Study on accuracy and interobserver reliability of the assessment of odontoid fracture union using plain radiographs or CT scans

Heiko Koller · Klaus Kolb · Juliane Zenner · Jeremy Reynolds ·
Marcel Dvorak · Frank Acosta · Rosemarie Forstner · Michael Mayer ·
Mark Tauber · Alexander Auffarth · Anton Kathrein · Wolfgang Hitzl

Received: 15 December 2008 / Revised: 21 July 2009 / Accepted: 4 August 2009 / Published online: 28 August 2009
© Springer-Verlag 2009

Abstract In odontoid fracture research, outcome can be evaluated based on validated questionnaires, based on functional outcome in terms of atlantoaxial and total neck rotation, and based on the treatment-related union rate. Data on clinical and functional outcome are still sparse. In contrast, there is abundant information on union rates, although, frequently the rates differ widely. Odontoid union is the most frequently assessed outcome parameter and

therefore it is imperative to investigate the interobserver reliability of fusion assessment using radiographs compared to CT scans. Our objective was to identify the diagnostic accuracy of plain radiographs in detecting union and non-union after odontoid fractures and compare this to CT scans as the standard of reference. Complete sets of biplanar plain radiographs and CT scans of 21 patients treated for odontoid fractures were subjected to interobserver assessment of fusion. Image sets were presented to 18 international observers with a mean experience in fusion assessment of 10.7 years. Patients selected had complete radiographic follow-up at a mean of 63.3 ± 53 months. Mean age of the patients at follow-up was 68.2 years. We calculated interobserver agreement of the diagnostic assessment using radiographs compared to using CT scans, as well as the sensitivity and specificity of the radiographic assessment. Agreement on the fusion status using radiographs compared to CT scans ranged between 62 and 90% depending on the observer. Concerning the assessment of non-union and fusion, the mean specificity was 62% and mean sensitivity was 77%. Statistical analysis revealed an agreement of 80–100% in 48% of cases only, between the biplanar radiographs and the reconstructed CT scans. In 50% of patients assessed there was an agreement of less than 80%. The mean sensitivity and specificity values indicate that radiographs are not a reliable measure to indicate odontoid fracture union or non-union. Regarding experience in years of all observers taking part in the study, there were no significant differences for specificity ($P = 0.88$) or sensitivity ($P = 0.26$). Further analysis revealed that if a non-union was judged present by an observer then, on average, each observer changed decision regarding the presence of a ‘stable’ or ‘unstable non-union’ in 4.2 of all the 21 cases (range 0–8 changes per observer). We investigated the interobserver reliability of the assessment of fusion in

H. Koller (✉) · M. Mayer · M. Tauber · A. Auffarth
Department for Traumatology and Sports Medicine,
Paracelsus Medical University, Salzburg, Austria
e-mail: heiko.koller@t-online.de

K. Kolb
Department for Trauma and Reconstructive Surgery,
Katharinenhospital, Stuttgart, Germany

J. Zenner
German Scoliosis Center Bad Wildungen,
Werner-Wicker-Klinik, Bad Wildungen 34517, Germany

J. Reynolds · M. Dvorak
Department for Spine Surgery, Vancouver General Hospital,
Vancouver, Canada

F. Acosta
Department of Neurological Surgery, University of California,
San Francisco, USA

R. Forstner
Institute for Radiodiagnostics, Paracelsus Medical University,
Salzburg, Austria

A. Kathrein
Department for Traumatology and Spine Surgery,
District Hospital Zams, Zams, Austria

W. Hitzl
Research Office, Biostatistics, Paracelsus Medical University,
Salzburg, Austria

odontoid fractures using biplanar radiographs compared to CT scans. A sensitivity of 77% and a specificity of 62% for the radiographs resemble a substantial lack of agreement if different observers evaluate odontoid union. Biplanar radiographs are judged not a reliable measure to detect odontoid fracture union or non-union. The union rates of odontoid fractures have to be revisited and CT scans as the endpoint anchor in outcome studies of treatment related union rates are recommended.

Keywords Odontoid fracture · Non-union · Interobserver reliability · Fusion · Assessment

Introduction

In odontoid fracture treatment there are three major determinants to which outcomes can be compared: (1) clinical outcome in terms of validated measures, (2) functional outcome in terms of atlantoaxial rotation and total cervical spine motion, and (3) treatment-related union rate. There is sparse data on clinical and functional outcome, but fusion rates have been analyzed and discussed abundantly [3, 13, 18]. The last review of the literature in 2006 [22] identified a mean non-union rate of 35% following non-surgical treatment. Prior studies [3, 13, 18] had difficulties in identifying differences in non-union rates with different non-surgical treatment methods because samples analyzed suffered from heterogeneous characteristics concerning fracture subtypes and severity, patients' injury characteristics, age, and comorbidities, as well as the technique selected for immobilization. Koivikko et al. [19] reported the detailed results of a rather homogeneous sample of patients with odontoid type II fractures. Non-union rate was 54% when treated with a halo vest. This data suggested that several patients would benefit from surgery to improve fusion rates, e.g., using anterior odontoid screw fixation (AOSF), most often for odontoid fractures (OF) of Type II. In 1990, Böhler et al. [5] reported a union rate of 80% in 81 patients using AOSF. In 2000, Apfelbaum et al. [3] reported results of a multicenter study yielding a union rate of 88% in 147 OF. Recently, in 2007, Platzer et al. [27] reported a union rate of 93% in 110 patients using AOSF. Several series [1, 3, 16, 27] suggest a reproducible high union rate of about 90% using AOSF for OF and some claim even 100% [24]. Posterior fusion C1–2 is an alternative treatment and even in elderly patients fusion rates were reported as high as 100% [12, 22]. As result of a higher union rate when using surgery the most recent review of literature [22] concluded that surgical intervention is becoming the mainstay of treatment for the majority of patients' age groups, using AOSF in the majority of cases and C1–2

