A0	0
A1	1
A2	2
A3	3
A4	5
B1	5
B2	6
B3	7
C	8
Neurological subgroup	
N0	0
N1	1
N2	2
N3	4
N4	4
Nx	3
Case-specific modifier subgroup	
M1	1
M2	0

Source : Kepler, C.K., et al. *(31)*

Table VIII. — Surgical algorithm for TL AOSIS

TL AOSIS Score	Recommended Intervention
≤ 3	Conservative
>5	Surgery

Source : Vaccaro, A.R., et al. *(32)*

Table IX. — Thoracolumbar Injury Classification and Severity score (TLICS)

Parameter	Points
Mechanism	
Compression	1
Burst	1
Translational/rotational	3
Distraction	4
Neurological status	
Intact	0
Nerve root injury	2
Spinal cord/conus medullaris injury	
Complete	2
Incomplete	3
Cauda equina	3
Posterior ligamentous complex	
Intact	0
Indeterminate	2
Disrupted	3

Source : Vaccaro, A.R., et al. *(11)*

moderate reliability, and B1 and B2 subtypes mainly lie within the fair-moderate range, evidence suggests that ATLICS is reliable for the majority of injuries [Table V].

The importance of ATLICS being reliable is to enable decision making in clinical practice. Since 2013, a treatment algorithm based on Kepler et al.'s Thoracolumbar AOSpine Injury Score (TL AOSIS) *(31)* has been incorporated into the system. This gives a certain number of points according to the individual grades of the morphological, neurological and modifier subgroups. The total injury score is then used to guide whether surgery should be considered [Table VII, VIII] and could potentially predict the best timing of operation as outlined in Du et al. study *(1)*. Since its predecessor [Table IX], there have been an expansion of nerve root injury, including transient and radicular symptoms, and upgrading of complete cord injury to the same level as incomplete. This evolution is important because of the growing evidence to act swiftly for all grade A injuries according to the American Spinal Injury Association Impairment Scale (AIS) *(33)* as there is a chance to improve AIS grade by two levels *(34)* when operations are done within 24 hours. Therefore, a complete cord injury is no longer considered as hopeless as used to be, compared to an incomplete cord injury. Patients with an unstable spine fracture who cannot be examined should be assumed to have a neurological injury and work towards fixing them rather than waiting to see if they will become paralysed. In that respect, there is also the addition of the "+ modifier" which delineates the existence of ongoing cord compression. Most recent literature *(35)* concludes that treatment recommendations of the TL AOSIS are more reliable than those of the TLICS. However, TL AOSIS can only be clinically useful in deciding the most appropriate management of thoracolumbar injuries if ATLICS itself, which is the basis of the treatment algorithm, is accurate and reliable.

This study was not able to fully assess the reliability of the modifiers and neurological classification due to lack of published evidence. Moreover, the reliability of different management methods for each fracture type was not explored as it was outside the scope of this study. There-fore, future research in these areas might be needed.

One of the studies analysed included paediatric thoracolumbar injuries *(22)* (<19yrs old), presenting an inter- and intra-observer reliability in line with the rest of the adult studies assessed. However, the number of its paediatric cohort was unknown. Considering the biological aspects of bone metabolism and healing timing in the paediatric population, further studies are recommended to better assess the reliability of ATLICS in paediatric thoracolumbar injuries.

CONCLUSION

This systematic review provides an up-to-date assessment of the current literature on ATLICS, including three further studies not evaluated previously. Furthermore, the inter-specialty reliability among non-spine specialists was analysed, representing a novel approach and an additional validation step compared to previous studies. An efficient inter-specialty communication, in the context of the daily challenges of an acute and emergency department, should represent the ultimate pragmatic target of every classification system.

This systematic review concludes that the ATLICS is reliable for the majority of the thoracolumbar injuries and is clinically more useful compared to previous classification systems and severity scores. However, more input is needed for the application of the ATLICS in a paediatric context. Furthermore, this review has highlighted an important inter-observer reliability variation between studies, particularly of Type B and subtypes A1 and B3, documenting findings not previously published. The important subtype-variability noticed suggests that further research is needed 1) to determine the reliability of subtypes that showed a large inter-observer disparity between studies 2) to assess the needs of users in order to increase familiarity with the ATLICS 3) the possible necessity to include more subtype-specific criteria into the system.

