

TRABAJO FIN DE MÁSTER:

Máster de Iniciación a la Investigación en Medicina

CONCORDANCIA Y REPRODUCIBILIDAD DE LA NUEVA CLASIFICACIÓN AOSPINE PARA FRACTURAS VERTEBRALES TORACOLUMBARES



Julián García Maroto

UNIVERSIDAD DE ZARAGOZA

Facultad de medicina: Departamento de



Universidad Zaragoza

*Concordancia y reproducibilidad de la nueva clasificación
AOSpine para fracturas vertebrales toracolumbares*

-

*Reliability and reproducibility of the new AOSpine
classification for thoracolumbar vertebral fractures*

Director: José María Artigas Martín

Profesor asociado de la Universidad de Zaragoza

Departamento de pediatría, Radiología y Medicina Física

Jefe de sección de Radiología de Urgencias del HUMS

ÍNDICE

RESUMEN.....	1
1. INTRODUCCIÓN	3
2. HIPÓTESIS Y OBJETIVOS	8
3. MATERIAL Y MÉTODOS.....	9
4. RESULTADOS.....	19
5. DISCUSIÓN	23
6. CONCLUSIONES.....	29
7. ANEXOS	30
8. BIBLIOGRAFIA.....	32

RESUMEN

Concordancia y reproducibilidad de la nueva clasificación AOSpine para fracturas vertebrales toracolumbares.

Objetivo: Analizar el grado de concordancia interobservador e intraobservador en la clasificación de fracturas vertebrales toracolumbares traumáticas según la clasificación AOSpine.

Diseño: Estudio de concordancia

Métodos: Se analizaron las imágenes de tomografía computarizada (TC) de 60 pacientes con fracturas traumáticas vertebrales toracolumbares. Dos radiólogos con distinta experiencia en patología espinal (un neurorradiólogo y un radiólogo general) clasificaron las fracturas según el modelo propuesto por AOSpine evaluando cada fractura en dos ocasiones con un intervalo de tiempo mínimo de dos meses entre ellas. Se utilizó el coeficiente Kappa para evaluar los grados de concordancia global y porcentajes de concordancia para el análisis de concordancia por tipos.

Resultados: La concordancia interobservador global de las fracturas clasificadas por sus tipos (A, B, C) fue excelente con valores $k = 0.803$ (0.712-0.9). La concordancia interobservador global por subtipos fue moderada con un valor $k = 0.606$ (0.306-1). La concordancia intraobservador fue excelente para ambos evaluadores (0.859 y 0.802).

Conclusión: La clasificación de las fracturas vertebrales toracolumbares mediante el sistema propuesto por AOSpine tiene unos niveles de concordancia y reproductibilidad adecuados para su empleo por radiólogos con distinto nivel de experiencia. Estudios futuros con un mayor número de casos y observadores podrían llegar a obtener resultados más representativos.

Palabras clave: Fractura vertebral, clasificación de fracturas vertebrales, toracolumbares, estudio de concordancia.

ABSTRACT

Reliability and reproducibility of the new AOSpine classification for thoracolumbar vertebral fractures

Objective: To analyze the degree of interobserver reliability and intraobserver reproducibility in the classification of traumatic thoracolumbar vertebral fractures according to the AOSpine classification.

Design: Agreement study.

Methods: Computed tomography (CT) images of 60 patients with traumatic thoracolumbar vertebral fractures were analyzed. Two radiologists with different experience in spinal pathology (a neuroradiologist and a general radiologist) classified the fractures according to the model proposed by AOSpine evaluating each fracture on two occasions with a minimum interval of two months between them. The Kappa coefficient was used to evaluate the global concordance grades and concordance percentages was used for the concordance analysis by type.

Results: The interobserver reliability was excellent when considering the fracture type (A, B, C), with a $k = 0.803$ (0.712-0.9). The interobserver agreement when considering the subtypes was moderate, $k = 0.606$ (0.306-1). The intraobserver reproductibility was excecellent for both evaluators (k 0.859 and 0.802).

Conclusions: The classification of thoracolumbar vertebral fractures using the system proposed by AOSpine has adequate levels of concordance and reproducibility for use by radiologists with different levels of experience. Future studies with a greater number of cases and observers could get more representative results.

Key words: spine fracture , thoracolumbar , agreement study , spinal injury classification

1. INTRODUCCIÓN

Las fracturas del territorio toracolumbar constituyen más del 75% de todas las fracturas vertebrales¹. Esta predilección puede explicarse por las características morfológicas de cada segmento vertebral. Funcionalmente, la columna toracolumbar puede dividirse en tres segmentos²: columna torácica, unión toracolumbar y columna lumbar, cada uno de estos segmentos está diseñado para soportar distintos tipos de fuerzas. La columna torácica, comprendida entre los cuerpos vertebrales T1 y T10 presenta la mayor rigidez de todo el eje vertebral toracolumbar al ser la encargada de mantener la posición erecta del cuerpo y servir de protección, junto a las costillas, de los órganos alojados en la caja torácica. Esta rigidez se consigue a partir de unos discos intervertebrales más delgados que los de otros segmentos de la columna y mediante unas facetas articulares con orientación coronal que las hacen especialmente resistentes a las fuerzas ejercidas en el eje antero-posterior. El segmento vertebral que se ve más frecuentemente dañado es la denominada unión toracolumbar², constituida por las vértebras comprendidas entre los niveles T11 a L2. Su vulnerabilidad es debida a la transición de la rígida curva cifótica torácica a la lordosis lumbar. En esta región, el eje de carga se ve desplazado del cuerpo vertebral, amplio y estable, a los elementos posteriores de la columna, se pierde el soporte estructural que ofrecen las costillas y las facetas articulares sufren un cambio de orientación del plano coronal hacia el plano sagital. Por último, el segmento lumbar comprendido entre L3 y L5, presenta la mayor flexibilidad de todo el eje toracolumbar contando con unas facetas en disposición sagital y un disco intervertebral ancho. Las lesiones neuronales son menos frecuentes en este último segmento al contar con un canal medular más amplio que en los niveles superiores.

Para poder entender los mecanismos de lesión vertebral correctamente es fundamental introducir el concepto de estabilidad. La estabilidad clínica es definida³ como la capacidad de la columna para mantenerse cohesiva y asegurar unos desplazamientos vertebrales normales durante los movimientos fisiológicos del cuerpo evitando daños por irritación del cordón espinal o por cambios estructurales.

El conocimiento de las características biomecánicas de la columna vertebral se debe en gran medida a la aportación de las técnicas de imagen. Desde las primeras descripciones por radiología convencional de las fracturas vertebrales en los años 40⁴ y especialmente tras el descubrimiento de la tomografía computarizada, numerosos autores han tratado de establecer un modelo clasificatorio que refleje de manera óptima el mecanismo, morfología y pronóstico de la lesión. Cada nuevo modelo ha tomado influencias de estudios anteriores y ha tratado de actualizarlo con descripciones más precisas, o bien adaptarlo para facilitar su aplicabilidad en la práctica clínica diaria.

