The Spine Journal 19 (2019) 1324–1330

Clinical Study

Reliability of radiological measurements of type 2 odontoid fracture

Brian A. Karamian, MD, Ning Liu, MD, MPH, Remi M. Ajiboye, MD, Ivan Cheng, MD, Serena S. Hu, MD, Kirkham B. Wood, MD*

Department of Orthopaedic Surgery, Stanford University School of Medicine, 450 Broadway St, Redwood City, CA 94063, USA

Received 2 January 2019; revised 29 April 2019; accepted 30 April 2019

Abstract

BACKGROUND CONTEXT: It is recognized that radiological parameters of type 2 dens fractures, including displacement and angulation, are predictive of treatment outcomes and are used to guide surgical decision-making. The reproducibility of such measurements, therefore, is of critical importance. Past literature has shown poor interobserver reliability for both displacement and angulation measurements of type 2 dens fractures. Since such studies however, various advancements of radiological review systems and measurement tools have evolved to potentially improve such measurements.

PURPOSE: To re-examine the inter-rater reliability of measuring displacement and angulation of type 2 dens fractures using modern radiological review systems. Besides quantitative measurements, the reliability of raters in identifying diagnostic classifications based on translational and angulational displacement was also examined.

STUDY DESIGN: Radiographic measurement reliability and agreement study.

PATIENT SAMPLE: Thirty-seven patients seen at a single institution between 2002 and 2017 with primary diagnosis of acute type 2 dens fracture with complete computed tomography (CT) imaging.

OUTCOME MEASURES: Radiological measurements included displacement and angulation. Diagnostic classifications based on consensus-based clinical cutoff points were also recorded.

METHODS: Measurements were performed by five surgeons with varying years of experience in spine surgery using the hospital's electronic medical record radiological measuring tools. The radiological measurements included displacement and angulation. Diagnostic classifications based on consensus-based clinical cutoff points were also recorded. Each rater received a graphic demonstration of the measurement methods, but had the autonomy to select a best cut from the sagittal CT to measure. All raters were blinded to patient information.

RESULTS: Measurements for displacement and angulation among the five raters demonstrated "excellent" reliability. Intra-rater reliability was also "excellent" in measuring displacement and angulation. The reliability of diagnostic classification of displacement (above vs. below 5 mm), was found to be "very good" among the raters. The reliability of diagnostic classification of angulation (above vs. below 11°) demonstrated "good" reliability.

CONCLUSIONS: Advancement of radiological review systems, including review tools and embedded image processing software, has facilitated more reliable measurements for type 2 odontoid fractures.

Keywords: Dens fracture; Measurement; Observer reliability; Radiographic classification; Type 2 odontoid fracture; Computed tomography

FDA device/drug status: Not applicable.

Author disclosures: ***BAK:*** Nothing to disclose. ***NL:*** Nothing to disclose. ***RMA:*** Nothing to disclose. ***IC:*** Royalties: Nuvasive (F), Spine Wave (B), SpineCraft (A); Stock Ownership: Nuvasive (B); Private Investments: Cytonics (D), Spine Innovations (D), SpinalCyte (D); Consulting: Stryker Spine (B), Nuvasive (D); Scientific Advisory Board/Other Office: Notogen. ***SSH:*** Royalties: OnPoint (A), Nuvasive (G); Trips/Travel (B); American Orthopaedic Association: Board or committee member; Global Spine Journal: Editorial or governing board; Scoliosis Research Society: Board or committee member. ***KBW:*** Royalties: Globus (B); Consulting: Life Spine (B), Alphatec Spine (B), K2M (B).

* Corresponding author. Department of Orthopaedic Surgery, Stanford University School of Medicine, 450 Broadway St MC 6342, Redwood City, CA 94063, USA. Tel.: (650)721-7620, fax: (650)721-3470.

