

临床论著

三种胸腰段骨折分类法的可信度和可重复性研究

吕 游, 张大明, 李 瑞, 张学东, 贾 斌

(北京德尔康尼骨科医院骨三科 100143 北京市)

【摘要】目的:探讨并比较 AO 分型、Denis 分类、TLICS 评分分类系统在胸腰段骨折诊断中的可信度和可重复性。**方法:**选择临床及影像学资料(X 线片、CT、MRI)完整的 31 例胸腰段骨折患者,将该 31 例患者的资料提供给 6 名骨科医生,分别采用 AO 分型、Denis 分类、TLICS 评分分类三种方法进行脊柱骨折分类。3 个月后进行重复分类。采用加权 Cohen's Kappa 系数评价观察者间可信度和观察者内可重复性。**结果:**AO 分型、Denis 分类、TLICS 评分分类的观察者间平均 Kappa 系数分别为 0.517、0.639、0.713;三种分类方法的观察者内平均 Kappa 系数分别为 0.766、0.832、0.804。**结论:**三种胸腰段骨折分类方法比较,TLICS 评分分类方法的可信度和可重复性较高,Denis 分类方法次之,AO 分型方法较差,前者更具临床实用价值。

【关键词】胸腰段骨折;胸腰椎损伤分类及损伤程度评分系统;AO 分型;Denis 分类;一致性检验

doi:10.3969/j.issn.1004-406X.2011.07.09

中图分类号:R683.2 文献标识码:A 文章编号:1004-406X(2011)-07-0566-03

Study of reliability and reproducibility of AO, Denis and TLICS classification systems for thoracolumbar fracture/LÜ You, ZHANG Daming, LI Rui, et al//Chinese Journal of Spine and Spinal Cord, 2011, 21(7):566-568

【Abstract】 Objective: To investigate and compare the reliability and reproducibility of AO, Denis and TLICS classification systems in the diagnosis of thoracolumbar fracture. **Method:** Thirty-one thoracolumbar fracture patients with complete clinical data and radiologic data (including X-ray films, CT and MRI) were enrolled. Six observers, who were orthopaedic surgeons, determined the classifications according to AO, Denis and TLICS systems respectively. The classification was repeated three months later. The Cohen Kappa test was used to determine inter-observer reliability and intra-observer reproducibility. **Result:** The average inter-observer Kappa values for AO, Denis and TLICS Systems were 0.517, 0.639 and 0.713 respectively. And the intra-observer agreements were 0.766, 0.832 and 0.804 respectively. **Conclusion:** Among the three classification systems, TLICS system has the greatest reliability and reproducibility, and thus has more value in clinical practice.

【Key words】 Thoracolumbar fracture; TLICS; AO classification; Denis classification; Consistency test

【Author's address】 The Third Department of Orthopaedics, Beijing DCN Orthopaedic Hospital, Beijing, 100143, China

近年来,随着我国工矿业发展、交通高速化和人口老龄化,脊柱骨折发生率有逐年增加趋势,特别是胸腰段骨折约占脊柱骨折的 80%^[1]。胸腰段骨折因其解剖特点,在诊断、治疗及预后等方面,均具有特殊性,骨科专家一直在寻求更合理实用的分类方法,以便更有效地指导临床治疗。国内目前最常使用的是 AO 分型^[2]和 Denis 分类系统^[3],而美国脊柱损伤研究小组 2005 年提出了有诸多

改进且与前两者不同的胸腰段骨折 TLICS 评分系统^[4]。本研究收集了资料完整的 31 例胸腰段骨折病例,分别让 6 名骨科医生采用上述 3 种分类方法进行分类,研究其观察者间可信度和观察者内可重复性,以此评价其在胸腰段骨折诊断中的临床应用价值。

1 资料与方法

1.1 一般资料

本组患者 31 例,均为本院 2007 年 8 月~2010 年 6 月间因胸腰段骨折入院接受治疗者。其

第一作者简介:男(1981-),主治医师,医学硕士,研究方向:脊柱外科

电话:(010)52713366 E-mail:travelyou@sdu.edu.cn

中男 17 例,女 14 例,年龄 30~85 岁,平均 54.5 岁。致伤原因包括高处坠落伤 7 例,交通伤 11 例,摔伤 13 例。纳入研究的所有患者入院后除体格检查外,均具有胸腰段正侧位 X 线片、伤椎 CT、胸腰段 MRI 检查图像,将专科查体结果与影像资料收集整理编号。