Para que una clasificación de fracturas vertebrales pueda ser adoptada de manera global tanto por radiólogos como por neurocirujanos y traumatólogos ha de ser lo suficientemente precisa para reflejar de manera correcta cada tipo de fractura y, a su vez, tener un planteamiento relativamente sencillo que reduzca la variabilidad que pueda existir entre los distintos observadores que decidan aplicarla. En 1983 Denis propuso una nueva clasificación de fracturas toracolumbares basándose en las imágenes por tomografía computarizada (TC) de 412 pacientes⁵. Con ella, Denis introdujo su famosa división de las “3 columnas”, delimitando una porción anterior, una media y una posterior en cada vértebra. Según el grado de afectación de cada segmento vertebral y la relación que establecían las vértebras fracturadas entre sí, Denis clasificó las fracturas en 4 tipos mayores (compresión, estallido, tipo “cinturón de seguridad” y fractura-luxación”) añadiendo 16 subdivisiones capaces de describir de la manera más precisa posible las características de cada fractura. La afectación de la columna media fue propuesta como el factor determinante para la pérdida de la estabilidad condicionando la actitud quirúrgica final. La experiencia posterior ha demostrado que fracturas que involucraban este segmento vertebral y que no acarrearán afectación neurológica pueden ser tratadas de manera conservadora. Además, pese a lograr una concordancia interobservador moderada en la clasificación de los 4 tipos mayores de fractura, ésta se perdía al intentar determinar cada subtipo⁶.

La siguiente contribución sustancial a la clasificación de las fracturas toracolumbares llegó en el año 1994 cuando Magerl propuso un modelo clasificatorio jerarquizado con 3 grupos mayores (A: Compresión; B: Distracción y C: Rotación) y un total de 53 subtipos para reflejar la morfología de cada fractura⁷. Pese a la gradación de la severidad que transmitía la clasificación de Magerl, ésta tuvo poco impacto en el algoritmo terapéutico de cada fractura y su gran complejidad teórica se veía penalizada con unos niveles de concordancia demasiado pobres⁸.

Desde su publicación en 2005, la clasificación con mayor aceptación es la TLICS (Thoracolumbar injury classification system)⁹. Como novedad se recalca la importancia del denominado Complejo Ligamentoso Posterior (CLP), estructura compuesta por el ligamento supraespinoso, ligamento interespinoso, ligamento amarillo y la cápsula de las articulaciones facetarias. El objetivo de esta clasificación es determinar la necesidad de tratamiento quirúrgico. Para ello, se asigna una puntuación a la fractura dependiendo de su morfología, de la presencia de sintomatología neurológica y de la lesión del CLP; de este modo, una puntuación superior a 4 indica la necesidad cirugía y una puntuación menor a 4 un tratamiento conservador. Hasta su publicación, TLICS era la única clasificación validada externamente con tasas de concordancia sobre la recomendación quirúrgica final del 96%¹⁰. Como crítica, a esta clasificación se le reprocha cierta falta de capacidad resolutoria en las fracturas que han recibido una puntuación de 4, donde deja la decisión quirúrgica en manos del juicio del propio cirujano que según su experiencia y la comorbilidad del paciente ha de decidir el tratamiento a seguir¹¹. También se ha señalado la dificultad de valorar el estado del CLP mediante TC, recomendando en caso de duda la realización de una resonancia magnética (prueba que demostró posteriormente un rendimiento diagnóstico limitado en la valoración del CLP en un estudio llevado a cabo por los propios autores de la TLICS¹²).

En 2013, el equipo que desarrolló la TLICS publicó una nueva clasificación de fracturas toracolumbares, la AOSpine TLICS¹³. En esta nueva clasificación se recuperaron y adaptaron conceptos de la clasificación de Magerl simplificándola a 9 patrones de fractura repartidos en tres grupos (A, B y C). Al igual que en la TLICS de 2005, se propone un sistema de clasificación jerarquizado ordenado de menor a mayor gravedad; con cuatro subtipos de fracturas "A", tres tipos de fracturas "B" y un solo tipo de fracturas

“C”. Además, esta nueva clasificación contempla dos modificadores (M1 y M2) que pueden condicionar el manejo terapéutico: la existencia de una banda de tensión no identificada visualizada por resonancia magnética o mediante la exploración física (M1) y la comorbilidad específica de cada paciente que pueda impedir la realización óptima de la cirugía como por ejemplo la espondilitis anquilosante o la existencia de quemaduras profundas en la zona afectada (M2). La clasificación también recoge la gravedad del déficit neurológico consecuencia de la fractura (Variable N) en cuatro niveles distintos de afectación.

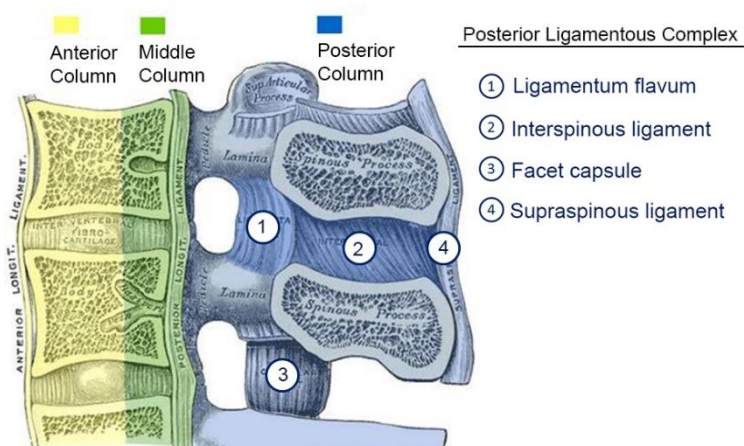


Figura 1: División de las tres columnas según Denis y representación del CLP

La clasificación AOSpine fue publicada junto a un estudio de fiabilidad en el que se cuantificaba la variabilidad inter e intraobservador entre nueve cirujanos clasificando un total de 110 casos¹³. Se obtuvo un coeficiente Kappa interobservador global de 0.64; la concordancia por tipos fue de 0.72 para las fracturas de tipo A, 0.58 para el tipo B y 0.7 para el tipo C. La concordancia en los subtipos de las fracturas fue algo menor, el peor resultado la obtuvo el tipo B2 con un coeficiente Kappa de 0.32. Los niveles de concordancia intraobservador remedaron los resultados obtenidos en el análisis intraobservador con valores Kappa algo superiores. Estudios posteriores han confirmado la buena concordancia existente en la aplicación de esta escala entre cirujanos^{14,15} y también han demostrado que los niveles de concordancia no varían sustancialmente según la experiencia del cirujano¹⁶.