REFERENCES

1. **Du, J.P., Fan, Y., Liu, J.J.,** *et al.* Decompression for Traumatic Thoracic/Thoracolumbar Incomplete Spinal

Cord Injury : Application of AO Spine Injury Classification System to Identify the Timing of Operation. *World Neurosurg.* 2018 ; 116 : e867-e873.

2. **Mirza, S.K., Mirza, A.J., Chapman, J.R., and Anderson, P.A.** Classifications of Thoracic and Lumbar Fractures : Rationale and Supporting Data. *J. Am. Acad. Orthop. Surg.* 2002 ; 10(5) : 364-377

3. **Srinivasan, P.** A Review of Thoracolumbar Spine Fracture Classification Systems. *Indian Spine J.* 2018 ; 1(2) : 71-78.

4. **Oner, F.C., Ramos, L.M.P., Simmermacher, R.K.J., et al.** Classification of thoracic and lumbar spine fractures : problems of reproducibility A study of 53 patients using CT and MRI. *Eur. Spine J.* 2002 ; 11(3) : 235-245.

5. **Watson-Jones, R.** The result of postural reduction of fractures of the spine. *J. Bone Joint Surg.* 1938 ; 20(3) : 567-586.

6. **Nicoll, E.A.** Fractures of the dorso-lumbar spine. *Bone Joint J.* 1949 ; 31-B(3) : 376-394.

7. **Holdsworth, F.** Fractures, dislocations, and fracture-dislocations of the spine. *J. Bone Joint Surg.* 1963 ; 45-B(1) : 6-20.

8. **Denis, F.** The three-column spine and its significance in the classification of acute thoracolumbar spinal injuries. *Spine.* 1983 ; 8(8) : 817-831.

9. **Magerl, F., Aebi, M., Gertzbein, S.D., Harms, J., Nazarian, S**. A comprehensive classification of thoracic and lumbar injuries. *Eur. Spine J.* 1994 ; 3(4) : 184-201.

10. **Vaccaro, A.R., Oner, F.C., Kepler, C.K., et al.** AOSpine thoracolumbar spine injury classification system : fracture description, neurological status, and key modifiers. *Spine.* 2013 ; 38(23) : 2028-2037.

11. **Vaccaro, A.R., Lehman Jr, R.A., Hurlbert, R.J., et al.** A new classification of thoracolumbar injuries : the importance of injury morphology, the integrity of the posterior ligamentous complex, and neurologic status. *Spine.* 2005 ; 30 : 2325-2333.

12. **Lee, J.Y., Vaccaro, A.R., Lim, M.R., et al.** Thoracolumbar injury classification and severity score : a new paradigm for the treatment of thoracolumbar spine trauma. *J. Orthop. Sci.* 2005 ; 10(6) : 671-675

13. **Barcelos, A.C.E.S., Joaquim, A.F., Botelho, R.V.** Reliability of the evaluation of posterior ligamentous complex injury in thoracolumbar spine trauma with the use of computed tomography scan. *Eur. Spine J.* 2016 ; 25(4) : 1135-1143.

14. **Rajasekaran, S., Kanna, R.M., Schroeder, G.D., et al.** Does the Spine Surgeon's Experience Affect Fracture Classification, Assessment of Stability, and Treatment Plan in Thoracolumbar Injuries? *Global Spine J.* 2017 ; 7(4) : 309-316.

15. **Astolfi, M.M., Millhouse, P., Murphy, H., Schroeder, G.D., Vaccaro, A.R.** Thoracolumbar Fracture Classifications : Pros and Cons of the TLICS and AOSpine Classification Systems. Contemp. *Spine Surg.* 2018 ; 19(1) : 1-7.

16. **Kaul, R., Chhabra, H.S., Vaccaro, A.R., et al.** Reliability assessment of AOSpine thoracolumbar spine injury classification system and Thoracolumbar Injury Classification and Severity Score (TLICS) for thoracolumbar spine injuries : results of a multicentre study. *Eur. Spine J.* 2017 ; 26(5) : 1470-1476.

17. **Kepler, C.K., Vaccaro, A.R., Koerner, J.D., et al.** Reliability analysis of the AOSpine thoracolumbar spine injury classification system by a worldwide group of naïve spinal surgeons. *Eur. Spine J.* 2016 ; 25(4) : 1082-1086.

18. **Lopes, F., Ferreira, A., Santos, R., Macaneiro, C.** Intraobserver and interobserver reproducibility of the old and new classifications of the toracolombar fractures. *Rev. Brazil Orthop.* 2018 ; 53(5) : 521-526.