fusion in selected fracture subtypes. However, although the published world literature and common orthopedic understanding would suggest odontoid fracture union to be an important target, there is no single consensus on how to assess fusion in odontoid fracture outcome studies yet. Koivikko et al. [19] employed measurements on flexion-extension films with anatomical boundaries of C1 and C2 drawn on transparent films. The authors reported that the measuring technique had enabled the differentiation between union, non-union, and fibrous union. Likewise, in the study of Apfelbaum et al. [3], lateral radiographs were used to assess fusion and flexion-extension films enabled further delineation of the fusion including the identification of fibrous stable and unstable non-unions. In the series of Platzer et al. [27] only in eight patients bony fusion could not be determined on static radiographs and non-union was diagnosed by cervical CT scans. In another report from that author [26] on a similar sample "fine-cut computed tomography scans were routinely acquired 12 months after surgery or earlier if the adequacy of fusion could not be determined on the standard radiographs" [28]. In a review of literature by Julien et al. [18] the majority of authors were noted to use flexion-extension views with only two performing CT scans in a total of 35 patients at all with patients having various C2-fractures.

The literature provides the impression that CT scanning is not used by most surgeons as part of a standardized diagnostic follow-up protocol for patients treated for odontoid fracture. We identified Fountas et al. [11] who reported a 87% union rate following AOSF in 31 patients. CT scans were reported to be performed at the 6-months follow-up. Using flexion-extension radiographs, only 68% of patients were judged to have a radiographically proven union after follow-up at 12–18 months, respectively. Similarly, Strohm et al. [35] reported a close follow-up of odontoid type II and III fractures treated with AOSF and halo fixation or single usage of the halo. The authors used CT scans at 4, 8, and 12 weeks follow-up. Radiological evidence of complete osseous bridging was only seen in 12% at 12 weeks after trauma. The authors concluded that conventional radiography is not the most suitable technical means to evaluate osseous healing in OF.

Using dynamic radiographs, motion of the distal odontoid fragment can be measured comparing changes of the odontoid-tilt-angle (OTA) between flexion and extensions films. The OTA is frequently used for determination of dynamic instability in odontoid non-unions. We are not aware of any study reporting the interobserver reliability of assessing odontoid fragment motion on flexion-extension films as an indicator for fusion. Additionally, after AOSF motion is frequently lacking and failure of osteosynthesis is often only heralded by screw loosening.

The author found the high union rates reported in literature difficult to reproduce assuming that non-union rates after treatment with the semi-rigid collar, the halo, or surgery using AOSF are higher than previously reported because most data gathered are derived from biplanar radiographic films in neutral position or in flexion and extension. Gross non-union or a highly mobile non-union would probably be detected by most observers. However, atlantoaxial degenerative changes, slight atlantoaxial rotation, implants in the odontoid or lateral masses, as well as a malaligned X-ray beam challenge the determination of a fusion process at the anatomic level of odontoid type II and III fractures. In contrast, CT scans are deemed to be the standard of reference for the assessment of spinal fusion [8, 35].

Until the advent of the validated standard, clinical outcomes which may include fusion rate, accurate and reliable assessment of fusion status is critical when comparing treatment options.

The authors designed an interobserver study including several experienced surgeons and radiologists to determine the interobserver reliability of the assessment of OF union using plain radiographs compared to CT scans. We hypothesized that the reliability of radiographs for determining fusion to be inadequate and consequently that previously reported fusions rates for a treatment modality in the absence of CT follow-up are of reduced value.

Materials and methods

This radiologic interobserver study was performed as part of an extended outcome investigation of the treatment of C2-fractures. For the purpose of the study the author's database of C2-fractures treated surgically or non-surgically in a 10 year period at a trauma center was reviewed. A total of 21 patients treated for OF were selected including the whole spectrum of OF. For the purpose of our study the 21 cases selected fulfilled the following criteria: (a) isolated OF type II, IIa, or III; (b) age of patient between 18 and 90 years at injury; (c) injury transoral and lateral radiographs as well as injury CT scans to classify the detailed OF subtype according to Anderson and Hadley [17]; (d) diagnostic follow-up including transoral and lateral radiographs as well as CT scans that did not differ >6 days from the time of radiographs; (e) minimum follow-up of 6 months; (f) treatment with motion preserving surgical or non-surgical techniques; (g) no injury to the lateral mass of C1 or C2 as shown on reconstructed CT scans; and (h) no injury associated with neoplastic disease, ankylosing spondylitis, and DISH. Meticulous care was provided by reviewing injury CT scans of the patients to prevent inclusion of

those OF having concomitant fractures of the lateral mass. Fractures of the lateral mass of C2 can cause vertebral body asymmetry and distortion of the lateral atlantoaxial joints and, thus, hamper radiographic projection of the odontoid neck or its base.