Previamente se ha mencionado el papel indispensable que ha jugado la TC para el conocimiento anatómico y biomecánico de la columna vertebral. Gracias a su excelente capacidad para la visualización de las estructuras óseas y a los avances tanto en software como en hardware de las últimas dos décadas somos capaces de obtener imágenes con una resolución espacial milimétrica y generar reconstrucciones en todos los planos del espacio, así como en tres dimensiones. Es este el motivo por el que la TC constituye a técnica de imagen de elección para el diagnóstico de las fracturas vertebrales. Una vez obtenida la tomografía, el radiólogo es el encargado del post-procesado e interpretación de la imagen y su diagnóstico condicionará la decisión terapéutica que se adopte con el paciente.

A pesar de los numerosos trabajos que han estudiado la fiabilidad de la AOSpine TLICS, todos ellos han comprobado el grado de concordancia existente entre cirujanos. Para poder validar esta escala en nuestro entorno y garantizar su capacidad de aplicabilidad en un contexto de urgencias hospitalarias es preciso conocer si radiólogos con un perfil generalista son capaces de identificar y clasificar las fracturas vertebrales de manera correcta. Este estudio pretende dar respuesta a esta pregunta mediante un estudio de concordancia inter-observador entre un radiólogo que no ha recibido formación especializada en neurorradiología y un neurorradiólogo especialista con varios años de experiencia en el diagnóstico patología del sistema nervioso.

2. HIPÓTESIS Y OBJETIVOS

2.1. HIPÓTESIS

La existencia de un grado de concordancia suficiente en la clasificación de fracturas vertebrales toracolumbares traumáticas según el sistema clasificatorio AOSpine entre un neurorradiólogo especialista y un radiólogo de perfil general que permita su aplicación fiable en un contexto de urgencias hospitalarias.

2.2. OBJETIVOS

- **Objetivo principal:**
 - Determinar el grado de concordancia interobservador en la clasificación AOSpine de fracturas vertebrales toracolumbares entre radiólogos con distinto nivel de experiencia..

- **Objetivos secundarios:**
 - Determinar el grado de variabilidad intraobservador.
 - Describir las variables demográficas de las fracturas vertebrales traumáticas del territorio toracolumbar.
 - Describir las características morfológicas de las fracturas vertebrales toracolumbares según la clasificación AOSpine.
 - Describir el número y características de las fracturas identificadas por cada uno de los observadores.
 - Describir el manejo terapéutico llevado a cabo en los distintos tipos de fracturas vertebrales toracolumbares.

3. MATERIAL Y MÉTODOS

3.1. CONSIDERACIONES ÉTICAS

El diseño del estudio ha seguido las directrices de la Declaración de Helsinki para investigaciones biomédicas en humanos y el protocolo de estudio fue aprobado por El Comité Ético de Investigación Clínica de Aragón (CEICA) (Anexo I). No fue necesaria la solicitud de permiso a los pacientes para su inclusión en el estudio ni tampoco la firma de consentimiento informado. Para la obtención de datos no se realizó ninguna prueba diagnóstica ni ningún procedimiento terapéutico que no se incluyera en el protocolo médico establecido. Los datos fueron registrados en una base de datos codificada de forma que los pacientes no puedan ser identificados y su información personal no quede registrada.

3.2. DISEÑO DEL ESTUDIO

Retrospectivo, analítico observacional. Estudio de concordancia.

3.3. TAMAÑO MUESTRAL

La muestra fue escogida por conveniencia, al tratarse de un estudio retrospectivo. El número de pacientes fue determinado por el tamaño muestral empleado en estudios similares a este.

3.4. MUESTRA

Se analizaron un total de 451 fracturas de 60 pacientes que acudieron al servicio de urgencias por distintos tipos de traumatismo que pudieran afectar a la columna vertebral. Se llevó a cabo el análisis de concordancia de 60 fracturas. En aquellos pacientes que presentaban múltiples fracturas, se señaló la fractura con mayor gravedad y fue ésta la incluida en el análisis. El nivel vertebral afectado fue también recogido. Para asegurar que ambos observadores evaluaban el mismo nivel de afectación, una vez acabada la lectura de todas las fracturas, las clasificaciones realizadas por cada uno fueron puestas en común y se acordó el nivel a analizar.

40 de los 60 pacientes fueron recogidos consecutivamente entre los meses de octubre de 2017 y febrero de 2018. Dada la gran mayoría de fracturas leves existentes en ese intervalo de tiempo, y para asegurar que nuestra muestra contara con el espectro completo de lesiones que contempla esta clasificación, los 20 casos restantes fueron rescatados de los archivos del servicio.

3.4.1. Criterios de inclusión:

- Mecanismo traumático de la lesión.
- Lesión de curso agudo.
- Existencia de cortes de cortes finos (al menos de 1mm) de TC en el PACS del hospital.

3.4.2. Criterios de exclusión:

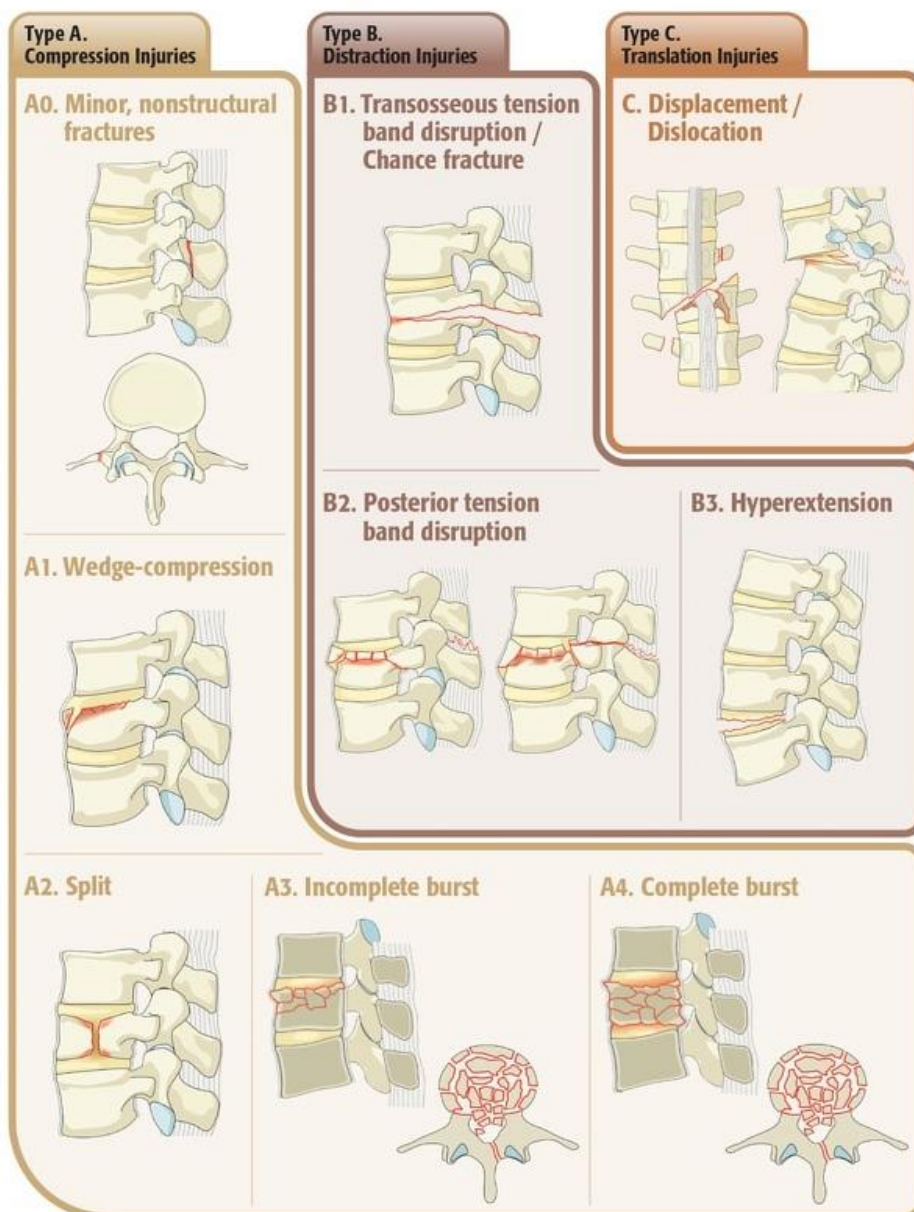
- Fracturas crónicas.
- Fracturas patológicas.
- Fracturas por impacto de objetos
- Fracturas sobre material de artrodesis.
- Presencia de malformaciones vertebrales.