19. **Pishnamaz, M., Balosu, S., Curfs, I., et al.** Reliability and Agreement of Different Spine Fracture Classification System : An Independent Intraobserver and Interobserver Study. *World Neurosurg.* 2018 ; 115 : e695-e702

20. **Abedi, A., Mokkink, L.B., Zadegan, S.A., et al.** Reliability and Validity of the AOSpine Thoracolumbar Injury Classification System : A Systematic Review. *Global Spine J.* 2019 ; 9(2) : 231-242.

21. **Curfs, I.** Decision making in the treatment of thoracolumbar fractures. Available from : https ://www.orthopeden.org/downloads/669/proefschrift-inez-curfs.pdf#page=23 [Accessed on : 25th December 2019].

22. **Morgonskold, D., Warkander, V., Savvides, P.,** et al. Inter- and intra-rater reliability of vertebral fracture classifications in the Swedish fracture register. *World J. Orthop.* 2019 ; 10(1) : 14-22.**Moher, D., Shamseer, L., Clarke, M., et al.** Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Cochrane Database Syst. Rev.* 2015 ; 4(1) : 1.

23. **Azimi, P., Mohammadi, H.R., Azhari, S., Alizadeh, P., Montazeri, A.** The AOSpine thoracolumbar spine injury classification system : A reliability and agreement study. *Asian J. Neurosurg.* 2015 ; 10(4) : 282-285.

24. **Cheng, J., Liu, P., Sun, D., Qin, T., Ma, Z., Liu, J.** Reliability and reproducibility analysis of the AOSpine thoracolumbar spine injury classification system by Chinese spinal surgeons. *Eur. Spine J.* 2017 ; 26(5) : 1477-1482.

25. **Sadiqi, S., Oner, F.C., Dvorak, M.F., Aarabi, B., Schroeder, G.D., Vaccaro, A.R.** The Influence of Spine Surgeons' Experience on the Classification and Intraobserver Reliability of the Novel AOSpine Thoracolumbar Spine Injury Classification System-An International Study. *Spine* 2015 ; 40(23) : e1250-e1256.

26. **Urrutia, J., Zamora, T., Yurac, R., et al.** An independent interobserver reliability and intraobserver reproducibility evaluation of the new AOSpine Thoracolumbar Spine Injury Classification System. *Spine.* 2015 ; 40(1) : e54-e58.

27. **Yacoub, A.R., Joaquim, A.F., Ghizoni, E., Tedeschi, H., Patel, A.A.** Evaluation of the safety and reliability of the newly-proposed AO spine injury classification system. *J. Spinal Cord. Med.* 2017 ; 40(1) : 70-75.

28. **CASP** - Critical Appraisal Skills Programme. CASP Checklists - CASP - Critical Appraisal Skills Programme.

[online] Available at : https://casp-uk.net/casp-tools-checklists/ [Accessed 9 Dec. 2018]

29. **Landis, J.R., Koch, G.G.** The measurement of observer agreement for categorical data. *Biometrics.* 1977 ; 33(1) : 159-174.

30. **Kepler, C.K., Vaccaro, A.R., Schroeder, G.D.,** *et al.* The Thoracolumbar AOSpine Injury Score. *Global Spine J.* 2016 ; 6(4) : 329-334.

31. **Vaccaro, A.R., Schroeder, G.D., Kepler, C.K.,** *et al.* The surgical algorithm for the AOSpine thoracolumbar spine injury classification system. *Eur. Spine J.* 2016 ; 25 : 1087-1094.

32. **Roberts, T.T., Leonard, G.R., Cepela, D.J.** Classifications In Brief : American Spinal Injury Association (ASIA) Impairment Scale. *Clin. Ortho. Relat. Res*. 2017 ; 475 : 1499-1504.

33. **Fehlings, M.G., Vaccaro, A.R., Wilson, J.R.,** *et al.* Early versus Delayed Decompression for Traumatic Cervical Spinal Cord Injury : Results of the Surgical Timing in Acute Spinal Cord Injury Study (STASCIS). *PLOS ONE* 7(2) : e32037.

34. **An, Z., Zhu, Y., Wang, G., Wei, H., Dong, L.** Is the Thoracolumbar AOSpine Injury Score Superior to the Thoracolumbar Injury Classification and Severity Score for Guiding the Treatment Strategy of Thoracolumbar Spine Injuries? *World Neurosurg.* 2020 ; 137 : e493-e498.