For the purpose of calculating the interobserver reliability of the assessment of OF union, those transoral and lateral radiographs were selected that had no significant characteristics of a misaligned beam on the transoral odontoid or lateral view, as defined at the vertebral body of C2 (duplicity of parts of the Harris-Ring-C2, posterior vertebral body cortex, deviation of the C2-spinous process in a–p view,...). However, in some cases the performance of transoral radiographs was only moderate resembling the 'real' conditions encountered in the treatment and follow-up of OF, particularly of the elderly patients. All radiographs were taken on a digital X-ray system (*Vertex 3D-III unit, Siemens, Erlangen/Germany*) and stored (*PACS Magic View VC 42, Rel A, Siemens, Germany*). The radiographs were processed using a commercial DICOM-viewing program (*Escape Medical Viewer V3, Escape, Greece*). All cervical spine CT scans were performed on a 16-row helical CT scanner (*SOMATOM Volume Zoom, Siemens, Germany*) using the following imaging parameters: 1 mm detector collimation, 1.25 mm slice thickness; 4/mm rotation table speed; and pitch: 1. According to our trauma protocol transaxial and sagittal reconstructions were performed for the cervical spine with a 2 mm slice thickness and a data reconstruction interval of 2 mm. Coronal images of the odontoid were reconstructed with 1.3 mm slice thickness and a reconstruction interval of 1 mm.

The radiographs and CT scans were shown to a total of 18 orthopedic surgeons, spine surgeons and radiologists with different experience in diagnosing odontoid unions, but at least having 1-year training with orthopedic radiographs. The authors who prepared all cases did not take part in the investigation of interobserver reliability. The study was conducted first at the authors' institution with a total of 11 observers with the images of a PACS work station displayed with a beamer that is used for daily orthopedic diagnostic sessions. Hence, the images were demonstrated both on an Agfa work station (Barco monitors with a resolution of 2 megapixel) and by projection with a high resolution beamer (Canon Xeed SX60) with a resolution of 1.5 megapixel). This group was expressed as Group 1.

Second, the author invited several surgeons to take part in the study and the set of images could be evaluated using a personal computer. As there, theoretically, would be a bias concerning control of these observers regarding moving forward or backward in the image gallery or by different on-screen quality of the radiographs and

Fig. 1 A 52-years-old patient at 25-months follow-up after AOSF for an odontoid type II fracture. Transoral a–p and lateral radiographs did not show clear radiographic criteria for odontoid non-union. Same day CT scans with coronal and sagittal reformations of axial CT scans revealed odontoid non-union and a halo surrounding the screws' shaft. Notably, the patient had excellent functional outcome and no pain suggesting a tight, fibrous union

CT-images shown, the second group is expressed as 'Group 2' for the purpose of statistical analysis.

With special focus on the odontoid, the author processed the radiographs and CT scans using a commercial software program (*Adobe Photoshop*, *Adobe*, *US*) as follows: the digital transoral and lateral radiographs were scaled down with the field of interest placed on the odontoid. No computerized effects were used for further processing regarding image quality. Original DICOM-data with lossless compression were used. The radiographic quality shown to the observers equaled that when using a beamer or commercially available Dicom-viewer software. Likewise, the reconstructed CT scans were processed with focus on the odontoid and no changes regarding quality of scans (brightness, contrast, etc.) were made. To display the complete odontoid on the CT scans, sagittal and coronal reformations were obtained using the CT console's software producing three images through the odontoid in each direction (Fig. 1) resulting in a total of six CT sketches displayed for each patient. Finally, a complete set of

images for each of the 21 patients consisted of biplanar radiographs and six images of the CT scans as illustrated in Fig. 2.

For the Group 2, the digital images were processed without further changes to their quality with a publishing software (*Microsoft Powerpoint*) with the radiographs and CT scans on separated files (Fig. 2). First, the radiographs of one patient were shown for about 1 min, then the CT scans were shown.

The observers were subjected to a query where they noted union/non-union for the radiographs and subsequent CT scans. Additionally, the observers were asked whether they judge the non-union, if present, to be 'stable'. If the observer decided for a non-union he was asked whether this decision was related to the presence of a 'dislocation or angulation', a detected lysis zone, a gap, or lucency. We noted if the observers changed their opinion regarding the presence of a stable or unstable odontoid non-union on the biplanar radiographs based on re-evaluation of the non-union by the subsequent CT scans.