3.5. INSTRUMENTOS DE MEDIDA

- Clasificación AOSpine para fracturas toracolumbares (*AOSpine Classification group*):



AOSpine Thoracolumbar Classification System



Tomado de: Kepler CK, et al. The Thoracolumbar AOSpine Injury Score (TL AOSIS)

- Fracturas tipo A: Fracturas por compresión del cuerpo vertebral.
 - Afectación exclusiva de los elementos anteriores (Cuerpo vertebral/disco). Se dividen en 5 tipos:
 - A0: Fracturas sin relevancia clínica que afectan a apófisis espinosas o transversas.
 - A1: Acuñaamiento con afectación de un único platillo vertebral sin afectación del muro posterior.
 - A2: Fracturas por división o tipo “pincer” donde la línea de fractura afecta ambos platillos vertebrales sin involucrar el muro posterior.
 - A3: Afectación de un único platillo y del muro posterior. La lámina vertebral puede verse afectada sin conllevar un cambio en la clasificación de la fractura.
 - A4: Afectación de los dos platillos vertebrales y del muro posterior. La lámina vertebral puede verse afectada sin conllevar un cambio en la clasificación de la fractura.



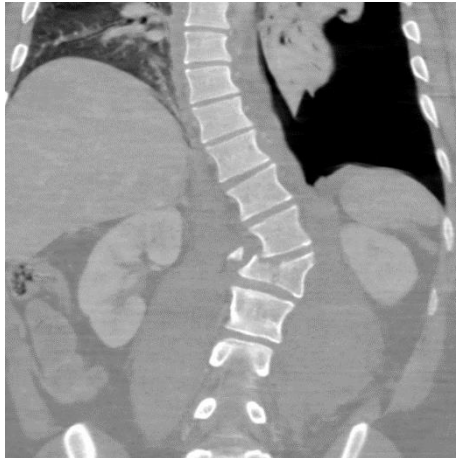
Figura 2: Fractura tipo “split” A2. Involucra ambos platillos vertebrales respetando el muro posterior

- Fracturas tipo B: Daño de la banda de tensión.
 - Daño de las bandas de tensión anterior (Ligamento longitudinal anterior + annulus fibrosus) o posterior (CLP).
 - B1: Daño óseo monosegmentario extendiéndose la fractura exclusivamente a lo largo de la superficie ósea sin daño de la banda de tensión posterior.
 - B2: Disrupción de la banda de tensión posterior pudiendo existir o no afectación ósea.
 - B3: Disrupción de la banda de tensión anterior (ligamento longitudinal anterior). Puede afectar al disco intervertebral o a un nivel vertebral exclusivamente.



Figura 3: Fractura por hiperextensión (B3) en un paciente con enfermedad de Forestier. Las fracturas por hiperextensión son características de procesos anquilosantes vertebrales.

- Fracturas tipo C: luxación vertebral.
 - Desplazamiento del nivel vertebral superior respecto al inferior superando el rango fisiológico. Debe existir afectación de la banda de tensión anterior y posterior.



Figuras 4 y 5: Fracturas por dislocación donde puede apreciarse el desplazamiento vertebral en plano coronal (Fi. 4) y la ruptura del complejo ligamentario anterior y posterior (Fig.5)

- Otras variables:
 - Variable N: Estado neurológico.
 - N0: Sin afectación neurológica.
 - N1: Déficit neurológico transitorio.
 - N2: Radiculopatía.
 - N3: Daño espinal incompleto/síndrome de cola de caballo.
 - N4: Daño espinal completo.
 - Variable M: Modificadores específicos.
 - M1: fracturas con daño indeterminado en las bandas de tensión basadas en pruebas de imagen complementarias como la resonancia magnética o por la exploración física.
 - M2: Comorbilidad específica de cada paciente que puede condicionar el resultado de la cirugía como la espondilitis anquilosante, la osteoporosis o las quemaduras de la piel sobre la porción vertebral dañada.



Figura 6: Paciente con espondilitis anquilosante (M2) y fractura de un sindesmofito.

3.6. PROTOCOLO DE RECOGIDA DE DATOS

La información referente a cada fractura fue recogida por uno de los dos radiólogos que llevaron a cabo el estudio. Éste se hizo responsable de la idoneidad de cada caso y de proporcionar al otro radiólogo la información necesaria para poder revisar las imágenes de TC de cada caso. También se encargó de anonimizar a cada paciente codificando los números de historia clínica siguiendo un criterio que sólo él y el otro observador conocían con el objetivo de mantener el máximo grado de confidencialidad.

3.6.1. *Primera fase: Búsqueda de casos y datos demográficos.*

Se revisaron de manera retrospectiva los escáneres toracoabdominales realizados de urgencias con antecedente traumático. Se buscó información demográfica y clínica en la historia clínica electrónica de los pacientes diagnosticados en la TC de urgencias de fractura vertebral toracolumbar entre los meses de octubre de 2017 y febrero de 2018.