1.2 评分标准

将 AO 分型、Denis 分类、TLICS 评分分类 3 种标准以书面文档形式提供给 6 名临床工作 2 年以内的骨科住院医师。经统一培训后,将已编号的 31 例病例资料隐藏患者姓名及原诊断资料,分发给 6 名医生,分别独立采用 AO 分型、Denis 分类、TLICS 评分分类 3 种方法分别进行分类。3 个月后将 31 例原病例资料再次提供给 6 名医生,再次独立采用 3 种分类方法分别进行分类。收集总结两次分类的结果,通过对 6 名观察者之间的分类情况进行统计比较,评价观察者之间的可信度,通过对同一观察者两次分类情况进行统计比较,确定观察者内的可重复性。

1.3 统计学分析

采用一致性检验,使用 SPSS 软件包(版本号:16.0)计算加权 Cohen's Kappa 系数,评价观察者间和观察者内的相符程度。Kappa 系数的 k 值取值范围在(-1,1)之间,采用 Landis 和 Koch 标准^[5],将一致性强度分为极佳($k > 0.80$)、高度($0.60 < k \leq 0.80$)、中度($0.40 < k \leq 0.60$)、低度($0.2 < k \leq 0.40$)、轻微($0 \leq k \leq 0.20$)和差($k < 0$),分析观察者间和观察者内的一致性。

2 结果

在两次分类中,6 名医生对 31 例胸腰段骨折进行分类,每种分类方法均做出 372 例次的分类。采用 AO 分型系统进行分类中:A 型 275 例次,占 73.93%;B 型 62 例次,占 16.67%;C 型 35 例次,占 9.41%。采用 Denis 分类系统进行分类中:A 型 212 例次,占 56.99%;B 型 103 例次,占 27.69%;C 型 26 例次,占 6.99%;D 型 31 例次,占 8.33%。采用 TLICS 评分系统进行分类中:TLICS 评分 < 4 分者 182 例次,占 48.93%;TLICS 评分=4 分者 66 例次,占 17.74%;TLICS 评分 > 4 分者 124 例次,占 33.33%。

2.1 观察者间可信度

AO 分型 3 种类型层面的平均 Kappa 系数为

0.517,相符性属中度;Denis 分类 4 种类型层面的平均 Kappa 系数为 0.639,相符性属高度;TLICS 评分系统 3 种类型层面的平均 Kappa 系数为 0.713,相符性属高度。进一步对 TLICS 评分系统各亚类(骨折形态、神经状态、后方韧带复合体状态)进行一致性检验,各亚型 Kappa 系数均位于中度以上,其中神经状态 Kappa 系数最高为 0.901,后方韧带复合体状态 Kappa 系数较低为 0.497(表 1)。

2.2 观察者内可重复性

AO 分型观察者内平均 Kappa 系数为 0.766,相符性属高度;Denis 分类系统观察者内平均 Kappa(k)指数为 0.832,相符性属极佳;TLICS 评分系统观察者内平均 Kappa 系数为 0.804,相符性亦属极佳(表 1)。

表 1 AO 分型、Denis 分类、TLICS 评分分类一致性检验结果

	观察者间一致性检验 Kappa 系数	观察者内一致性检验 Kappa 系数
AO 分型系统	0.517	0.766
Denis 分类系统	0.639	0.832
TLICS 评分系统	0.713	0.804
骨折形态	0.652	
神经状态	0.901	
后方韧带复合体状态	0.497	

3 讨论

胸腰段骨折因其解剖位置和生理意义的特殊性,在临床诊断与治疗上存在诸多问题,相关的学术争议也较多。从 1949 年提出的 Nicoll 胸腰椎骨折分类,到 1963 年由 Holdsworth 提出双柱概念,再到 1983 年 Denis 的脊柱三柱理论,关于胸腰段骨折的分类研究一直在争论中发展。目前国内临床较多采用的是 AO 分型^[2]和 Denis 分类系统^[3],而国外则较多采用 Gertzbein 分类或 AO 分型系统。2005 年,以 Vaccaro 为首的美国脊柱损伤研究小组制定出一套针对胸腰段脊柱脊髓损伤程度的评分系统,即胸腰椎损伤分类及损伤程度评分系统(TLICS)^[4],也被用于胸腰段骨折的诊断分类。以上众多分类方法各有优劣,但尚无一种方法能够得到广泛接受和应用。

本研究采用一致性检验,对 3 种分类方法观察者间的可信度和观察者内的可重复性进行比较研究。结果发现,TLICS 评分系统的观察者间可信度优于 AO 分型和 Denis 分类系统,这可能与 AO