Fig. 2 Two examples of the diagnostic images offered to the 18 observers

Statistical analysis

Statistical analyses included descriptive statistics such as means, standard deviations, and cross-tabulation tables. Independent two-sided Student *t*-test was used for hypothesis testing. A *P* value less than 5% was considered to indicate a statistically significant effect. To measure the performance of the binary classification tests (fusion detected: yes/no), the following computations were made: (1) the number of observers with fusion-positive results for each patient was computed separately for biplanar radiographs and CT scans and illustrated in histograms, (2) overall agreement: For each patient, the proportion of all raters in agreement between biplanar radiographs and CT scans were computed for each patient (e.g., 100% agreement was achieved if 18 out of 18 raters agreed), and (3) With the CT scans considered to be the standard of reference we calculated sensitivities and specificities levels:

- Specificity: The probability (%) of detection of a non-union on radiographs if there is no fusion (in CT scans).
- Sensitivity: The probability (%) of detection of a fusion on radiographs if there is a fusion (in CT scans).

Because of the high number of observers ($n = 18$) and the number of OF/unions presented ($n = 21$) we could not apply kappa statistics.

All analyses were done using Statistica 6.1 (StatSoft, Tulsa, OK) and SPSS 11.0 (SPSS Inc., Chicago). Statistical analysis was performed by one of the authors (WH).

Results

Eighteen observers took part in the study. The observers were from institutions in Canada ($n = 4$), the US ($n = 1$), from Germany ($n = 4$), and from Austria ($n = 9$). Mean experience of the observers with orthopedic radiographs was 10.7 years on average (range 1–22 years).

The 21 patients selected for assessment of interobserver reliability had a mean age of 68.2 ± 18 years (range 21–89 years) at follow-up. Mean time between trauma and index treatment and the radiographic follow-up using radiographs and CT scans was 63.3 ± 53 months (range 6–150 months). There were 9 female (45%) and 12 male patients (55%). Sixteen patients (80%) had type II and IIa fractures and 4 (20%) had type III fractures. Combined fractures including stable anterior and posterior ring fractures of C1 (stable Jefferson burst fracture) or posterior arch fractures were present in five patients (25%). Fifteen patients (75%) were treated by single or double AOSF and the remainder using a halo vest or a Philadelphia collar. Main patient characteristics are summarized in Table 1. This table also presents the percentage of agreement regarding odontoid union for each case as assessed on the CT scans by the 18 observers. Taking into account all observers, according to a majority rule (3:1) the fractures was defined as fused/not fused.

The percentage of overall agreement, as illustrated in Table 2, between radiographic and computer-tomographic findings concerning fusion/non-union ranged between 62 and 90% depending on the observer. If calculating percentage interobserver agreement for the Group 1 (mean

Table 1 Main characteristics of 21 preselected patients after treatment of an odontoid fracture subjected to interobserver assessment of osseous union

No.	Sex	Age FU	Time FU–CT (months)	Odontoid fracture type and assoc #	Tx (No. screws)	Fusion/Non-union ^b	% Fusion in CT scans ^a	% Non-union in CT scans ^a
Show case	F	76	38	II	AOSF (2)			
1	M	52	25	II	AOSF (2)	Non-union	0	100
2	F	89	6	IIa	HTV	Non-union	19	81
3	M	66	54	II and stable JBF	HTV	Non-union	9.5	90.5
4	M	79	6	IIa	AOSF (1)	Non-union	0	100
5	F	30	24	III	Philadelphia C	Fusion	95.2	4.8
6	F	84	16	II and stable JBF	HTV	Non-union	0	100
7	M	61	48	II	AOSF (2)	Non-union	0	100
8	F	87	12	II and post arch C1	AOSF (2)	Fusion	100	0
9	F	77	6	III	HTV	Non-union	0	100
10	M	74	21	II	AOSF (2)	Non-union	0	100
11	F	21	67	II	HTV	Non-union	0	100
12	M	77	128	II and post arch C1	AOSF (2)	Fusion	90.5	9.5
13	M	78	121	III	AOSF (2)	Fusion	100	0
14	F	71	136	II	AOSF (2)	Non-union	4.8	95.2
15	M	86	57	II	AOSF (1)	Non-union	14.3	85.7
16	M	65	111	II	AOSF (2)	Fusion	95.2	4.8
17	M	64	7	II	AOSF (1)	Non-union	0	100
18	M	55	60	II	AOSF (2)	Non-union	0	100
19	M	60	125	III	AOSF (2)	Fusion	100	0
20	F	73	150	II and stable JBF	AOSF (2)	Fusion	100	0
21	F	82	149	III	AOSF (2)	Fusion	95.2	4.8

HTV halo thoracic vest, JBF Jefferson burst fracture of C1 and equivalents, AOSF anterior odontoid screw fixation, in brackets number of lag screws used

^a Percentage (%) of 18 observers rating for the presence of fusion or a non-union based on the CT reconstructions

^b Fusion and non-union as determined by a majority rules (3:1). Show case to the observers, used for explanation of the study design and techniques used

Table 2 Percentage (%) of fusion-positive agreement between radiographic and computer tomography for each of the 18 observers

Obs. 1	Obs. 2	Obs. 3	Obs. 4	Obs. 5	Obs. 6	Obs. 7	Obs. 8	Obs. 9	Obs. 10	Obs. 11
Group 1 (%)										
62	86	90	76	71	86	65	75	86	67	81
Obs. 12	Obs. 13	Obs. 14	Obs. 15	Obs. 16	Obs. 17	Obs. 18				
Group 2 (%)										
76	76	67	81	81	71	67				

Obs Observer

70%) and Group 2 (mean 74%) there were no significant intergroup differences ($P = 0.51$). That is, demonstrating the images with different modalities in Group 1 and 2 had no impact on sensitivity and specificity values.