Fueron incluidos todos los pacientes con diagnóstico de fractura toracolumbar de manera cronológica, sin importar la severidad de su fractura en un periodo de cuatro meses. A su vez se comprobó si las imágenes de dicha prueba poseían grosores de corte de como mínimo 1 mm para permitir la correcta identificación de las fracturas.

Con el fin de obtener una muestra con casos de mayor severidad para así abarcar todo el espectro de fracturas recogido por la clasificación AOSpine, se añadieron 20 casos pertenecientes al archivo de la sección de neurorradiología. Del mismo modo, la información demográfica y clínica referente a estos pacientes fue recogida de la historia clínica electrónica.

3.6.2. *Segunda fase: Primera vuelta de lectura y búsqueda de casos de archivo.*

Cada radiólogo contaba con una ficha de recogida de datos para la clasificación de las fracturas. Los radiólogos no revelaron el contenido de sus fichas hasta el final del estudio para evitar la posible influencia del uno sobre el otro.

3.6.3. *Tercera fase: Segunda vuelta de lectura y puesta en común de los datos.*

Las fracturas clasificadas en la segunda vuelta de lectura se analizaron al menos dos meses después de la primera lectura con el objetivo de minimizar en la medida de lo posible el sesgo que supondría recordar la anterior clasificación otorgada a cada caso.

3.6.4. *Cuarta fase: Puesta en común de resultados.*

Para el análisis de concordancia inter e intraobservador se utilizó únicamente el tipo de fractura según AOSpine, con independencia del nivel vertebral afectado. Las fracturas de menor severidad concomitantes en un mismo nivel fueron obviadas para el análisis de concordancia. Para asegurar que ambos observadores evaluaban el mismo nivel de afectación, en aquellos casos que presentaban fracturas múltiples se consensuó el nivel de la fractura a analizar. Una vez obtenidos los resultados de concordancia, ambos radiólogos pusieron en común los casos discrepantes para compartir qué les había llevado a clasificar de esa manera las fracturas y visualizaron las imágenes de resonancia magnética de esos pacientes para comprobar si la información obtenida en esta prueba modificaba la clasificación otorgada mediante TC.

3.7. ANÁLISIS ESTADÍSTICO

El análisis estadístico se llevó a cabo con el programa SPSS (*IBM Corporation, Somers, NY, United States*).

La descripción de las características de cada fractura, su localización y la información referente al sexo, clínica, factores modificadores, etiología y tratamiento de cada fractura fue expresada mediante porcentajes. La descripción de la edad se llevó a cabo mediante el valor mínimo y máximo, media, mediana y desviación estándar.

El análisis de concordancia interobservador global, la concordancia por tipos y por subtipos y la concordancia intraobservador global fue llevado a cabo mediante el coeficiente Kappa. Se consideró estadísticamente significativo un valor p menor de 0.05. Los coeficientes Kappa fueron interpretados siguiendo el sistema de gradación propuesto por Landis y Koch¹⁷:

- $\kappa \leq 0.2$ → Concordancia insignificante
- $\kappa > 0.2-0.4$ → Concordancia discreta
- $\kappa > 0.4-0.6$ → Concordancia moderada
- $\kappa > 0.6-0.8$ → Concordancia sustancial
- $\kappa > 0.8$ → Concordancia excelente.

El grado de concordancia interobservador por tipos (A, B y C) y por subtipos (p.ej. A1, A2, A3, A4...) se expresó mediante porcentajes de concordancia

4. RESULTADOS

4.1. ESTADÍSTICA DESCRIPTIVA

Se analizaron fracturas pertenecientes a 60 pacientes distintos. El neurorradiólogo especialista identificó 105 fracturas distintas y el radiólogo general 124. Si un paciente presentaba múltiples fracturas se escogió la de mayor gravedad para llevar a cabo el análisis de concordancia, contando por lo tanto con 60 fracturas principales.

La edad media de los pacientes incluidos en el estudio fue de 48.5 años, con una mediana situada de 45 años y una desviación estándar de 21,6. El paciente más joven era de 7 años y el más longevo de 93. El 50% de los pacientes fueron varones y el 50% mujeres.

El grado de daño neurológico se recogió siguiendo el método propuesto por la clasificación AOSpine¹⁰. El 75% de los pacientes no mostraron afectación neurológica. También fue recogido el tratamiento empleado en cada paciente. La mayoría de los pacientes (66,7%) fueron tratados de manera conservadora.

		Clínica			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	N0	45	75,0	75,0	75,0
	N1	2	3,3	3,3	78,3
	N2	2	3,3	3,3	81,7
	N3	2	3,3	3,3	85,0
	N4	3	5,0	5,0	90,0
	Nx	2	3,3	3,3	93,3
	NX	4	6,7	6,7	100,0
	Total	60	100,0	100,0	

Tabla 1: Grado de afectación clínica según AOSpine

El segmento más afectado del conjunto de fracturas toracolumbares fue la unión toracolumbar (comprendida entre los niveles T11 y L2). 27 de las 60 fracturas etiquetadas como principales afectaron este segmento (un 45%). El segmento dorsal (D1 a D10) fue el segundo más afectado con un total de 21 fracturas contabilizadas (35%). El nivel vertebral más afectado fue D11, con 11 fracturas en esta vértebra. Pueden consultarse todos los niveles afectados en el Anexo II.

En cuanto a su severidad, la mayoría de las fracturas identificadas por ambos radiólogos fueron clasificadas como A. El neurorradiólogo clasificó como A 45 fracturas (el 75%) y el radiólogo general 43 (71,6%). Por subtipos, la fractura más identificada por ambos radiólogos fue la A4 con 11 y 17 fracturas clasificadas como tales por el neurorradiólogo y el radiólogo general, respectivamente.

Se contabilizó el número, tipo y cantidad de pacientes con fracturas múltiples (aquellas producidas a la vez que la fractura principal). El neurorradiólogo identificó en ambas lecturas 45 fracturas múltiples en 22 pacientes distintos (el 36%), mientras que el radiólogo general identificó 57 fracturas múltiples en la primera lectura y 64 en la segunda lectura en 31 pacientes distintos (el 51%). El subtipo de fractura múltiple más frecuente fue el A0 con una prevalencia del 49 y del 56% en las revisiones del neurorradiólogo y del radiólogo general respectivamente.

Subtipo	Neurorradiólogo	Radiólogo gral.
A0	22 (49%)	36 (56%)
A1	14 (31%)	21 (33%)
A2	2 (4%)	2 (3%)
A3	4 (9%)	0
A4	2 (4%)	2 (3%)
B1	1 (2%)	1 (1%)
B2		2 (3%)
B3		
C		
Total	45	64

Tabla 2: Fracturas múltiples por subtipo

4.2. ESTADÍSTICA ANALÍTICA: ANÁLISIS DE CONCORDANCIA

4.2.1. Variabilidad interobservador:

El análisis del grado de concordancia interobservador fue calculado de bal mediante el coeficiente Kappa. Como se ha mencionado en el análisis estadístico, la interpretación del grado de concordancia se llevó a cabo siguiendo la propuesta por Landis y Koch.

