An Original Study

Measurement of Thoracolumbar Kyphosis After Burst Fracture: Evaluation of Intraobserver, Interobserver, and Variability of 4 Measurement Methods

Jerome G. Enad, MD, Joseph B. Slakey, MD, and Patrick S. McNulty, MD

Abstract

There are various methods for measuring kyphosis after thoracolumbar burst fracture. The reliability and reproducibility of these methods are not well defined. In the study reported here, we examined 4 commonly used measurement methods in order to determine intraobserver variability, interobserver variability, and variability between measurement methods. All 4 methods were found to be accurate and reproducible when used by 4 observers on 2 occasions. One method, in comparison with the others, tended to overestimate degree of kyphosis. Understanding the methods for measuring kyphotic deformity after thoracolumbar burst fracture is essential in making decisions about prognosis and treatment.

Management of thoracolumbar burst fractures is influenced by both clinical and radiographic findings. Surgical treatment consisting of open reduction, fixation, and arthrodesis confers the advantage of deformity correction, early mobilization, and reduced reliance on orthotic bracing and is considered in the event of neurologic deficit or when the initial sagittal deformity is large or progresses during the course of nonoperative treatment.[1-4] However, several studies have also demonstrated that severity of initial or residual kyphotic deformity does not correlate with symptoms at follow-up, thereby bringing into question the efficacy of radiographic indications for surgery.[5-9]

Additionally, the numerous methods described for measuring thoracolumbar kyphosis after burst fracture suggest that no single method is widely accepted.[5,6,10-20] Kuklo and colleagues[16] examined 5 different methods for measuring thoracolumbar fracture kyphosis and found 1 that exhibited significantly higher interobserver and intraobserver reliability. This method measured from the superior endplate of the vertebral body 1 level above the injured vertebra to the inferior endplate of the vertebral body 1 level below. In that study, however, the absolute values obtained with each individual method were not compared with the values obtained with the other methods. Thus, though a single method may produce the most reliable and repeatable result, it should also be known whether that result may be significantly different from that obtained with other methods or whether each method that is proved to be accurate yields similar measurement values. If decision making depends in part on the absolute or relative value of the measured kyphosis, it is important to recognize whether a single method, in comparison with other methods, consistently results in a significantly different measured value.

In the study reported here, we evaluated 4 commonly used methods in order to determine intraobserver variability, interobserver variability, and variability between measurement methods.

Materials and Methods

Sixteen lateral spine standard cassette x-rays of single-level thoracolumbar burst fractures (range, T10–L2) treated at our institution were selected from the radiology department files. X-rays were selected for clarity, adequate rotation, and centering of the vertebral level on the film. Four high-quality hard copies of the x-rays were made and numbered by an independent assistant who did not perform the measurements. Four board-certified orthopedic surgeons (2 orthopedic spine surgeons, 2 pediatric orthopedic surgeons trained in scoliosis surgery) were identified as observers. The individual observers were blinded to patient identity, history, diagnosis, and treatment. Each examiner independently performed 4 measurements on a copy of each x-ray using a radiographic marking pencil and a standard Cobb ruler (United States Manufacturing Co, Pasadena, Calif). All radiographic markings and lines were completely erased after each measurement was recorded. The films were renumbered and measurements retaken 3 weeks after the first iteration.

The observers were instructed on the 4 methods most commonly used at our institution: (1) measuring from

CDR Jerome G. Enad, MC, USN, and CAPT Joseph B. Slakey, MC, USN, are Assistant Professors of Surgery, Uniformed Services University of the Health Sciences, Bethesda, Maryland.
Dr. McNulty is Staff Spine Surgeon, Nevada Orthopedic and Spine Center, Las Vegas, Nevada.

Address correspondence to: Jerome G. Enad, MD, 620 John Paul Jones Circle, Portsmouth, VA 23708 (tel, 757-953-1884; fax, 757-953-0362; e-mail, jerome.enad@med.navy.mil).

Am J Orthop. 2008;37(4):E60-E63.

