ORIGINAL ARTICLE

Concordance in the radiological diagnosis of thoracolumbar spine fractures ☆

M. Hirschfeld [a,*], **M. Rodriguez** [a], **A.M. Cerván** [a], **J.A. Ortega** [a], **F. Rivas-Ruiz** [b], **E. Guerado** [a]

[a] *Departamento de Traumatología y Cirugía Ortopédica, Hospital Universitario Costa del Sol, Universidad de Málaga, Marbella, Málaga, Spain*
[b] *Unidad de Investigación, Agencia Sanitaria Costa del Sol, Red de Investigación en Servicios de Salud y Enfermedades Crónicas REDISSEC, Marbella, Málaga, Spain*

Abstract
Introduction: Thoracolumbar spine fractures are frequent and severe. Early diagnosis and appropriate treatment to obtain good clinical results is essential, with many classifications being proposed for this purpose.
Objective: To determine the external validity of radiographic and computed tomography (CT) measurements for the most used classifications, and decide on the type of treatment required. The working hypothesis is the existence of external validity of radiographic measurements.
Material and method: A sample of patients with thoracolumbar fracture was selected. Three spine specialists and a resident performed measurements on anteroposterior and lateral radiographic images as well as coronal, sagittal and axial CT slices. Fractures were classified as stable or unstable, evaluating the degree of intra-and interobserver agreement based on a standard observer. Sagittal index of Farcy, lateral wedging, Beck Index, traumatic regional angulation and channel occupancy were studied.
Results: All indicators studied, except the lateral wedging, showed a high degree of concordance.
Conclusions: Instability determinants studied with radiographs and CT, which had obtained statistical significance, are reliable and accurate for the classification of thoracolumbar fractures and, therefore, to indicate an appropriate treatment.
© 2014 SECOT. Published by Elsevier España, S.L.U. All rights reserved.

PALABRAS CLAVE
Fracturas de raquis;
Diagnóstico;
Variabilidad
interobservador;
Variabilidad
intraobservador

Concordancia en el diagnóstico radiológico de las fracturas del raquis toracolumbar

Resumen

Introducción: Las fracturas del raquis toracolumbar son muy frecuentes y graves, siendo imprescindible su diagnóstico precoz y tratamiento adecuado para obtener buenos resultados clínicos. Para este propósito se han descrito numerosas clasificaciones e índices.

Objetivo: Estudiar la validez externa de las mediciones radiográficas y de la tomografía computarizada (TC) para las clasificaciones más usadas y decidir el tipo de tratamiento que precisan. La hipótesis operativa consiste en la existencia de validez externa de las mediciones radiográficas.

Material y método: Se seleccionó una muestra de pacientes con fractura toracolumbar. Tres especialistas de raquis y un médico residente realizaron mediciones sobre imágenes radiográficas anteroposterior y lateral, así como sobre cortes coronales, sagitales y axiales de TC, clasificándose las fracturas como estables o inestables, evaluándose el grado de concordancia intra e interobservador, basados en un observador estándar. Se estudiaron las variables índice sagital de Farcy, acuñamiento lateral, índice de Beck, angulación regional traumática y ocupación del canal.

Resultados: Todos los indicadores estudiados, excepto el acuñamiento lateral, presentaron un alto grado de concordancia.

Conclusiones: Los determinantes de inestabilidad estudiados en radiografías simples y TC que han obtenido significación estadística son fiables y precisos para caracterizar las fracturas toracolumbares y, por tanto, para indicar un tratamiento adecuado.

© 2014 SECOT. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Introduction

Spine fractures are frequent and present high morbidity, with an incidence of 64 cases, 4 with spinal cord injuries, per 100,000 inhabitants per year. Overall mortality is from 4–5%[1] with a higher prevalence in males between 15 and 25 years of age, due to traffic accidents.[2] Thoracolumbar fractures (TL) are the most common.

Accurate, early diagnosis is necessary in order to administer appropriate treatment. To meet this objective there are several variables and severity classifications for thoracic and TL fractures based on radiological imaging,[3–7] although, to our knowledge as authors, there are no concordance studies to enable determination of their external validity.