E-mail address: kbwood@stanford.edu (K.B. Wood).

https://doi.org/10.1016/j.spinee.2019.04.020

Introduction

The recognized incidence of odontoid (dens) fractures has been increasing caused by both societal aging and improvement of diagnostic techniques [1,2]. Type 2 odontoid fractures, where the fracture takes place in the junctional region between the dens and the C2 vertebral body, are clinically challenging as nonunion rates following nonoperative management range from 24% and 88% [3,4]. Although few standardized strategies exist guiding its management, it is commonly recognized that the radiological parameters of type 2 dens fractures, including displacement and angulation, are important predictors of treatment outcomes and are necessary to guide surgical decision-making [3,5].

Given the aforementioned clinical relevance, the reliability of radiological measurements for dens fractures is a subject of obvious importance. In 2010, Bono et al. [6] found poor inter-rater reliability for both displacement and angulation measurements for type 2 dens fractures even when experienced spine surgeons were given the "best case" scenario with a provided lateral cervical radiograph, single sagittal computed tomography (CT) cut, and instructional diagram. However, as the authors indicated, the study conditions may have been somewhat strict as the reliability was evaluated solely by comparing quantitative measurements (eg, distance in millimeters) as continuous data and did not analyze the reliability of classifying odontoid fractures using clinical cutoff values for angulation or displacement. In addition, the raters were also required to measure from both radiographs and CT sagittal images with calibration methods used to standardize measurements between the two modalities that may have impacted measurements.

The interpretation of any measurement study is subject to the accuracy and facility of its measurement tool. Since Bono et al.'s work, various new advanced hospital radiological systems for both capturing and reviewing have been adopted by many facilities nationwide. Software capabilities including artifact reduction and interface features such as zoom, window, and measurement readjustment aid in the accuracy of measurement and have allowed users to become more facile with measuring. Improvements of CT scanners and advancements in imaging processing software, such as automated exposure control, dual energy, and iterative reconstruction, have led to more precise characterization and quantification of connective tissue architecture.

Accordingly, the current study aims to re-examine the reliability of measuring the displacement and angulation of type 2 dens fractures using a modern radiological review system that incorporates the abovementioned technological improvements. In addition to the continuous quantitative measurements, the reliability in identifying diagnostic classifications (which are based on neurosurgical and orthopedic spine consensus for increased risk of nonunion: greater than 5 mm for displacement and 11° for angulation) [3,7–9] regarding translational and angulational stability was also examined. Finally, raters with varying years of experience in spinal surgery participated to assess the influence of practice years as well as to include the range of surgeons who would likely be involved in assessing actual clinical scenarios.

Materials and methods

After IRB approval, the cases for measurement were obtained from patients seen at our hospital between 2002 and 2017 through a medical record searching for "type II dens fracture" in primary diagnosis or the Current Procedural Terminology codes (22318, 22319, and 22548) that suggest relevant surgeries. This search strategy harvested a total of 91 cases. After excluding irrelevant or duplicate cases, cases with congenital deformity, malignancy affecting cortical bone, and cases with incomplete images, 37 patients diagnosed with acute type 2 odontoid fractures with complete CT images were enrolled. Computed tomography scans were obtained either from at outside facility or our institution. Computed tomography scans obtained at our facility were performed on LightSpeed, Discovery, or Revolution (General Electric, Boston, MA, USA) as well as Somatom (Siemens, Munich, Germany) CT scanners.

The radiological measurements included displacement and angulation defined according to the consensus of the Spine Trauma Study Group as shown in Fig. 1 [10]. For displacement and angulation, a diagnostic classification (above vs. below 5 mm and above vs. below 11°, respectively) based on a consensus-based clinical cutoff points was also recorded [3,7–9]. The measurements were performed using the hospital's web-based picture archiving and communication system, Centricity (General Electric, Boston, MA, USA), integrated into the medical record system. In this system, different slices of the cervical spine sagittal and axial CT can be scanned fluently for sequential review, and the images can be magnified as needed with the embedded calibration scale readapted automatically.