The mean specificity was 62% and mean sensitivity was 77%. Specificities were 69% in the Group 1 and 51% in the Group 2. There were no significant differences regarding agreement on fusion whether assessed in the Group 1 or the

Group 2 (independent, two-sided Student t -test: $P = 0.16$). Sensitivities were 75% in the Group 1 and 80% in the Group 2 with no significant intergroup differences ($P = 0.81$). Statistical analysis revealed an agreement between the biplanar radiographs and the reconstructed CT scans as follows: 80–100% agreement in only 48% of cases, 60–80% agreement in 33% of cases, 40–60% agreement in 14%, 20–40% agreement in 0%, and 0–20%

Fig. 3 Histograms of fusion-positive results of the 18 observers rating radiographs and CT scans for fusion/non-union of 21 patients. Histogram of CT scans (*red columns*) show more cases in the lower and upper tails as compared to radiographs (*blue columns*)

agreement in 5% of cases. In other words, in about 50% of cases assessed there was an agreement of less than 80%. The number of raters with fusion-positive results for each patient is illustrated for the biplanar radiographs and CT scans in Fig. 3. The histogram shows more cases in the lower (ranging from 0 to 4) and upper tails (ranging from 16 to 18) when fusion was assessed on CT scans as compared to radiographs. Results indicate that the CT scans ended up with more stable results as compared to radiographs.

The results of our statistical analysis of the sensitivity and specificity of the assessment of fusion using biplanar radiographs compared to CT scans are demonstrated in Table 3. The mean specificity of 62% and the mean sensitivity of 77% indicate that radiographs are not a reliable tool when assessing for union. This clinical problem is illustrated in Fig. 4 where the individual sensitivities and specificities of some observers are scattering around the line of random guess. Regarding experience in years of all observers taking part in the study, there were no significant differences regarding specificity ($P = 0.88$) and sensitivity ($P = 0.26$) when fusion was assessed by the Group 1 or the Group 2.

Fig. 4 Illustration of the individual level of sensitivities and false positive rates of using radiographs for fusion assessment by 18 observers in odontoid fractures (CT scan is standard of reference)

We noted if the observers amended their opinion regarding the presence of a stable or unstable odontoid non-union based on the subsequent CT scans. On average, each observer changed his decision in 4.2 of 21 cases with a range from 0 to 8 changes per observer. Assuming the non-union rate of our artificial sample to be 62% (13 cases), then a change of diagnosis from stable non-union to unstable non-union based on the CT scans occurred in about one-third of non-unions. That is, in each third non-union patient the diagnosis ‘instability’, that might indicate subsequent surgical treatment, would be added because a CT scan was performed.

Discussion

With any kind of proper immobilization, most fractures of the odontoid heal by the formation of endosteal bone [31]. Therefore, follow-up after treatment of OF includes radiographic control for the progression of osseous healing. Although it has been argued that the radiographic determination of fusion may be difficult and subject to observer variability [1, 34], it seemed to be the most appropriate outcome measure besides functional and clinical outcome and treatment related complications. The term union rate

Table 3 Sensitivity and specificity values of the assessment of biplanar radiographs for evidence of fusion compared to the CT scan in 18 observers

	Mean	SD	CI -95% (%)	CI +95% (%)	Minimum	Maximum	Percentile 10	Percentile 90
Specificity	62	26	49	75	23	92	23	92
Sensitivity	77	16	69	85	43	100	50	100

An overview of the 18 individual sensitivity and specificity values is illustrated in Fig. 2

has been used as a surrogate marker of success when comparing treatment modalities [1]. Fusion is thought to be solid when there is no motion on flexion-extension films and when bridging trabeculae are seen across the fracture level [7, 36]. However, in OF it can be very difficult to identify motion when there is a tight pseudoarthrosis, resembling a ‘fibrous union’ or ‘fibrous non-union’. In particular, if the non-union is adjacent to anterior odontoid screws, a fractured lateral mass or oblique in one or both planes with the X-ray beam not angulated perfectly, then non-union is difficult to detect and flexion-extension films will not reveal mobility when screws still halter motion and when screws do not show gross loosening or cut-out. In addition, assessment of the OTA can be difficult on plain radiographs. In a recent study by the authors [20] analyzing the interobserver reliability of morphometrical measurements at the C1–2 junction, the intraclass coefficient (ICC) for the OTA was low (ICC = 0.37). In this context it is of note that risk factors for non-union have been investigated with great eager [2, 19, 22] as well as the interrater variability of the OF classification according to Anderson [4]. The latter study showed that for fracture classification CT scans are mandatory. In contrast, studies focusing on the accuracy of diagnostic techniques for fusion assessment are lacking. In a recent study by Strohm et al. [34] the authors compared the radiographic course of 25 patients with mean age of 72 years suffering from OF treated with either AOSF and a halo or the halo only. At follow-up of 4, 8, and 12 weeks patients were subjected to biplanar radiographs and CT scans. On a scale of 1 (no osseus bridging) to 5 (complete osseus bridging) the patients yielded an average of 2.9 and 2.0 using radiographs and CT scans at the 12 weeks interval. The authors concluded that radiographs often showed false-positive result regarding fusion assessment and CT scans were recommended for specific diagnosis in OF. However, data on accuracy of the radiographs compared to CT scans were not provided.