La concordancia interobservador del conjunto de las fracturas principales clasificadas por su tipo (A, B o C) fue de 0.803, con un p-valor menor a 0.001, lo que supone un grado de concordancia excelente. Por el tipo B obtuvo la menor concordancia con un valor kappa de 0.712, concordancia moderada.

Tipo	Fracturas Neurorradiólogo	Fracturas Rad. Gen.	Número de acuerdos	Kappa	Significación
A	45 (75%)	43 (71.6%)	42	0.830	0.000
B	10 (16%)	11 (18.3%)	8	0.712	0.000
C	5 (8.3%)	6 (10%)	1	0.900	0.000
Total	60 (100%)	60 (100%)	3	0.803	0.000

Tabla 3: Concordancia por tipos

La concordancia interobservador del conjunto de las fracturas clasificadas por subtipos fue de 0.606, con un p-valor menor a 0.001, lo que supone un grado de concordancia sustancial. El rango de valores Kappa fue amplio, obteniendo una menor concordancia en las fracturas tipo A3 (k 0.306, concordancia discreta) y una concordancia máxima en el tipo A0 (k 1, concordancia excelente). El valor de significación estadística para las fracturas del tipo B2 fue mayor a 0.005, por lo que se considera estadísticamente no significativo.

Subtipo	Fracturas Neurorradiólogo	Fracturas Rad. Gen.	Número de acuerdos	Kappa	Significación
AO	5 (8.3%)	5 (8.3%)	5	1	0.000
A1	10 (16.6%)	17 (28.3%)	10	0.672	0.000
A2	2 (3.3%)	3 (5%)	1	0.375	0.003
A3	12 (20%)	4 (6.7%)	3	0.306	0.004
A4	17 (28.3%)	14 (23.3%)	11	0.610	0.000
B1	2 (3.3%)	3 (5%)	1	0.375	0.003
B2	4 (6.6%)	6 (10%)	2	0.348	0.006
B3	3 (5%)	2 (3.3%)	2	0.792	0.000
C	5 (8.3%)	6 (10%)	5	0.900	0.000
Total	60(100%)	60(100%)	40	0.606	0.000

Tabla 3: Concordancia por subtipos.

4.2.2. Variabilidad intraobservador:

Para comprobar la reproductibilidad de la clasificación se analizó la concordancia intraobservador del total de fracturas principales clasificadas por subtipos. La concordancia entre las propias mediciones de ambos radiólogos fue excelente. El neurorradiólogo obtuvo un valor kappa de 0.859 y el radiólogo general de 0.802 con valores de significación estadística menores a 0.001.

5. DISCUSIÓN

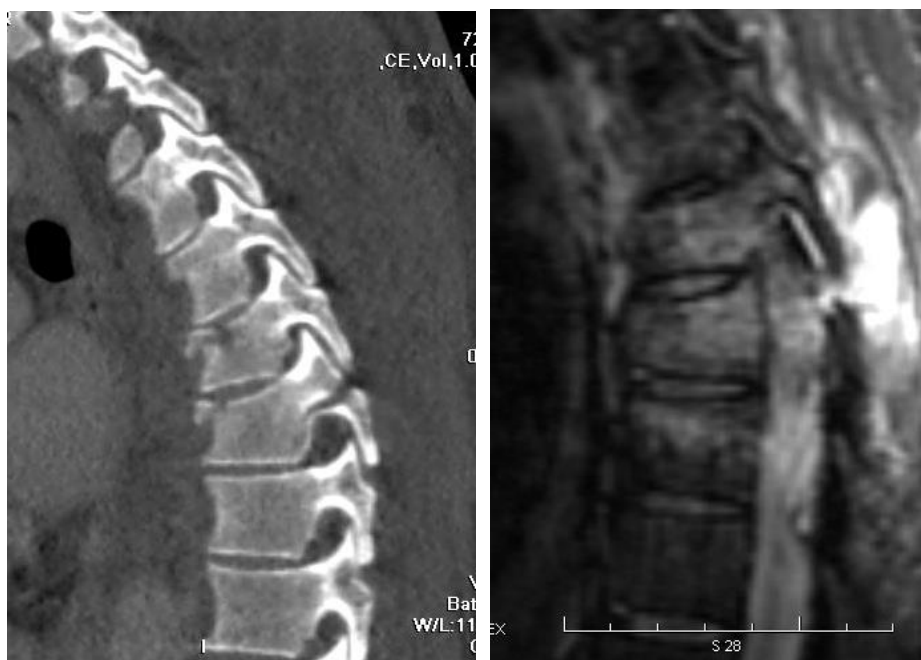
La correcta identificación y clasificación de las fracturas vertebrales traumáticas supone un reto diagnóstico para el radiólogo de urgencias. Es primordial que, independientemente de su nivel de experiencia, cualquier radiólogo este familiarizada con esta patología y con los sistemas clasificatorios empleados por neurocirujanos y traumatólogos para poder traducirles los hallazgos de las pruebas de imagen de una manera concisa y unificada.

Otros trabajos han estudiado la fiabilidad de esta clasificación entre cirujanos de distinto nivel de experiencia^{14,16} observándose resultados dispares en los valores de concordancia de estos estudios. En nuestro trabajo hemos obtenido niveles de concordancia “excelentes” en la clasificación de las fracturas por tipos (k 0.803) y “sustanciales” (k 0.606) en la clasificación por subtipos.

La clasificación de las fracturas tipo B ha obtenido el menor valor de concordancia (k 0.712, concordancia moderada). Este es un resultado observado en todos los estudios revisados sobre la fiabilidad de esta clasificación. La causa de esta menor concordancia es conocida y se debe a la dificultad de la evaluación del CLP mediante TC¹⁸. La visualización directa del CLP no es posible mediante TC y su ruptura tiene que sospecharse mediante signos indirectos. Como se ha mencionado, la integridad de este complejo es fundamental para la estabilidad de la columna y puede llegar a condicionar el manejo quirúrgico de la fractura. En estos casos la resonancia magnética puede contribuir a determinar la ruptura del CLP aunque la información proporcionada por esta técnica no debe ser usada de manera aislada, al proporcionar resultados poco fiables en las lesiones de menor gravedad¹⁹.

Nuestros resultados superan a los obtenidos por Cheng¹⁴, que es el estudio con un diseño más similar al nuestro.

Aunque no era un objetivo del estudio, una vez obtenidos los resultados de las pruebas de concordancia, ambos radiólogos pusieron en común los casos discrepantes, fundamentalmente aquellos relacionados con desacuerdos en clasificaciones del tipo B. Se revisaron las imágenes de resonancia magnética de estos pacientes visualizando el estado del CLP comprobando la influencia que podría tener la resonancia magnética en la decisión final sobre la gradación de las fracturas llegando a modificar el grado otorgado previamente.