The aim of our study is to determine the external validity of the most frequently described variables of conventional X-ray and computed technology (CT) measurements with an operational hypothesis of a high degree of intra-and interobserver concordance.

Material and method

Data

A retrospective study was performed, the main variable of which was a diagnosis of spine fracture according to clinical criteria of trauma history and pain and the radiological criteria which is explained later on. Variables studied include gender, age, fracture level and type of treatment received (conservative or surgical).

A patient sample of 147 patients was obtained with the use of a prospective database on the HP Doctor v.2.22 (Hewlett Packa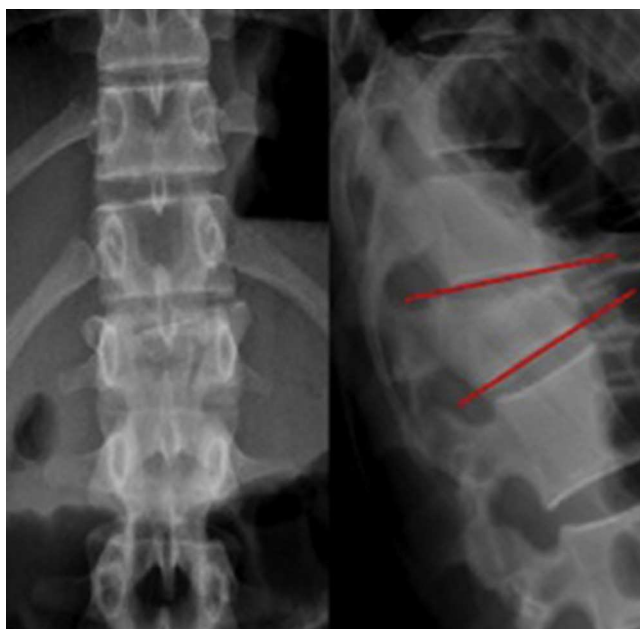rd) software application between 2009 and 2012, from which 79 patients were selected whose fracture level was localized in the TL segment (T11-L2). Data handling was covered by the personal Data Protection Act.[8]

Design

One of the researchers made an evaluation study which included the anteroposterior and lateral X-rays as well as coronal, sagittal and axial CT imaging sections for each patient in the sample. The observers were 3 specialists from the spine unit of our hospital, and one 4th year orthopaedic and trauma resident surgeon. All of them worked in our hospital. Two measurements were taken, 6 months apart, from these images using the Centricity (GE Healthcare) programme measuring tool, of the instability markers described below (quantitative variables). The observers defined each fracture as stable or unstable (dichotomous qualitative variable), according to the results of these measurements. The intra- and interobserver correlation was thus evaluated respectively. The measurements made by the observer with the greatest experience were taken as the benchmark measurements, with the intention of comparing the qualitative ''instability'' variable with that made by the other observers with respect to this standard, thus determining the level of validity of the diagnostic testing.

Instability markers

Sagittal index of Farcy[3]