Assessing fusion of an OF on radiographs is difficult, similar to the subaxial spine. However, in comparison to OF union, classification and grading systems exist for the subaxial spine fusion and literature shows that there is increasing research on the accuracy of fusion assessment done at the subaxial spine [6, 8, 36, 37]. Likewise, in a previous investigation [21] on fusion rates in ACDFP for subaxial cervical spine injuries, the author noted a Kappa value of 0.5 and 0.64 for the interobserver assessment of the cephalad and caudad graft-endplate junction of the fusion-level when using the classification of Vavruch [37]. Using the classification of Bridwell [6], the Kappa statistic was 0.33, resembling only fair agreement. Ploumid et al. [29] noted after a review of 47 patients 9–15 months after cervical fusion surgery that CT scan assessment led to a higher rate of diagnosed pseudoarthrosis than plain

radiographs, with a pseudoarthrosis rate of 13–31% for CT scans depending on the observer compared with 2–16% for plain radiographs. The author noted that interobserver agreement was higher for CT scans than for plain radiographs as with the current study. Buchowski et al. [7] assessed the reliability of plain radiographs, CT, and MRI to detect a pseudoarthrosis after ACDF compared with intraoperative exploration [7]. The authors observed that findings on CT scans were in agreement with intraoperative assessment of fusion status in 78.6–85% of cases depending on the observer with a Kappa statistic of 0.81 (range 0.71–0.87, $P < 0.05$). When radiographs and CT scans were used in concert the agreement between the radiographic studies and intraoperative findings increased from 78.6 to 92.3 and the Kappa statistic increased to 0.85 (range 0.71–1.00, $P < 0.05$). CT scans were judged highly accurate for determining the fusion status. Similarly, Epstein et al. [10] reported an alleged fusion rate of 50% when using plain radiographs and 83% when using CT scans. The authors concluded that CT scans were a more accurate assessment tool of fusion status than plain radiographs.

The results of the studies by Ploumid, Buchowsky, and Epstein are echoed by the current investigation. We observed a high percentage overall agreement in identifying union and non-union within 18 observers in 21 cases using CT scans. However, the interrater agreement on fusion or non-union based on the radiographs was disappointing compared to the CT scans, with a sensitivity of 77% and a specificity of 62% only for the radiographs. As illustrated in Table 1 and the histogram of Fig. 3, using CT scans caused a rather homogeneous rating regarding the presence of fusion and a high rate of interrater accuracy with agreement greater than 90% in 19 of 21 cases (90.5%). Interestingly, the raters’ experience in years had no impact on the accuracy of assessing the fusion using radiographs compared to CT scans. Results indicate that assessment of fusion in OF is rather an issue of the diagnostic instrument used (radiographs versus CT scan) than related to the observers’ experience.

Studies on OF reported the non-union ranging from 26 to 80% [22] with one rather homogeneous series reporting a radiographical non-union rate in type II fractures of 35% following the treatment with the halo [32]. On the one hand, different non-union rates with identical techniques refer to the heterogeneous characteristics of the samples studied. But, on the other hand, we suggest that differences in non-union rates also refer to the inadequate accuracy of radiographs to define OF union. In transoral radiographs the assessment of the odontoid can be hampered by, e.g., the teeth, the jaw, a Mach-phenomenon (air entrapment below the tongue) mimicking an OF or non-union gap; tilting of the axis and the fracture trace/non-union gap

being projected off the X-ray beam (particularly in oblique fractures) can mislead the observer. In lateral radiographs tilting of the axis vertebra causes asymmetry of the Harris-Ring-C2 [14, 15, 23, 25, 33] thereby obscuring the base of the odontoid neck. In addition, atlantoaxial degeneration and osteophytic changes, rotation and tilting of the lateral mass of C1 as well as osteosynthesis material in the odontoid can hamper diagnostic of a union or non-union following OF.