Figuras 7a y 7b: Fractura clasificada por TC como B1 (Fig. 8a). La resonancia posterior demuestra ruptura del CLP con afectación del ligamento amarillo y el ligamento interespinoso (Fig. 8b).

A pesar del relativo buen grado de concordancia en la clasificación global por subtipos (k 0.606), el rango de valores de concordancia ha sido amplio obteniendo valores de concordancia de 1 y 0.9 en las fracturas de tipo A0 y C respectivamente, hasta valores muy discretos como los obtenidos en la clasificación de las fracturas A3, la cual ha obtenido el menor valor de concordancia con un k de 0.306. Los valores máximos de concordancia han sido obtenidos precisamente en las fracturas de menor y mayor gravedad. El alto grado de concordancia para las fracturas A0 es un hallazgo que se repite en varios de estudios de fiabilidad de la clasificación

AOSpine^{14,15,16} y es debido a la poca dificultad que entrañan las fracturas de las apófisis transversas o de espinosas. Sin embargo, la excelente concordancia de las fracturas del tipo C (k 0.9), las más graves, es un hallazgo que no se comparte con otros trabajos. Cheng obtuvo un valor k de 0.432 y Urrutia y Vaccaro de 0.7. La ruptura de los complejos ligamentosos anterior y posterior ofrece una imagen fácilmente identificable, este es un hecho de especial trascendencia al tratarse de fracturas de alto riesgo neurológico cuyo tratamiento ha de ser siempre quirúrgico.



Figuras 8a y 8b: Fractura L3-L4 C (Fig. 7a) con fractura L4 B1 asociada (Fig. 7b).

Dentro de las fracturas del tipo A, el subtipo que ha obtenido resultados más discordantes es el tipo A3 (k 0.306). La concordancia en la clasificación del grado A3, es pobre en todos los trabajos revisados^{13,14,15}. Schroeder investigó la dificultad que conlleva la correcta diferenciación de las fracturas A3 respecto a las A4²⁰ señalando la importancia que comporta disponer de imágenes reconstruidas en planos sagital y coronal para la correcta visualización de ambos platillos vertebrales, hallazgo que condiciona el paso de un grado A3 a A4.



Figura 9: Importancia del reformateo en los tres planos del espacio. Se visualiza fractura por estallido con afectación del muro posterior y aparentemente afectación única del platillo superior, A3 (Fig. 9b). En el reformateo coronal se puede visualizar como afecta a ambos platilos vertebrales, A4 (Fig.9c).

Cabe señalar la diferencia en el número de fracturas visualizadas por ambos radiólogos donde el radiólogo de perfil general logró identificar un mayor número de fracturas que el neurorradiólogo, 64 frente a 45. Este resultado podría explicarse por la posible mayor “cautela” con la que el radiólogo general estudió las imágenes de TC fruto de una menor experiencia. Del mismo modo, las fracturas principales captaron la atención del neurorradiólogo y se enfrentó a ellas con una mayor seguridad pasando por alto posibles fracturas menores a otros niveles. No obstante, el mayor porcentaje de fracturas múltiples fueron clasificadas con los grados de menor severidad de toda la escala: A0 y A1 y no hubieran condicionado en exceso el manejo terapéutico del paciente.



Figura 10: Paciente que presenta múltiples fracturas de apófisis transversas izquierdas lumbares. La presencia de fracturas vertebrales múltiples de poca relevancia clínica (AO) es un hallazgo frecuente.

La concordancia intraobservador fue excelente en ambos radiólogos. El neurorradiólogo obtuvo una concordancia algo mayor que el radiólogo general, con un valor kappa de 0.859 frente a 0.802 por parte radiólogo general. Estos resultados superan ligeramente a los obtenidos por Vaccaro¹³ y reflejan la buena reproducibilidad de esta escala.

Este estudio cuenta con una serie de limitaciones que es necesario señalar. La principal limitación a la que se enfrenta este estudio respecto a otros trabajos similares estriba en el reducido número de observadores que han realizado la clasificación de las fracturas. El análisis de concordancia sería más representativo si hubiéramos contado con mayor número de radiólogos de perfil general y subespecialistas en neurorradiología.

La escasa cantidad de fracturas graves frente a la abundancia de las más leves ha producido que contemos con pocos casos en determinados subtipos. Hubiera sido necesario contar con un mayor número de fracturas graves ya que son estas las que implican una mayor dificultad en su diagnóstico y clasificación y los niveles de concordancia globales podrían haberse visto afectados.

Del mismo modo, determinados subtipos de fracturas han sido particularmente escasos por lo que los resultados del análisis de concordancia de estos casos podrían verse modificados si contáramos con un mayor número de fracturas.

Pese a contar en todos los casos con cortes de escáner de 1mm que permitían el reformateo de las imágenes a planos axiales y coronales, pocos casos contaban con las características técnicas idóneas para una reconstrucción ósea de máxima precisión (cortes de 0.5mm con filtro óseo y field of view centrado en la columna vertebral) pudiendo dificultar la interpretación de algunas imágenes.

6. CONCLUSIONES

La clasificación de las fracturas vertebrales toracolumbares mediante el sistema propuesto por AOSpine tiene unos niveles de concordancia y reproductibilidad adecuados. Los resultados obtenidos en este trabajo muestran que la clasificación AOSpine podría ser aplicada de manera fiable por radiólogos con distinto nivel formativo al existir niveles de concordancia excelentes en ambos tipos de evaluadores.

Esta concordancia es más elevada en la clasificación de las fracturas según su tipo, siendo un hecho relevante al tener implicaciones pronósticas y terapéuticas. La valoración del complejo ligamentoso posterior mediante tomografía computarizada es limitada produciendo resultados discordantes en las clasificaciones de los examinadores para la identificación de las fracturas tipo B.

Las fracturas de menor y mayor gravedad son identificadas con facilidad por ambos radiólogos con unos niveles de concordancia excelentes. La reconstrucción de las imágenes en los planos sagital y coronal tiene especial relevancia para la valoración morfológica de la vértebra permitiendo apreciar detalles que pueden modificar la pertenencia de una fractura a un determinado subtipo.

Ambos radiólogos han logrado niveles de concordancia intraobservador excelentes reflejando la buena reproducibilidad de la clasificación

El grado de experiencia del radiólogo puede limitar la identificación de un mayor número de fracturas posiblemente por una mayor confianza y familiaridad con esta patología. No obstante, las fracturas no visualizadas no tienen interés pronóstico.