This is measured in the sagittal plane and determines the wedging level in this plane. It is an angle comprised of 2 lines resulting from projections of the lower discs of the
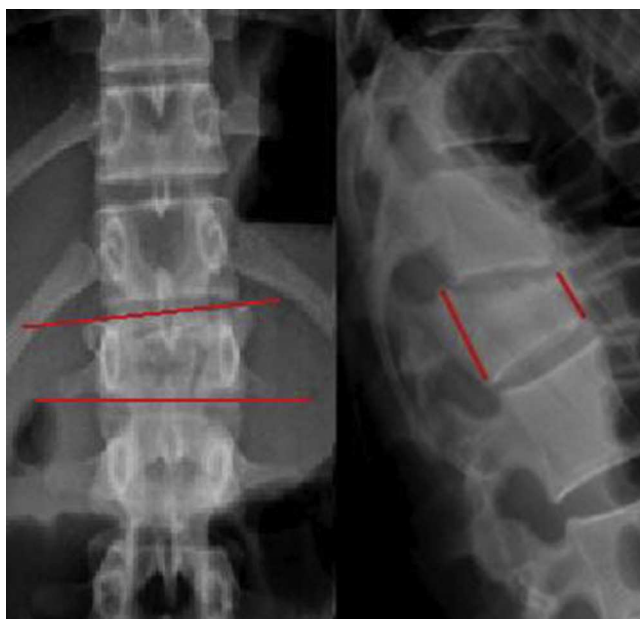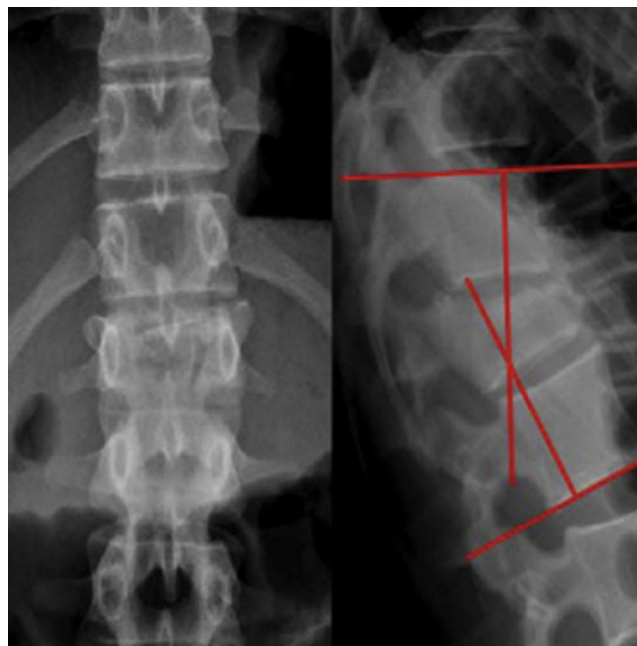
Figure 1   Sagittal index of Farcy.

cranial vertebra to the fractured vertebra and from the latter. It is considered unstable when the value is higher than 16°, corrected according to the level of the fracture. Correction is made by deducting 5° from the measurement in the case of thoracic level fractures and adding 10° at lumbar level; fractures localized at junction level do not require correction (Fig. 1).

Lateral wedging[4]
This is measured in a coronal plane. It is an angle formed by projections from the upper and lower discs of the fractured

Figure 2   Lateral wedging (A). Berk index (B).

vertebra. It is considered unstable when the angle exceeds 20° (Fig. 2, left).

Beck Index[5]
This is measured in the sagittal plane and defines the vertebral wedging percentagewise, measuring the difference in height in the anterior wall compared with the posterior wall in the fractured vertebra. Instability values are determined of above 50% (Fig. 2, right).

Regional traumatic angulation[6]
This determines the deformity produced in the injured vertebral segment. It is obtained in the sagittal plane by obtaining an angle the vectors of which are the projections of the upper disc from the cranial vertebra and the lower disc of the caudal vertebra to the fractured vertebra. This is considered an unstable fracture when there is regional traumatic angulation (RTA) above 20°, corrected according to the level of the fracture (Fig. 3).

Spinal canal occupation[7]
Defined percentagewise by the level of invasion of the spinal canal. The McAfee instability criteria include a compromise of the spinal canal by free bone fragments, demonstrated by CT scan.[9] For the purposes of our study we consider a compromise above 50% as unstable (Fig. 4).

Statistical analysis

A descriptive analysis with measurements of central tendency was performed, with dispersion for quantitative variables and frequency distribution for qualitative variables. For quantitative variable concordance evaluation between the 4 observers the intraclass correlation coefficient was used for individual measurements and the interelement correlation matrix, and for the qualitative
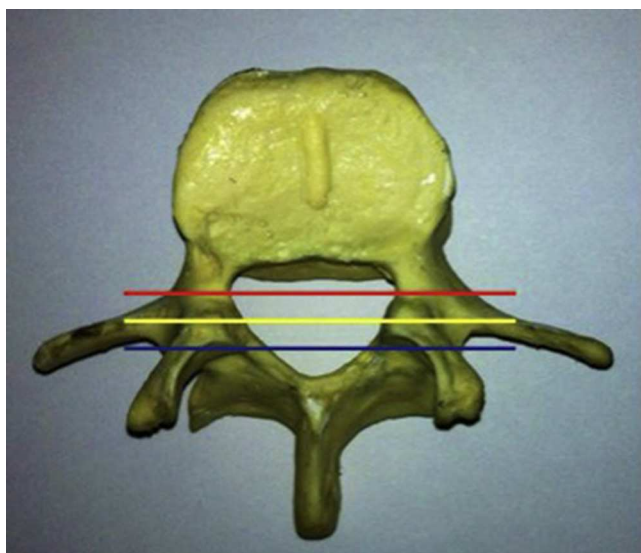