As a consequence of the current study the authors call for odontoid outcome studies that use CT scans for the analysis of treatment-related union rate. Union rates have to be revisited and an increased non-union rate following surgical or non-surgical approaches will be brought to light. Anectodically, it was thought that a patient who has intact ligamentous structures despite poor bone healing at the odontoid might be stable but prone to a disabling injury with further trauma [30]. Although osseous union seems desirable, in many cases fibrous union is clinically acceptable and several authors experienced that OF can be stable despite the lack of osseous union on radiographs by achieving a fibrous union [9, 19, 28, 32]. Seybold et al. [32] reported a good functional outcome that was not dependent on radiographic criteria for union and more dependent on return to pain-free, independent living, regardless of union rate. Only 19% of all non-unions required late posterior fusion C1–2 because of symptomatic pseudoarthrosis. Likewise, Clark and White recommended no further treatment in case of a stable fibrous non-union [9]. Without CT scans Platzner et al. [28] selected 14 patients, out of a total of 90 elderly patients with type II OF, that were suspected to have a radiographic non-union. The suspicion was verified on CT scans in all cases. Notably, non-union was not necessarily equivalent to a bad result and in more than 50% of their patients with non-union there was a stable and asymptomatic fibrous union. It seems that the stable non-union can be accepted, particularly in the elderly patient [22], but the definition of what is a ‘stable’ fibrous union demands further long-term analysis using CT scans in concert with dynamic films. In addition, e.g., the biomechanical testing of the stability of in vivo generated odontoid pseudoarthroses and fusions in animals has the potential to bring light into the definition of stability and the treatment of odontoid non-unions.

Limitations

Although data on intraobserver agreement would be worthy, the observers were not able to commit enough time to collect data enabling calculation of both inter- and intraobserver reliability and additional information was judged of limited clinical value. The cases were selected due to

their completeness (radiographs and CT scans) and according to our definition of sufficient quality of the radiographs. Accordingly, a selection bias might exist, but, only favouring the union assessments when using the radiographs. In contrast, one feature of the sample of observers was a mean experience of 10.7 years resembling a clinically experienced group of mainly senior radiologists and surgeons. In addition, 18 observers rated on 21 radiographic cases resembling an extensive approach to analysis of interobserver reliability.

In this study the radiographs were presented to 11 clinicians using the common setup of a diagnostic meeting with the images displayed on a high-fidelity beamer (Group 1). The other group of eight clinicians made diagnosis on a computer workstation (Group 2). The percent agreement was not significantly different between both groups. That is, we can exclude bias related to the observers in the Group 2 flipping backwards and forwards in their image presentation and related to a theoretically reduced quality of image presentation in Group 2.

It is possible that there would be better agreement between radiographs and CT scans if the author would have offered physical copies of the radiographs to all observers, but technically this was not possible and is judged unlikely to change the main results.

Acknowledgments Special thanks go to H. Resch, Prof.; M. Eppel, MD; Dr. C. Meyer, MD; F. Romeder, MD; S. Paquette, MD, S. Kingwell; MD, D. Savaranja, MD, P. Grützner, PD for significant support and participation in the interobserver study.

References

1. American Association of Neurological Surgeons (2002) Isolated fractures of the axis in adults. *Neurosurgery* 50(suppl 3):140–147
2. AO Spine International (2008) Type II Odontoid fractures. Evidence-based spine surgery 2:1–8
3. Apfelbaum RI, Lonser RR, Veres R, Casez A (2000) Direct anterior screw fixation for recent and remote odontoid fractures. *J Neurosurg (Spine 2)* 93:227–236
4. Barker L, Anderson J, Chesnut R, Nesbit G, Tjauw T, Hart R (2006) Reliability and reproducibility of dens fracture classification with use of plain radiography and reformatted computer-aided tomography. *J Bone Joint Surg* 88-A:106–112
5. Böhler JB, Poigenfürst J, Gaudernak W, Hintringer T (1990) Odontoid screw osteosynthesis. *Operat Orthop Traumatol* 2:75–83
6. Bridwell KH, Lenke LG, McEnery KW, Baldus C, Blanke K (1995) Anterior fresh frozen structural allografts in the thoracic and lumbar spine. Do they work if combined with posterior fusion and instrumentation in adult patients with kyphosis or anterior column defects? *Spine* 20:1410–1418
7. Buchowski JM, Liu G, Bunmaprasert T, Rose PS, Riew D (2008) Anterior cervical fusion assessment. Surgical exploration versus radiographic evaluation. *Spine* 33:1185–1191
8. Carreon LY, Glassman SD, Schwender JD, Subach BR, Gornet MF, Ohno S (2008) Reliability and accuracy of fine-cut computed