Five surgeons with varying years of experience in spine surgery measured the full list of the 37 patients separately. They were affiliated with the same academic spine center and include two attending surgeons [Attending (A) and Attending (B)], one spine fellow, one orthopedic resident, and a researcher with spine surgeon training. Each rater received a graphic demonstration and text guideline with a detailed description of the measurement methods (Fig. 1), but had the autonomy to select a best slice from the sagittal CT of each patient to measure [10]. All raters were blinded to other patient information. Two of the raters, attending and researcher, randomly selected 15 cases from the initial 37 cases, and their intra-rater reliability was measured.

The radiological measurements of odontoid fractures included both quantitative measurements of displacement and angulation, as well as qualitative evaluation of displacement and angulation based on clinical cutoff values. Intraclass correlation coefficient (ICC) was used to evaluate

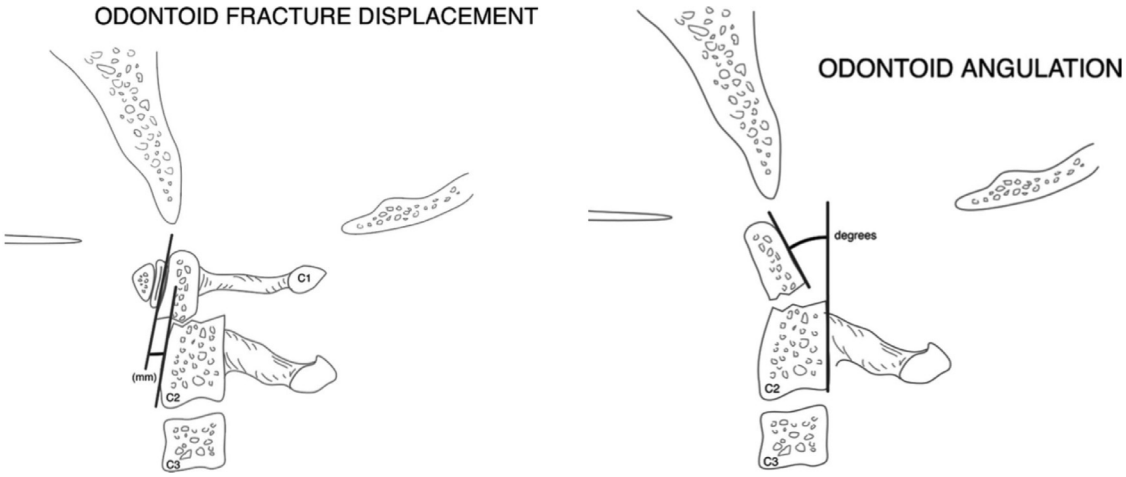

Fig. 1. Graphic demonstration of measurement technique provided to raters for both displacement and angulation of type 2 odontoid fractures. Copyright purchased. Bono et al. [10].

the inter-rater reliability when measuring displacement and angulation as continuous data. An ICC score greater than 0.75 indicates "excellent" reliability, between 0.4 and 0.75 indicates "good" reliability, and below 0.4 indicates "poor" reliability based on recommendations from Shrout and Fleiss [11]. Given its categorical nature, Kappa statistic was used for the reliability of the diagnostic classification of displacement (above vs. below 5 mm). Kappa statistic was interpreted with a scale taken from Altman, with greater than 0.8 indicating "very good" and between 0.6 and 0.8 indicating "good" reliability [12].

The sample size of patients needed was estimated using methodology previously described by Donner and Rotondi to ensure that the obtained Kappa value had an acceptable degree of precision [13]. Taking the Kappa value for displacement classification among raters as an example, we hypothesized that the observer reliability regarding classification would be very good (Kappa=0.8) with preferred precision level of Kappa value being 0.2 (meaning the estimated Kappa value would not differ by more than 0.2 from the "true" Kappa value). Thus, with five raters and a 30% evidence-based proportion of significant displacement (≥5 mm) in type 2 dens fracture [3], the minimal sample size needed was 22 patients to achieve an acceptable degree of precision. R statistical software version 3.3.2 (R Foundation for Statistical Computing) was used for analysis. The α value was set at 0.05.