7. ANEXOS

ANEXO I



Informe Dictamen Favorable Trabajos académicos

C.P. - C.I. P118/237

25 de julio de 2018

Dña. María González Hinjos, Secretaria del CEIC Aragón (CEICA)

CERTIFICA

1º. Que el CEIC Aragón (CEICA) en su reunión del día 25/07/2018, Acta N° 14/2018 ha evaluado la propuesta del Trabajo:

Título: CONCORDANCIA Y REPRODUCTIBILIDAD DE LA NUEVA CLASIFICACIÓN AOSPINE PARA FRACTURAS DEL TERRITORIO TORACO-LUMBAR.

Alumno: Julián José García Maroto

Director: José María Artigas Martín

Versión protocolo: v1.1, 05/09/2018

2º. Considera que

- El proyecto se plantea siguiendo los requisitos de la Ley 14/2007, de 3 de julio, de Investigación Biomédica y los principios éticos aplicables.
- El Tutor/Director garantiza la confidencialidad de la información, el adecuado tratamiento de los datos en cumplimiento de la legislación vigente y la correcta utilización de los recursos materiales necesarios para su realización.

3º. Por lo que este CEIC emite **DICTAMEN FAVORABLE a la realización del proyecto.**

Lo que firmo en Zaragoza

GONZALEZ Firmado digitalmente
HINJOS MARIA por GONZALEZ HINJOS
DNI 03857456B MARIA - DNI 03857456B
Fecha: 2018.07.27
10:49:56 +02'00'

María González Hinjos
Secretaria del CEIC Aragón (CEICA)

ANEXO II

		Nivel				
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado	
Válido	D1	1	1.7	1.7	1.7	
	D10-D11	3	5.0	5.0	6.7	
	D11	1	1.7	1.7	8.3	
	D11-D12	1	1.7	1.7	10.0	
	D12	10	16.7	16.7	26.7	
	D12-L1	2	3.3	3.3	30.0	
	D2	1	1.7	1.7	31.7	
	D2-D3	1	1.7	1.7	33.3	
	D3	1	1.7	1.7	35.0	
	D4-D5	1	1.7	1.7	36.7	
	D5	1	1.7	1.7	38.3	
	D6	1	1.7	1.7	40.0	
	D7	3	5.0	5.0	45.0	
	D7-D8	2	3.3	3.3	48.3	
	D8	1	1.7	1.7	50.0	
	D9	3	5.0	5.0	55.0	
	L1	11	18.3	18.3	73.3	
	L2	5	8.3	8.3	81.7	
	L3	5	8.3	8.3	90.0	
	L3-L4	1	1.7	1.7	91.7	
	L4	1	1.7	1.7	93.3	
	L5	4	6.7	6.7	100.0	
	Total		60	100.0	100.0	

Anexo II: Niveles de afectación vertebral

8. BIBLIOGRAFIA

1. Hu R, Mustard CA, Burns C. Epidemiology of incident spinal fracture in a complete population. *Spine (Phila Pa 1976)* 1996;21:492–9.
2. Izzo R, Guarnieri G, Guglielmi G, Muto M. Biomechanics of the spine. Part I: spinal stability. *Eur J Radiol.* 2013;82(1):118-26.
3. Pope MH, Panjabi M. Biomechanical definitions of spinal instability. *Spine* 1985;10:255–6.
4. Nicoll EA. Fractures of the dorso-lumbar spine. *J Bone Joint Surg Br.* 1949;31:376–394.
5. Denis F. The three column spine and its significance in the classification of acute thoracolumbar spinal injuries. *Spine.* 1983;8:817–831.
6. Wood KB, Khanna G, Vaccaro AR, Arnold PM, Harris MB, Mehbod AA. Assessment of two thoracolumbar fracture classification systems as used by multiple surgeons. *JBone Joint Surg Am.* 2005;87(7):1423-9.
7. Magerl F, Aebi M, Gertzbein SD, Harms J, Nazarian S. A comprehensive classification of thoracic and lumbar injuries. *Eur Spine J.* 1994;3(4):184-201.
8. Oner FC, Ramos LM, Simmermacher RK, 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; 11:235–45.
9. Vaccaro AR, Lehman RA Jr, Hurlbert RJ, et al. A new classification of thoracolumbar injuries: the importance of injury morphology, the integrity of the posterior ligamentous complex, and neurologic status. *Spine (Phila Pa 1976)* 2005; 30:2325–33;
10. 30. Vaccaro AR, Baron EM, Sanfilippo J, et al. Reliability of a novel classification system for thoracolumbar injuries: the thoracolumbar injury severity score. *Spine (Phila Pa 1976)* 2006;31:S62–9
11. Schroeder GD, Harrop JS, Vaccaro AR. Thoracolumbar Trauma Classification. *Neurosurg Clin N Am.* 2017;28(1):23-29.
12. Vaccaro AR, et al. Injury of the posterior ligamentous complex of the thoracolumbar spine: a prospective evaluation of the diagnostic accuracy of magnetic resonance imaging. *Spine (Phila Pa 1976)*
13. Kepler CK, Vaccaro AR, Schroeder GD, et al. The Thoracolumbar AOSpine Injury Score (TL AOSIS). *Global Spine J* 2016;6(4):329–34; and Vaccaro AR, Oner C, Kepler CK, et al. AOSpine thoracolumbar spine injury classification system: fracture description, neurological status, and key modifiers. *Spine (Phila Pa 1976)* 2013; 38:2028–37.
14. 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.

15. Urrutia J, Zamora T, Campos M, Palma J, Mobarec S, Prada C. An independent interobserver reliability and intraobserver reproducibility evaluation of the new AOSpine Thoracolumbar Spine Injury Classification System. *Spine (Phila Pa 1976)*. 2015 J;40(1):E54-8.
16. Sadiqi S, Oner FC, Dvorak MF, Aarabi B, Schroeder GD, Vaccaro AR. 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 (Phila Pa 1976)*. 2015;40(23):E1250-6.
17. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174.
18. Barcelos AC, Joaquim AF, Botelho RV. 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):1135-43.
19. Vaccaro AR, JA, Saravanja D, Anderson DG, Hilibrand AS, Albert TJ, Fehlings MG, Morrison W, Flanders AE, France JC, Arnold P, Anderson PA, Friel B, Malfair D, Street J, Kwon B, Paquette S, Boyd M, Dvorak MF, Fisher C. Injury of the posterior ligamentous complex of the thoracolumbar spine: a prospective evaluation of the diagnostic accuracy of magnetic resonance imaging. *Spine (Phila Pa 1976)*. 2009;34(23):E841-7.
20. Schroeder GD, Kepler CK, Koerner JD, Chapman JR, Bellabarba C, Oner FC, Reinhold M, Dvorak MF, Aarabi B, Vialle L, Fehlings MG, Rajasekaran S, Kandziora F, Schnake KJ, Vaccaro AR. Is there a regional difference in morphology interpretation of A3 and A4 fractures among different cultures? *J Neurosurg Spine*. 2015; 9:1-8.

* Figura 1 tomada de: <https://upload.orthobullets.com/topic/2003/>