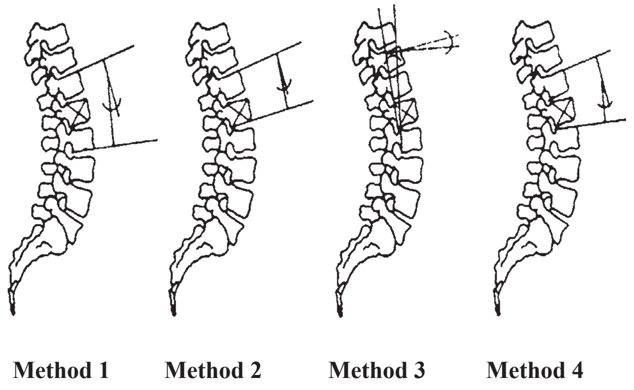

Figure. Four methods used for fracture kyphosis measurement: (1) measuring from the superior endplate of the vertebral body 1 level above the injured vertebral body to the inferior endplate of the vertebral body 1 level below, (2) measuring from the superior endplate of the vertebral body 1 level above the injured vertebral body to the inferior endplate of the injured vertebral body, (3) measuring the angle between the posterior vertebral body 1 level above the injured vertebral body and 1 level below the injured vertebral body, (4) measuring from the superior endplate 1 level above the injured vertebral body and the superior endplate 1 level below the injured vertebral body. Illustration by Jerome G. Enad, MD.

the superior endplate of the vertebral body 1 level above the injured vertebral body to the inferior endplate of the vertebral body 1 level below, (2) measuring from the superior endplate of the vertebral body 1 level above the injured vertebral body to the inferior endplate of the injured vertebral body, (3) measuring the angle between the posterior vertebral body 1 level above and below the injured vertebra, and (4) measuring from the superior endplate of the vertebral body 1 level above the injured vertebral body to the superior endplate of the vertebral body 1 level below (Figure).

Statistical Analysis

Data were analyzed with a statistical software program, SPSS 7.5 for Windows (SPSS Inc, Chicago, Ill). Analyses of variance (ANOVAs) for repeated measures were performed to calculate the intraclass correlation coefficient, ρ (rho), for estimating the intraobserver and interobserver reliability as described by Winer.[21] The intraobserver reliability assessed the reproducibility of each observer for each measurement method. In this study, each observer measured the same x-ray twice for each method. The interobserver reliabilities were obtained to assess overall agreement among the 4 observers for all methods and for each method. For analyzing interobserver reliability, the first measurement of each observer was entered into the ANOVA. The ANOVA generalized linear model for repeated measures was performed to determine the effect of the different methods on the intraobserver and interobserver reliability. The Duncan multiple range test was used to compare the mean values obtained from each method to determine variability between methods. (The test controls only for the type 1 comparison-wise error rate, not the experiment-wise error rate). Further, a post hoc power analysis for a sample size of 16 subjects (ie, x-rays) was performed to estimate the power of the difference in mean values obtained between methods.

Results

Intraobserver Reliability

Reproducibility for each observer was high in comparison of each of the 4 methods (Table I). The intraclass coefficient (ρ) varied from .979 to .996 for observer 1, from .874 to .939 for observer 2, from .978 to .992 for observer 3, and from .987 to .991 for observer 4. The intraclass coefficients were most consistent for method 1 (mean ρ, .978), measuring from the superior endplate of the vertebral body 1 level above the injured vertebral body to the inferior endplate of the vertebral body 1 level below. This was followed by method 4 (mean ρ, .978), measuring from the superior endplate of the vertebral body 1 level above the injured vertebral body to the superior endplate of the vertebral body 1 level below. Next was method 3 (mean ρ, .959), measuring the angle between the posterior vertebral body 1 level above and below the injured vertebra. Method 2, measuring from the superior endplate of the vertebral body 1 level above the injured vertebral body to the inferior endplate of the injured vertebral body, produced the lowest intraclass coefficients overall (mean ρ, .954).

Interobserver Reliability

Paired comparisons between observers had low variability, according to ρ intraclass correlation coefficients for each measurement method (Table II). Method 1 (mean ρ, .962) had the best interobserver reliability. Next were method 3 (mean ρ, .948) and method 4 (mean ρ, .945), followed by method 2 (mean ρ, .898). All methods resulted in excellent reliability (ρ>.80).[21]

Measurement Differences Between Methods

According to the Duncan multiple range test for variability ($\alpha = .05$), mean values for each x-ray using methods 1, 3, and 4 were not significantly different from each other. The mean value for each x-ray using method 2 was significantly different from ($P<.001$) and higher in absolute value than that using method 1, 3, or 4 (mean difference, range +8.0° to +9.8°). A post hoc power analysis, performed to detect a large effect difference between values from method 2 and values from methods 1, 3, and 4, yielded 91% power at a .05 significance level.