Results

All five raters completed the measurements for the 37 patients, 59.5% (22 of 37) of whom were men, with an average age at diagnosis of 69 (19−96) years. Of the 37 cases examined, 33 cases were acute odontoid fractures with median time from injury to CT imaging of 0 (0−3) days. Four cases were chronic fractures with average time from injury to CT imaging of 83 (58−120) days. Thirty-two percent of cases had their imaging obtained at an outside facility. Patient demographics and mechanism of injury are shown in Table 1. A descriptive summary of each rater's measurement results is shown in Table 2.

The ICCs of the measurements for displacement and angulation of the five raters was 0.940 (95% confidence interval [CI]: 0.899−0.963) and 0.849 (95% CI: 0.773−0.91), respectively, both indicating "excellent" reliability [6]. To visualize the underlying measurement data of raters, the measurements for displacement and angulation were also plotted pairwise in Figs. 2 and 3, respectively. These show similar patterns of pairwise reliability among the raters, which is in keeping with the ICCs.

For the reliability of diagnostic classification of displacement (above vs. below 5mm), the agreement percentages among raters are listed in Table 3. The corresponding Kappa value is 0.782 (95% CI: 0.73−0.95, p<.001), indicating "very good" reliability among the five raters.

The reliability of diagnostic classification of angulation (above vs. below 11°) is somewhat less than that for displacement, although the agreement percentages (Table 4) and the corresponding Kappa value 0.70 (95% CI: 0.58−0.85, p<.001) still indicate "good" reliability among the five raters.

The ICCs of intra-rater reliability for both displacement and angulation showed excellent reliability in measuring displacement and angulation with values 0.922 (95% CI: 0.791−0.973) and 0.901 (95% CI: 0.738−0.965), respectively, for attending remeasurements and 0.988 (95% CI: 0.967−0.996) and 0.907 (95% CI: 0.752−0.967), respectively, for researcher remeasurements.

Discussion

Type 2 dens fractures represent a broad anatomic spectrum of injury. Radiographic parameters of type 2 dens fractures, including displacement and angulation, are

Table 1
Patient demographics and injury mechanism

Patient	Age	Gender	Days from injury to image	Injury mechanism	CT location
1	85	M	0	Fall	Internal
2	19	M	0	MVC	Internal
3	30	F	0	MVC	External
4	21	M	1	MVC	Internal
5	24	M	0	MVC	Internal
6	57	M	0	Fall (ladder)	Internal
7	81	M	1	Fall (horse)	Internal
8	82	M	0	Fall	Internal
9	89	F	3	Fall	Internal
10	74	M	0	Bicycle crash	Internal
11	90	M	0	Fall	Internal
12	32	M	0	MVC	External
13	57	M	58	Neoplasm	Internal
14	80	M	120	Fall	External
15	87	F	0	Fall	Internal
16	23	F	0	MVC	Internal
17	77	M	0	MVC	External
18	25	M	96	MVC	Internal
19	61	F	0	Fall	External
20	69	F	59	Fall	External
21	77	M	0	Fall	External
22	59	M	1	MVC	External
23	76	M	0	Ped vs. Auto	External
24	70	F	0	Fall	Internal
25	89	F	1	Fall	External
26	77	F	0	Fall	Internal
27	96	M	0	Fall	External
28	64	F	0	Fall	Internal
29	85	M	2	Fall	Internal
30	87	F	0	Fall	Internal
31	83	F	1	Fall	Internal
32	75	M	0	Fall	Internal
33	90	F	0	Fall	Internal
34	93	M	0	Fall	External
35	96	M	0	Fall (horse)	Internal
36	89	F	0	Fall	Internal
37	72	F	0	Fall	Internal

Table details patient age (years) at the time of CT imaging, gender (M=male, F=female), time in days from injury to obtaining CT scan, mechanism of injury (Fall=ground level fall unless otherwise noted, MVC=motor vehicle collision, Ped vs. Auto=pedestrian vs. automobile), and whether imaging was obtained at our institution or an outside facility (Internal=our institution, External=outside institution).

important to guide surgical management and are predictive of treatment outcomes [3,5]. Although the clinical relevance of such parameters is well established in the literature, few studies have scrutinized the reliability of obtaining such measurements. Without establishing such reliability, the validity of previous studies evaluating measurements and outcomes following odontoid fractures may be called into question. Reliability is essential in order to best communicate pathology, classify accurately, and appropriately guide treatment.