Figure 3   Traumatic regional angulation.

Figure 4 Spinal canal occupancy.

instability variable the kappa Jacknife[10] was used. A 95% confidence interval was included in both tests. The levels of concordance according to the kappa value were zero ($K = 0$), insignificant ($K = 0-0.20$), medium ($K = 0.21-0.40$), moderate ($K = 0.41-0.60$), substantial ($K = 0.61-0.0$) and almost perfect ($K = 0.81-1.00$).[11]

A multivariate logistic regression was subsequently made for each of the evaluators against the ''instability'' result variable (assessed by the observer with the greatest experience), including as independent variables the tests which were reliable in interobserver analysis (Sagittal index of Farcy [IF], Beck [IB] index, RTA, spinal canal cord occupancy [CO]). We included the odds ratio (OR) in addition to the 95% confidence intervals (CI 95%) and the table of values observed compared with those predicted (predicted probability ranges between 0 and 1). The statistical significance level was established at $P < .05$.

Results

Univariate analysis

Age: mean patient age was 59.36, with a standard deviation of 17.66 years of age and minimum and maximum values of 16 and 87 years of age, respectively.

Gender: there were 25 males (31.6%) and 54 females (68.4%).

Treatment type: 40 patients (50.63%) received conservative treatment and the remaining 39 patients (49.37%) received surgical treatment.

Fracture level: there were 5 cases (6.32%) at T11 level, 20 cases (25.31%), at T12 level, 42 cases (53.16%) at L1 level and 12 cases (15.18%) at L2 level (Fig. 5).

Degree of interobserver correlation

Quantitative variables
All instability markers (IF 0.794; IB 0.658; OC 0.916; RTA 0.895) obtained a substantial degree of correlation, except for the lateral wedging, with a value of 0.393 (Fig. 6).
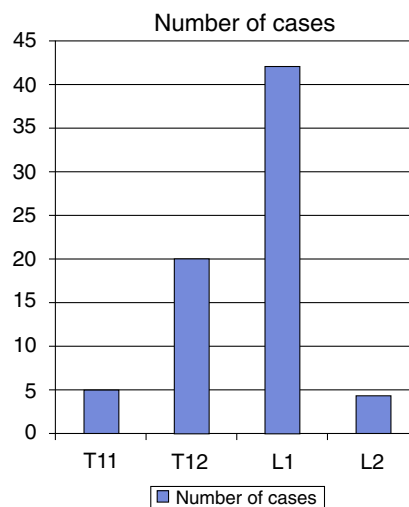

Figure 5 Descriptive frequency of fractured vertebra level.

Qualitative variable
The ''stability/instability'' variable obtained a value of 0.17 (Table 1).

Degree of intraobserver correlation

Quantitative variables
Comparing the measurements of each observer with those they had made 6 months previously we observed that all of them showed a high degree of concordance except one, which we ruled out for subsequent analysis as it was not stable over time.

Qualitative variable
All patients had a notable degree of concordance with regards to this variable: observer 1 (0.8361), observer 2 (0.6834), observer 3 (0.9071) and observer 4 (0.9738) (Fig. 7).