Discussion

The first goal of this study was to determine the best measurement method in terms of intraobserver reproducibility and interobserver reliability for thoracolumbar burst fracture. The second goal was to detect any discernible difference in values obtained from using each of the methods. Our data suggest that each of the 4 methods used to measure posttraumatic kyphosis in thoracolumbar burst fracture is reproduc-

Table I. Intraclass Correlation Coefficients (ρ) for Intraobserver Reliability for Each Observer Using Each of the 4 Measurement Methods

Observer	Method 1	Method 2	Method 3	Method 4
1	.996	.989	.979	.993
2	.931	.874	.938	.939
3	.992	.978	.931	.985
4	.991	.976	.987	.979

Table II. Paired Interobserver Reliability Calculated Using Intraclass Correlation Coefficients (ρ) Given a Specified Measurement Method

Observers	Method 1	Method 2	Method 3	Method 4
1 & 2	.935	.815	.932	.929
1 & 3	.995	.956	.973	.977
1 & 4	.989	.979	.974	.962
2 & 3	.939	.857	.948	.938
2 & 4	.926	.817	.903	.901
3 & 4	.991	.962	.957	.962

ible and reliable. However, a significant difference was found between the values obtained with method 2 and each of the other methods used in this study. A mean difference of +8.0° to +9.8° existed for method 2 versus the other methods. These findings are important in considerations of fracture treatment and long-term stability. Understanding the methods for measuring kyphotic deformity is essential in making decisions about prognosis and treatment.

In a similar study, examining 5 methods for measuring fracture kyphosis, Kuklo and colleagues[16] noted that 1 method had superior intraobserver and interobserver reliability. This method—measuring from the superior endplate of the vertebral body 1 level above the injured vertebral body to the inferior endplate of the vertebral body 1 level below—was identical to method 1 described in our study. In our study, we also found that this method had the most consistent intraobserver correlation coefficient and the best paired-comparison interobserver reliability of all methods examined.

Our study has some potential sources of variability. A standard radiographic method was not used to obtain images of the vertebral burst fracture. Of note, these injury films were each taken in the supine position and do not account for the load-bearing effects on the spine during upright posture. We did not examine fracture type based on the Denis classification of burst fractures.[22] Quality of duplicated x-rays may have distorted interpretation of endplate and posterior body landmarks. Our selection criteria allowed us to minimize these technical concerns but left us with only 16 eligible x-rays. This sample size was calculated by our statistician to be sufficient to yield a power of more than 90% in determining reliability and reproducibility of each measurement method and the differences between methods.

Conclusions

Invariably, the amount of posttraumatic kyphotic deformity is only one factor in determining overall treatment of thoracolumbar burst fracture. If this criterion is to be considered, however, the method for measuring the angle of kyphosis must be reproducible, accurate, and well defined. Our study demonstrated that each of the 4 methods described is reliable and reproducible. Method 1, measuring from the superior endplate of the vertebral body 1 level above the injured vertebral body to the inferior endplate of the vertebral body 1 level below, is the most consistent in terms of intraobserver and interobserver reliability, and its continued use is recommended. Method 2, measuring from the superior endplate of the vertebral body 1 level above the injured vertebral body to the inferior endplate of the injured vertebral body, has the lowest intraobserver and interobserver reliability. Method 2 also produces consistently higher values in comparison with the other methods, and its continued use is discouraged.

Some authors have suggested that the initial amount of fracture kyphosis is related to prognosis,[1-4,23] but there is no clear association between kyphotic deformity and reported pain at follow-up.[5-9] However, accurately measuring the angle of kyphosis remains important both for documentation of clinical assessment and for examination of parameters related to treatment outcomes. Therefore, one should strive to use a measurement method that is reliable and reproducible.

Authors' Disclosure Statement and Acknowledgments

The authors report no actual or potential conflict of interest in relation to this article.

The views expressed in this article are those of the authors and do not reflect the official policy or position of the US Department of the Navy, the US Department of Defense, or the US government.

The authors acknowledge the contributions of Dr. Jeffrey Knapp and Dr. Anthony Lapinsky as participating investigators in data collection.