Our analysis shows that displacement and angulation of type 2 odontoid fractures can now be very reliably measured via CT scans using modern integrated hospital radiological review systems. This contrasts with Bono et al.'s

Table 2
Summary of measurement results of 37 type 2 odontoid fracture patients by rater

	Displacement (mm)	Angulation (degree)	Proportion of ≥5 mm displacement	Proportion of ≥11° angulation
Attending (A)	2.89±2.71	15.32±12.25	18.9% (7/37)	64.9% (24/37)
Attending (B)	3.05±3.02	16.13±11.43	16.2% (6/37)	62.2% (23/37)
Fellow	3.09±3.33	18.62±13.95	21.6% (8/37)	67.6% (25/37)
Resident	3.04±2.90	18.08±13.36	18.9% (7/37)	67.6% (25/37)
RA	2.73±2.68	12.66±11.89	24.3% (9/37)	51.4% (19/37)

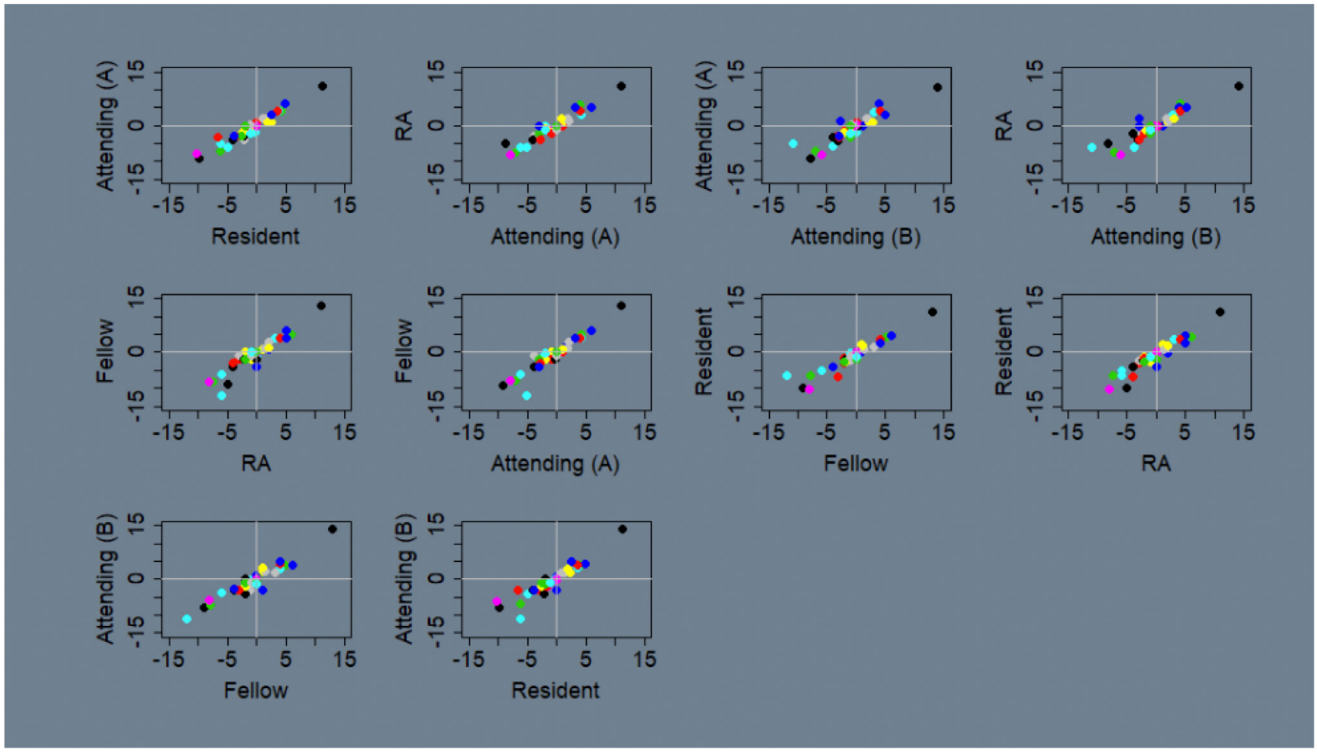

Fig. 2. Graphical representation of pairwise comparison of displacement measurements (in mm) of type 2 odontoid fractures in 37 patients among five raters of varying surgical experience.