分型和 Denis 分类的固有缺陷有关。根据本研究实践认为:AO 分型分 3 大类 9 组 27 型,分型方法繁琐,对初学者难以掌握,而且其中有些分型对指导骨折治疗并无实际价值;Denis 分类分 4 类,其中 A 类、B 类根据病理形态分类,C 类是特定称谓,D 类根据损伤机制分类,在临床应用中难免发生交叉混淆。TLICS 评分虽然涉及方面众多,但经过评分加分后,将最终结果以 4 分为界分为 3 类,能够大大提高评分的可信程度。孙天胜等^[6]对 TLICS 评分系统的可信度进行研究,结果同样发现,TLICS 总分从 1~10 分不等,在此层面观察者间可信度 Kappa 系数较低,仅为 0.48;而推荐治疗后,即以 4 分为界分为 3 个分类,此层面 Kappa 系数提高至 0.73。另外,Denis 分类的观察者间可信度也高于 AO 分型。观察者内 Denis 分类和 TLICS 评分系统均具有较高的可重复性,明显高于 AO 分型,说明 AO 分型在临床应用中复杂繁琐,可信度较低。这与 Oner 等^[7]的研究结果一致,后者证实 AO 分型的可信度和可重复性均明显低于 Denis 分类。

在 TLICS 评分系统中,各亚类 Kappa 系数差异较大。其中,骨折形态的 Kappa 系数与 Denis 分类系统较接近;神经功能状态的 Kappa 系数较高,可信度达到极佳;而后方韧带复合体的 Kappa 系数较低,可信强度仅为中度。TLICS 评分系统中,骨折形态分类方法与 Denis 分类较接近,均分为 4 个亚类,但不含特定称谓,并将旋转机制与牵张机制区分为两类,更易于临床评估判断。神经状态 Kappa 系数较高,此结果说明其临床判断方法简明清晰,易于临床应用。后方韧带复合体 Kappa 系数较低,这可能与后方韧带复合体定义不清、诊断难度较大有关。后方韧带复合体(PLC)与 TLICS 评分系统同时提出,其定义包含后纵韧带、棘上韧带、棘间韧带、黄韧带关节囊,其诊断需借助临床查体、X 线片、CT、MRI 等综合考虑,虽然 MRI 对其诊断敏感性较高^[8],但也存在由于血肿形成和黄韧带受压所导致的假阳性诊断^[9]。但不能否认,相比 AO 分型、Denis 分类单纯以 X 线和 CT 为诊断基础,TLICS 评分系统纳入了更先进的影像学诊断技术,使评估更符合骨折损伤的实际情况。

理想的分类方法应该具备以下特点:全面性、简易性、高信度以及对临床治疗和预后的指导性。

用于诊断胸腰段骨折的 AO 分型和 Denis 分类已经在临床得到了较广泛的应用,其临床应用价值也得到了学术界的公认。TLICS 评分系统在此基础上,能够更全面准确评估胸腰段骨折,并采用评分分类方法判定骨折损伤的严重程度,该方法不仅较容易判定,且具有更好的临床实用性。本研究着眼于 3 种分类系统的比较研究,虽证实 TLICS 评分系统具有较高的可信度和可重复性。然而受病例资料完整性的限制,导致本研究样本量较小,随访时间较短,有待于更大样本量、多中心和更长时间的随访研究,才能对这一评分系统作出更科学的评价。

4 参考文献

1. Verlaan JJ, Diekerhof CH, Buskens E, et al. Surgical treatment of traumatic fractures of the thoracic and lumbar spine: a systematic review of the literature of techniques complications and outcome[J]. Spine, 2004, 29(7): 803-814.
2. Magerl F, Aebi M, Gertzbein SD, et al. A comprehensive classification of thoracic and lumbar injuries[J]. Eur Spine J, 1994, 3(4): 184-201.
3. Denis F. The three column spine and its significance in the classification of acute thoracolumbar spinal injuries[J]. Spine, 1983, 8(8): 817-831.
4. Vaccaro AR, Lehman RA, Hurlbert J, et al. A new classification of thoracolumbar injuries: the importance of injury morphology, the integrity of the posterior ligamentous complex, and neurologic status[J]. Spine, 2005, 30(20): 2325-2333.
5. Landis JR, Koch GG. The measurement of observer agreement for categorical data[J]. Biometrics, 1977, 33(1): 159-174.
6. 孙天胜, 张志成. 胸腰椎损伤分类及损伤程度评分系统的评估及初步应用[J]. 脊柱外科杂志, 2007, 5(6): 325-329.
7. 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 [J]. Eur Spine J, 2002, 11(3): 235-245.
8. Lee HM, Kim HS, Kim DJ, et al. Reliability of magnetic resonance imaging in detecting posterior ligament complex injury in thoracolumbar spinal fractures[J]. Spine, 2000, 25(16): 2079-2084.
9. 徐建桥, 陈维善, 陈其昕, 等. MRI 诊断胸腰椎骨折后柱韧带复合体损伤的可信度研究[J]. 中华骨科杂志, 2006, 26(9): 602-605.

(收稿日期: 2011-01-26 修回日期: 2011-03-03)

(英文编审 赵景鑫/贾丹彤)

(本文编辑 彭向峰)