- tomography scans to determine the status of anterior interbody fusions with metallic cages. *Spine J* 8(6):998–1002
9. Clark CR, White AA (1985) Fractures of the dens: a multicenter study. *J Bone Joint Surg* 67-A:1340–1348
 10. Epstein NE, Silvergleide RS (2003) Documenting fusion following anterior cervical surgery: a comparison of roentgenogram versus two-dimensional computed tomographic findings. *J Spinal Disord Tech* 16:243–247
 11. Fountas KN, Kapsalaki EZ, Karampelas I, Feltas CH, Dimopoulos VG, Machinis TG, Nikolakakos LG, Boev AN, Choudhri H, Smisson HF, Robinson JS Jr (2005) Results of long-term follow-up in patients undergoing anterior odontoid screw fixation for type II and rostral type III odontoid fractures. *Spine* 30:661–669
 12. Frangen TM, Zilkens C, Muhr G, Schinkel C (2007) Odontoid fractures in the elderly: dorsal C1–2 fusion is superior to halo-vest immobilization. *J Trauma* 63:83–89
 13. Govender S, Maharaj JF, Haffajee MR (2000) Fractures of the odontoid process. *J Bone Joint Surg* 82-B:1143–1147
 14. Hähnle U, Wiesniewski TF, Craig JB (1999) Shear fracture through the body of the axis vertebra. *Spine* 24:2278–2281
 15. Harris JH et al (1984) Low (type III) odontoid fracture: a new radiographic sign. *Radiology* 153:353–356
 16. Harrop JS, Przybylski GJ, Vaccaor AR, Yalamanchili K (2000) Efficacy of anterior odontoid screw fixation in elderly patients with type II odontoid fractures. *Neurosurg Focus* 8(6):e6
 17. Hadley MN, Browner C, Sonntag VK (1985) Axis fractures: a comprehensive review of management and treatment in 107 cases. *Neurosurgery* 17:281–289
 18. Julien TD, Frankel B, Traynelis VC, Ryken TC (2000) Evidence based analysis of odontoid fracture management. *Neurosurg Focus* 8(6):1–6
 19. Koivikko MP, Kiuru MJ, Koskinen SK, Myllynen P, Santavirta S, Kivisaari L (2004) Factors associated with nonunion in conservatively treated type-II fractures of the odontoid process. *J Bone Joint Surg* 86-B:1146–1151
 20. Koller H, Acosta F, Tauber M, Komarek E, Fox M, Moursy M, Hitzl W (2009) C2-fractures—part I: Quantitative morphology of the C2 vertebra is a prerequisite for the radiographic assessment of posttraumatic C2-alignment and the investigation of clinical outcomes. *Eur Spine J* (online 03-2009)
 21. Koller H, Reynolds J, Zenner J, Forstner R, Hempfing A, Maislinger I, Kolb K, Tauber M, Resch H, Mayer M, Hitzl W (2009) Mid- to long-term outcome of instrumented anterior cervical fusion for subaxial injuries. *Eur Spine J* (E-pub 02-2009)
 22. Maak TG, Grauer JN (2006) The contemporary treatment of odontoid injuries. *Spine* 31:S53–60
 23. Mortelsmann LJ, Geusens EA et al (1999) Harris or Axis Ring: an aid in diagnosing low (type 3) odontoid fractures. *Eur J Surg* 165:1138–1141
 24. Olerud C (2008) Dens fractures. *Spineweek*, Geneva, Switzerland
 25. Pellei DD (2000) The fat C2 sign. *Radiology* 217:359–360
 26. Platzer P, Thalhammer G, Oberleitner G, Schuster R, Vecsei V, Gaebler C (2007) Surgical treatment of dens fractures in elderly patients. *J Bone Joint Surg* 89-A:1716–1722
 27. Platzer P, Thalhammer G, Ostermann R, Wieland R, Vecsei V, Gaebler C (2007) Anterior screw fixation of odontoid fractures comparing younger and elderly patients. *Spine* 32:1714–1720
 28. Platzer P, Thalhammer G, Sarahrudi K, Kovar F, Veksler G, Vecsei V, Gaebler C (2007) Nonoperative management of odontoid fractures using a halo-thoracic vest. *Neurosurgery* 61:529–530
 29. Ploumiss A, Mehbod A, Garvey T, Gilbert T, Transfeldt E, Wood K (2006) Prospective assessment of cervical fusion status: plain radiographs versus CT-scan. *Acta Orthop Belg* 72:342–346
 30. Polin RS, Szabo T, Bogaev CA, Replogle RE, Jane JA (1996) Nonoperative management of type II and III odontoid fractures: the Philadelphia collar versus the halo vest. *Neurosurgery* 38:450–456
 31. Schatzker J, Rorabeck CH, Waddell JP (1975) Non-union of the odontoid process. *Clin Orthop* 108:127–137
 32. Seybold EA, Bayley JC (1998) Functional outcome of surgically and conservatively managed dens fractures. *Spine* 23:1837–1846
 33. Smoker WR, Dolan KD (1987) The “fat” C2: a sign of fracture. *Am J Radiol* 148:609–614
 34. Strohm PC, Bley TA, Ghanem N, Scheck B, Südkamp NP, Müller CA (2006) Clinical and radiological findings after different treatment of odontoid fractures type Anderson II and III. *Acta Chir Orthop Traumatol Czech* 73:151–156
 35. Tan GH, Goss BG, Thorpe PJ, Williams RP (2007) CT-based classification of long spinal allograft fusion. *Eur Spine J* 16:1875–1881
 36. Tuli SK, Chen P, Eichler M, Woodard EJ (2004) Reliability of radiologic assessment of fusion: cervical fibular allograft model. *Spine* 29:856–860
 37. Vavruch L, Hedlund R, Javid D, Leszniewski W, Shalabi A (2002) A prospective randomized comparison between the Cloward procedure and a carbon fiber cage in the cervical spine. A clinical and radiologic study. *Spine* 27:1694–1701