References

1. Bradford DS, Akbarnia BA, Winter RB, Seljeskog EL. Surgical stabilization of fracture and fracture dislocations of the thoracic spine. *Spine*. 1977;2(3):185-196.
2. McAfee PC, Yuan HA, Ladda NA. The unstable burst fracture. *Spine*. 1982;7(4):365-373.
3. Krompinger WJ, Frederickson BE, Mino DE, Yuan HA. Conservative treatment of fractures of the thoracic and lumbar spine. *Orthop Clin North Am*. 1986;17(1):161-170.
4. Weidenbaum M, Farcy JPC. Surgical management of thoracic and lumbar burst fractures. In: Bridwell KH, Gordon RL, eds. *The Textbook of Spinal Surgery*. 2nd ed. New York, NY: Lippincott-Raven; 1997:1851.
5. Weinstein JN, Collalto P, Lehmann TR. Thoracolumbar "burst" fractures treated conservatively: a long-term follow-up. *Spine*. 1988;13(1):33-38.
6. Mumford J, Weinstein JN, Spratt KF, Goel BK. Thoracolumbar burst fractures. The clinical efficacy and outcome of nonoperative management. *Spine*. 1993;18(8):955-970.
7. Chow GH, Nelson BJ, Gebhard JS, Brugman JL, Brown CW, Donaldson DH. Functional outcome of thoracolumbar burst fractures managed with hyperextension casting or bracing and early mobilization. *Spine*. 1996;21(18):2170-2175.
8. Kraemer WJ. Schemitsch EH, Lever J, McBroom RJ, McKee MD, Waddell JP. Functional outcome of thoracolumbar burst fracture without neurological deficit. *J Orthop Trauma*. 1996;10(8):541-544.
9. Wood K, Buttermann G, Mehbod A, Garvey T, Jhanjee R, Sechriest V. Operative compared with non-operative treatment of a thoracolumbar burst fracture without neurological deficit. *J Bone Joint Surg Am*. 2003;85(5):773-781.
10. Akbarnia BA, Crandall DG, Burkus K, Matthews T. Use of long rods and a short arthrodesis for burst fractures of the thoracolumbar spine. *J Bone Joint Surg Am*. 1994;76(11):1629-1635.
11. Atlas SW, Regenbogen V, Rogers LF, Kim KS. The radiographic characterization of burst fractures of the spine. *AJR Am J Roentgenol.* 1986;147(3):575-582.
12. Chan DPK, Seng NK, Kaan KT. Nonoperative treatment in burst fractures of the lumbar spine (L2–L5) without neurologic deficits. *Spine*. 1993;18(3):320-325.
13. Davies WE, Morris JH, Hill V. An analysis of conservative (non-surgical) management of thoracolumbar fractures and fracture-dislocations with neural damage. *J Bone Joint Surg Am*. 1980;62(8):1324-1328.
14. Hardaker WT Jr, Cook WA, Friedman AH, Fitch RD. Bilateral transpedicular decompression and Harrington rod stabilization in the management of severe thoracolumbar burst fractures. *Spine*. 1992;17(2):162-171.
15. Kaneda K, Abumi K, Fujiya M. Burst fractures with neurologic deficits of the thoracolumbar-lumbar spine. *Spine*. 1984;9(8):788-795.
16. Kuklo TR, Polly DW Jr, 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(1):61-66.
17. McNamara MJ, Stephens GC, Spengler DM. Transpedicular short-segment fusions for treatment of lumbar burst fractures. *J Spinal Disord*. 1992;5(2):183-187.
18. Okuyama K, Abe E, Chiba M, Ishikawa N, Sato K. Outcome of anterior decompression and stabilization for thoracolumbar unstable burst fractures in the absence of neurologic deficits. *Spine*. 1996;21(5):620-625.
19. Stephens GC, Devito DP, McNamara MJ. Segmental fixation of lumbar burst fractures with Cotrel-Dubousset instrumentation. *J Spinal Disord*. 1992;5(3):344-348.
20. Willen J, Lindahl S, Nordwall A. Unstable thoracolumbar fractures. A comparative clinical study of conservative treatment and Harrington instrumentation. *Spine*. 1985;10(2):111-122.
21. Winer BJ. *Statistical Principles in Experimental Design*. 2nd ed. New York, NY: McGraw-Hill; 1971.
22. Denis F. Spinal instability as defined by the three-column spine concept in acute spinal trauma. *Clin Orthop*. 1984;(189):65-76.
23. Gertzbein SD. Scoliosis Research Society. Multicenter spine fracture study. *Spine*. 1992;17(5):528-539.