Degree of concordance vs. reference value

Multivariate logistic regression achieved high concordance results of 87.1% and 85.7% (Tables 2 and 3). Three independent variables studied with high correct evaluation
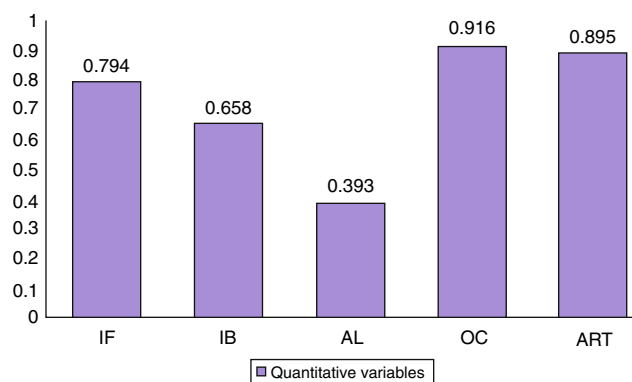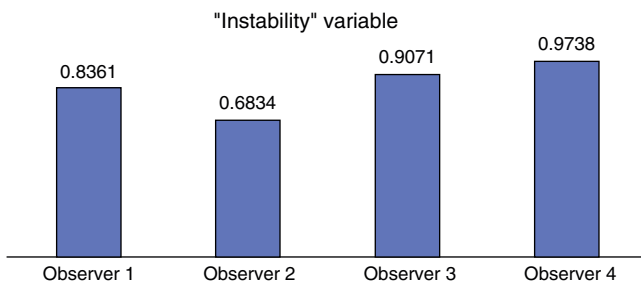

Figure 6 Instability markers evaluated using interclass correlation coefficient for individual measurements and correlation matrix of inter elements.

Table 1 Frequency and degree of consensus between observers on evaluating the ''stability/instability'' variable (kappa-Jacknife = 0.17).

	Frequency	Percentage	Valid percentage	Accumulated percentage
Valid				
2	15	19.0	19.2	19.2
3	8	10.1	10.3	29.5
4	55	69.6	70.5	100.0
Total	78	98.7	100.0	
Lost				
System	1	1.3		
Total	79	100.0		

Figure 7 Degree of intraobserver correlation for qualitative ''stability/instability'' variable.

percentages are included (IF, IB, OC), mainly in stable fractures (Tables 4 and 5).

Discussion

Almost 90% of vertebral fractures occur in the TL junction,[1] and it is important to have a universally accepted classification which serves as a guideline regarding the severity of the fracture and its stability, thereby supposedly leading to an indication of the most appropriate treatment.[12] Several classifications have appeared over time regarding spinal fractures, aimed at defining their level of deformity, stability and neurological risk,[13] according to vertebral morphology, the integrity of soft tissues, especially of the posterior ligamentous complex,[14] or the neurological risk. The references contain a great many articles which assess these classifications using correlation, concordance or reproducibility[15,16,12,17] and they usually conclude that there is a need for additional studies to prove the superiority of one of these classifications over the others.[18–20]

For the diagnosis of patients with spinal column trauma with suspicion of severity, CT is the method of choice to assess the degree of vertebral collapse and measure the diameter of the spinal canal. This imaging method is superior to the others in detecting vertebral fractures and

Table 2 Degree of concordance of observer 2 with regards to the reference value (multivariate logistic regression).

Observed	Predicted		
	Unstable 1		Correct percentage
	No	Yes	
Step 1			
unstable? 1			
No	39	5	88.6
Yes	8	18	69.2
Overall percentage			81.4
Step 2			
unstable? 1			
No	40	4	90.9
Yes	7	19	73.1
Overall percentage			84.3
Step 3			
unstable? 1			
No	40	4	90.9
Yes	5	21	80.8
Overall percentage			87.1

Table 3 Degree of concordance of observer 3 with regards to the reference value (multivariate logistic regression).

Observed	Predicted		
	Unstable 1		Correct percentage
	No	Yes	
Step 1			
Unstable? 1			
No	41	6	87.2
Yes	7	23	76.7
Overall Percentage			83.1
Step 2			
Unstable? 1			
No	40	7	85.1
Yes	8	22	73.3
Overall percentage			80.5
Step 3			
Unstable? 1			
No	42	5	89.4
Yes	6	24	80.0
Overall percentage			85.7

Table 4 Analysis of the independent variables included in the logistic regression (observer 2).