previous work that demonstrated poor interrater reliability for measurements. Using the modern technology previously described, excellent inter-rater reliability in the measurements for displacement and angulation was seen. For the reliability of diagnostic classification of displacement and angulation, the corresponding Kappa values indicated "very good" and "good" reliability among raters, respectively. Our data support similar patterns of pairwise reliability and

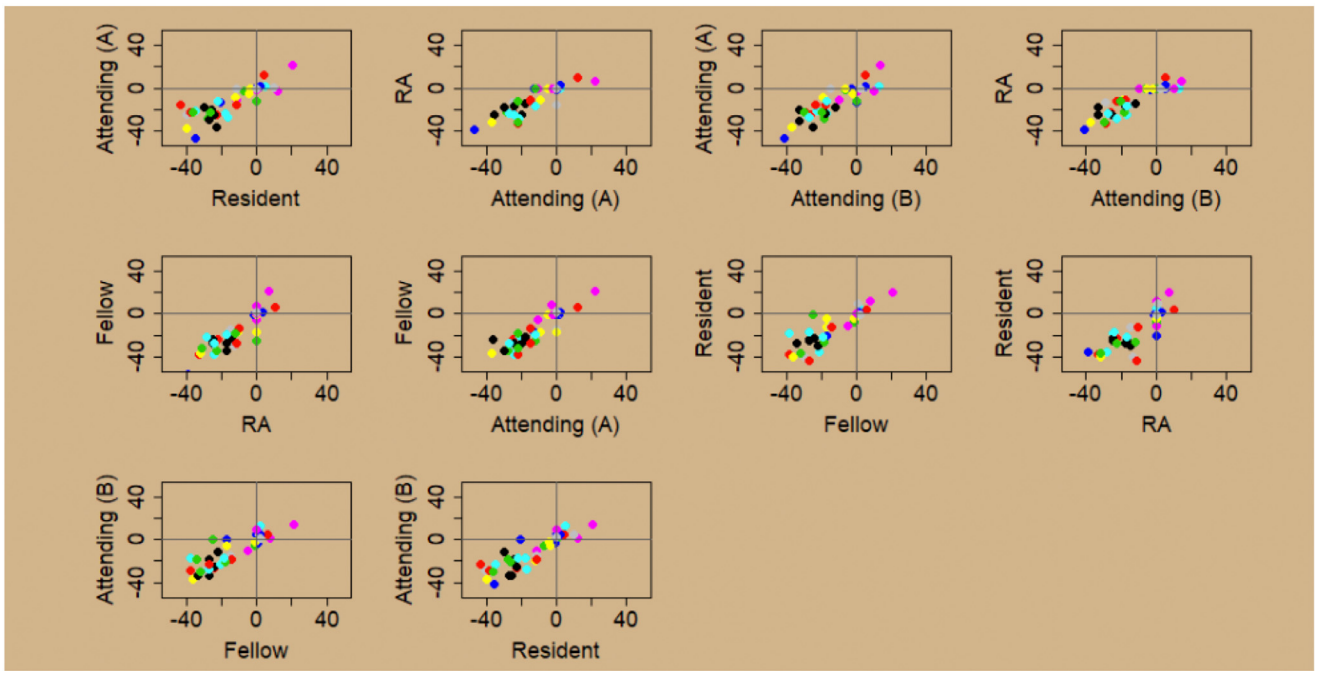

Fig. 3. Graphical representation of pairwise comparison of angulation measurements (in degrees) of type 2 odontoid fractures in 37 patients among five raters of varying surgical experience.

Table 3
Agreement percentages of diagnostic classification of displacement (above vs. below 5 mm) in 37 type 2 odontoid fracture patients among the five raters

	Attending (A)	Attending (B)	Fellow	Resident	RA
Attending (A)	NA	91.9%	97.3%	94.6%	94.6%
Attending (B)	91.9%	NA	89.2%	91.9%	91.9%
Fellow	97.3%	89.2%	NA	91.9%	97.3%
Resident	94.6%	91.9%	91.9%	NA	89.2%
RA	94.6%	91.9%	97.3%	89.2%	NA

Overall agreement percentage: 86.5%.

Table 4
Agreement percentages of diagnostic classification of angulation (above vs. below 11°) in 37 type 2 odontoid fracture patients among the five raters

	Attending (A)	Attending (B)	Fellow	Resident	RA
Attending (A)	NA	81.1%	86.5%	86.5%	81.1%
Attending (B)	81.1%	NA	89.2%	89.2%	89.2%
Fellow	86.5%	89.2%	NA	89.2%	83.8%
Resident	86.5%	89.2%	89.2%	NA	83.8%
RA	81.1%	89.2%	83.8%	83.8%	NA

Overall agreement percentage: 70.3%.

agreement percentages among all the raters, underscoring the ability of all training levels to reliably obtain measurements and stratify based on displacement and angulation.

Although both measurements for displacement and angulation had "excellent" reliability, there was greater ICC for displacement compared with that for angulation. We hypothesize this salient difference to be 2-fold. Although the pictorial representation for measurements demonstrates an odontoid fragment that is uniformly cylindrical in shape, true anatomy often reveals an hourglass-like shape of the dens. The narrowing of the diameter from its base leading up and into the dens may result in a slight variation in rater placement and angle of the posterior tangential line used for angulation measurement. Second, our study allows rater autonomy in selecting what they believed was the best sagittal slice for accuracy of measurements. Parasagittal cuts through the dens may also result in slight differences in placement of the posterior tangential line of the dens fragment. This autonomy may add variability in analysis of the CT imaging. Given that CT scan is a stand-alone modality most often used for cervical trauma, this scenario further mimics the typical clinical scenario more accurately. Despite such potential variability, the ICC for both variables remains "excellent" further underscoring the ability to reliably obtain measurements. We postulate that the same reasons for greater reliability of absolute measurements of displacement compared with angulation also hold true for the better reliability seen in the diagnostic classification of displacement compared with that of angulation.

Surgical management of odontoid fractures is considered appropriate when patients present with increased risk factors for nonunion. The degree and direction of dens fracture displacement has been shown in the literature to significantly correlate with nonunion rates prompting surgical intervention. Although the discussion of the clinical utility of such cutoff points is outside the scope of this study, our study shows the ability of raters of all training levels to stratify patients based on their fracture alignment under a minimum of "good" reliability. We can now be assured that we are appropriately communicating, classifying, and treating such pathology. We believe that the reason for such strong reliability may well be the interval technological advancements in radiographic software. Improvements in iterative reconstruction algorithms and microelectronic circuits with photon-counting detector technology, for example, have resulting in improved tissue composition and image optimization [14–16]. New software incorporates improved artifact reduction and improved signal-to-noise ratio software. Such advancement in the hospitals radiological review systems allow for not only better capturing of images, but also improvements in imaging review permitting more precise measurements.