	B	E.T.	Wald	gl	Sig.	Exp (B)	ıC 95.0% for EXP (B)	
							Inferior	Superior
Step 1 (a)								
Occupancy 2	0.144	0.033	18.860	1	0.000	1.155	1.082	1.233
Step 2 (b)								
Constant	−2.102	0.481	19.101	1	0.000	0.122		
Farcy Index 2	0.251	0.081	9.646	1	0.002	1.286	1.097	1.507
Occupancy 2	0.142	0.042	11.654	1	0.001	1.153	1.062	1.251
Constant	−5.492	1.427	14.820	1	0.000	0.004		
Step 3 (c)								
Farcy Index 2	0.199	0.077	6.658	1	0.010	1.221	1.049	1.420
Beck Index 2	0.059	0.026	5.038	1	0.025	1.061	1.008	1.117
Occupancy 2	0.147	0.044	11.048	1	0.001	1.158	1.062	1.263
Constant	−6.989	1.838	14.463	1	0.000	0.001		

unstable injuries. However, CT is of limited value in evaluating the degree of soft tissue injury.[1] Magnetic resonance (MR) imaging is the method of choice for evaluating spinal column injuries, vertebral bone marrow oedema and ligament injuries[21]; the latter is also considered a criterion of instability.

We have evaluated different instability markers in our work, studying the inter and intraobserver correlation level, to determine their reliability with diagnostic and therapeutic objectives and, by extension, their application to the most widely used classifications.

Initially we evaluated the degree of intra-and interobserver correlation of some instability parameters in thoracolumbar spine fractures. From this first analysis we rejected lateral wedging due to a lack of acceptable degree of correlation, and one observer who did not remain stable over time, due to a median concordance between their 2 measurements. We also evaluated the degree of interobserver concordance of a dichotomous qualitative variable (stable/unstable), obtaining a low Kappa value, which could be interpreted as an insignificant level of agreement, although it is attributable to the number of observers, since on studying the results in detail, we see that the 4 observers agree in 55 patients (70.5%).

Two conclusions were obtained in a secondary analysis by means of a multivariate logistic regression. The first was that out of the markers studied only 3 (IF, IB, OC) obtained high correct percentages of evaluation, especially in stable fractures. The second was that the degree of concordance compared with the reference value was high, above 85%.

In light of our findings we can confirm that the evaluated diagnostic testing and the IF, IB and OC instability markers are reliable and accurate for characterizing these fractures.

Table 5 Analysis of the independent variables included in the logistic regression (observer 3).

	B	E.T.	Wald	gl	Sig.	Exp (B)	ıC 95.0% for EXP (B)	
							Inferior	Superior
Step 1 (a)								
Farcy Index 3	0.287	0.064	20.016	1	0.000	1.332	1.175	1.510
Constant	−4.045	0.892	20.539	1	0.000	0.018		
Step 2 (b)								
Farcy Index 3	0.259	0.070	13.868	1	0.000	1.296	1.131	1.486
Occupancy 3	0.082	0.027	9.083	1	0.003	1.086	1.029	1.145
Constant	−4.564	1.051	18.850	1	0.000	0.010		
Step 3 (c)								
Farcy Index 3	0.189	0.078	5.831	1	0.016	1.209	1.036	1.409
Beck Index 3	0.057	0.029	3.844	1	0.050	1.059	1.000	1.121
Occupancy 3	0.074	0.027	7.847	1	0.005	1.077	1.023	1.135
Constant	−5.376	1.284	17.538	1	0.000	0.005		

X-ray images using simple anteroposterior and lateral radiography, together with CT imaging, have a high degree of diagnostic concordance.

Level of evidence

Evidence level V.

Ethical responsibilities

Protection of human beings and animals

The authors declare that no experiments were performed on humans or animals for this investigation.

Data confidentiality. The authors declare that no patient data appear in this article.

Right to privacy and informed consent. The authors declare that no patient data appear in this article.

Conflict de interests

The authors have no conflict of interests to declare.

References

1. Wood KB, Li W, Lebl DS, Ploumis A. Management of thoracolumbar spine fractures. Spine J. 2014;14:145–64.
2. Hu R, Mustard CA, Burns C. Epidemiology of incident spinal fracture in a complete population. Spine. 1996;21:492–9.
3. Farcy JP, Weidenbaum M, Glassman SD. Sagittal index in management of thoracolumbar burst fractures. Spine. 1990;5:958–65.
4. Masharawi Y, Salame K, Mirovsky Y, Peleg S, Dar G, Steinberg N, et al. Vertebral body shape variation in the thoracic and lumbar spine: characterization of its asymmetry and wedging. Clin Anat. 2008;21:46–54.
5. Diel P, Röder C, Perler G, Vordemvenne T, Scholz M, Kandziora F, et al. Radiographic and safety details of vertebral body stenting: results from a multicenter chart review. BMC Musculoskelet Disord. 2013;14:233.
6. Stagnara P, De Mauroy JC, Dran G, Gonon GP, Costanzo G, Dimnet J, et al. Reciprocal angulation of vertebral bodies in a sagittal plane: approach to references for the evaluation of kyphosis and lordosis. Spine. 1982;7:335–42.
7. Hashimoto T, Kaneda K, Abumi K. Relationship between traumatic spinal canal stenosis and neurologic deficits in thoracolumbar burst fractures. Spine. 1988;13:1268–72.
8. Ley Orgánica 15/1999 de 13 de diciembre de protección de datos de carácter personal. BOE n.° 298, de 14/12/1999 (BOE-A-1999-23750).
9. McAfee PC, Yuan HA, Lasda NA. The unstable burst fracture. Spine. 1982;7:365–73.
10. Cyr L, Francis K. Measures of clinical agreement for nominal and categorical data: the kappa coefficient. Comput Biol Med. 1992;22:239–46.
11. Legendre P. Species associations: the Kendall coefficient of concordance revisited. J Agric Biol Environ Stat. 2005;10:226–45.
12. Lenarz CJ, Place HM, Lenke LG, Alander DH, Oliver D. Comparative reliability of 3 thoracolumbar fracture classification systems. J Spinal Disord Tech. 2009;22:422–7.
13. Sethi MK, Schoenfeld AJ, Bono CM, Harris MB. The evolution of thoracolumbar injury classification systems. Spine J. 2009;9:780–8.
14. Vaccaro AR, Lehman RA Jr, Hurlbert RJ, Anderson PA, Harris M, Hedlund R, et al. A new classification of thoracolumbar injuries: the importance of injury morphology, the integrity of the posterior ligamentous complex and the neurologic status. Spine. 2005;30:2325–33.
15. Vaccaro AR, Baron EM, Sanfilippo J, Jacoby S, Steuve J, Grossman E, et al. Reliability of a novel classification system for thoracolumbar injuries: the thoracolumbar injury severity score. Spine. 2006;31(11S):S62–9.
16. Wood KB, Khanna G, Vaccaro AR, Arnold PM, Harris MB, Mehbod AA. Assessment of two column fracture classification systems used by multiple surgeons. JBJS Am. 2005;87A:1423–9.
17. Elzinga M, Segers M, Siebenga J, Heilbron E, de Lange-de Klerk ES, Bakker F. Inter- and intraobserver agreement on the load sharing classification of thoracolumbar spine fractures. Injury. 2012;43:416–22.
18. Oner FC, Ramos LM, Simmermacher RK, Kingma PT, Diekerhof CH, Dhert WJ, 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.
19. Kuklo TR, Polly DW, Owens BD, Zeidman SM, Chang AS, Klemme WR. Measurement of thoracic and lumbar fracture kyphosis: evaluation of intraobserver, interobserver, and technique variability. Spine. 2001;26:61–5.
20. Dai LY, Jin WJ. Interobserver and intraobserver reliability in the load sharing classification of the assessment of thoracolumbar burst fractures. Spine. 2005;30:354–8.
21. Parizel PM, van der Zijden T, Gaudino S, Spaepen M, Voormolen MH, Venstermans C, et al. Trauma of the spine and spinal cord: imaging strategies. Eur Spine J. 2010;19 Suppl. 1:S8–17.