This study is not without limitations. First, the study does not examine the relationship between measurement reliability and the clinical characteristics of patients' fractures, such as acuity of the fracture and patient age. These factors may influence the morphology of the fracture and therefore affect measurement reliability. Additionally, reliability values may still be limited, as the measurements for each patient were taken only once by each rater and compared for analysis. Interobserver values may improve if each rater were to measure images multiple times. However, typical "real world" clinical situations may not be realistically afforded the opportunity for multiple measurements. Moreover, although all patients were evaluated at a single institution, the CT imaging obtained by each patient was not obtained at a single institution with some scans obtained from various outside facilities. For that reason, we cannot comment specifically on the CT-related hardware, software, or other related imaging specifications used but

rather comment generally on the evolution of imaging technology as a whole. Furthermore, although our reliability may have improved compared with the previous studies caused by advancements of these technologies, such hardware and software may not be available in smaller facilities in the community setting. Therefore, our findings may not be universally applicable across smaller communities and health-care systems. Lastly, as with many observer reliability studies, the raters were not blinded as to their participation in the study; the raters were aware that their measurements would be analyzed and compared. The potential alteration of a raters' behavior in measurement technique due their awareness of being observed, also known as the Hawthorne effect, cannot be calculated.

Conclusion

This work reports the largest inter-rater reliability study looking at type 2 odontoid fractures, and finds that the use of modern CT equipment in association with modern image processing software appears to facilitate more reliable measurements of angulation and translation for type 2 odontoid fractures, even among surgeons with varying degrees of experience.

References

[1] Smith HE, Kerr SM, Fehlings MG, Chapman J, Maltenfort M, Zavlasky J, et al. Trends in epidemiology and management of type II odontoid fractures: 20-year experience at a model system spine injury tertiary referral center. J Spinal Disord Tech 2010;23:501–5.

[2] Gornet ME, Kelly MP. Fractures of the axis: a review of pediatric, adult, and geriatric injuries. Curr Rev Musculoskelet Med 2016;9:505–12.

[3] Hadley MN, Dickman CA, Browner CM, Sonntag VK. Acute axis fractures: a review of 229 cases. J Neurosurg 1989;71:642–7.

[4] Apuzzo ML, Heiden JS, Weiss MH, Ackerson TT, Harvey JP, Kurze T. Acute fractures of the odontoid process. An analysis of 45 cases. J Neurosurg 1978;48:85–91.

[5] Greene KA, Dickman CA, Marciano FF, Drabier JB, Hadley MN, Sonntag VK. Acute axis fractures. Analysis of management and outcome in 340 consecutive cases. Spine 1997;22:1843–52.

[6] Bono CM, Schoenfeld AJ, Anderson PA, Harrop JS, France J, Vaccaro AR, et al. Observer variability of radiographic measurements of C2 (axis) fractures. Spine 2010;35:1206–10.

[7] Muller C, Decker S, France J, et al. C2 fractures. Skeletal trauma: basic science, management, and reconstruction, 2-volume set. 5th ed. Philadelphia, PA: Saunders; 2014. p. 845–70.

[8] Anderson PA, Hah RJ. Occipitocervical and subaxial cervical trauma. Orthopaedic knowledge update: spine. 5th ed. Rosemont, IL: Amer Academy of Orthopaedic; 2017. p. 395–418.

[9] Ryken TC, Hadley MN, Aarabi B, Dhall SS, Gelb DE, Hurlbert RJ, et al. Management of isolated fractures of the axis in adults. Neurosurgery 2013;72(Suppl 2):132–50.

[10] Bono CM, Vaccaro AR, Fehlings M, Fisher C, Dvorak M, Ludwig S, et al. Measurement techniques for upper cervical spine injuries: consensus statement of the Spine Trauma Study Group. Spine 2007;32:593–600.

[11] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–8.

[12] Altman D. Practical statistics for medical research. London: Chapman and Hall; 1991.

[13] Donner A, Rotondi MA. Sample Size requirements for interval estimation of the kappa statistic for interobserver agreement studies with a binary outcome and multiple raters. Int J Biostat 2010;6:1–11.

[14] Johnson TRC, Krauss B, Sedlmair M, Grasruck M, Bruder H, Morhard D, et al. Material differentiation by dual energy CT: initial experience. Eur Radiol 2007;17:1510–7.

[15] Goo HW, Goo JM. Dual-energy CT: new horizon in medical imaging. Korean J Radiol 2017;18:555–69.

[16] Ginat D, Gupta R. Advances in computed tomography imaging technology. Annu Rev Biomed Eng 2014;16:431–53.