

RESEARCH ARTICLE

Does Computed Tomography Change our Observation and Management of Fracture Non-Unions?

Ydo V. Kleinlugtenbelt, MD; Vanessa A.B. Scholtes, PhD; Jay Toor, MD; Christian Amaechi, MD; Mario Maas, MD, PhD; Mohit Bhandari, MD, PhD; Rudolf W. Poolman, MD, PhD; Peter Kloen, MD, PhD; COAST Group

Research performed at Academic Medical Center, University of Amsterdam, The Netherlands

Received: 18 April 2016

Accepted: 07 August 2016

Abstract

Background: The purpose of this study was to determine whether Multi-Detector Computed Tomography (MDCT) in addition to plain radiographs influences radiologists' and orthopedic surgeons' diagnosis and treatment plans for delayed unions and non-unions.

Methods: A retrospective database of 32 non-unions was reviewed by 20 observers. On a scale of 1 to 5, observers rated on X-Ray and a subsequent Multi Detector Helical Computer Tomography (MDCT) scan was performed to determine the following categories: "healed", "bridging callus present", "persistent fracture line" or "surgery advised". Interobserver reliability in each category was calculated using the Interclass Correlation Coefficient (ICC). The influence of the MDCT scan on the raters' observations was determined in each case by subtracting the two scores of both time points.

Results: All four categories show fair interobserver reliability when using plain radiographs. MDCT showed no improvement, the reliability was poor for the categories "bridging callus present" and "persistent fracture line", and fair for "healed" and "surgery advised". In none of the cases, MDCT led to a change of management from nonoperative to operative treatment or vice versa. For 18 out of 32 cases, the treatment plans did not alter. In seven cases MDCT led to operative treatment while on X-ray the treatment plan was undecided.

Conclusion: In this study, the interobserver reliability of MDCT scan is not greater than conventional radiographs for determining non-union. However, a MDCT scan did lead to a more invasive approach in equivocal cases. Therefore a MDCT is only recommended for making treatment strategies in those cases.

Keywords: Computed Tomography, Non-union, Fracture, Reliability

Introduction

The most widely used tool for diagnosis of non-union is conventional radiography (1). Characteristic radiological findings of non-union are lack of bone bridging and persistence of the fracture line (2). The diagnosis of non-union on plain radiographs can be open to interobserver variability. To complicate matters further, surgeons disagree on when a fracture is healed (3). Overlying hardware and sclerosis can also obscure the original fracture lines and thus hamper diagnosis. Alternative imaging studies such as computer tomography (CT) are often used in clinical practice but their success rate has not yet been widely reported in the literature. Furthermore, there is an increased concern regarding underestimation of the oncogenic risks of radiation-based imaging (4,5).

Many physicians are convinced that CT evaluation

is impaired by metal hardware (6). However, it is our experience that bone consolidation (or lack thereof) is readily diagnosed with the use of Multi Detector Helical CT (MDCT) with Multi Planar Reconstructions (MPR) scanning despite present hardware. MDCT with MPR examination could therefore be suitable in the management of delayed and non-union fractures.

The goal of our study was to evaluate whether MDCT with MPR assessment can play a role in assisting orthopedic surgeons in diagnosing delayed and non-unions and, if necessary, adjusting treatment strategy. Firstly, we aimed to evaluate interobserver reliability in diagnosis of non-union in fractures of the appendicular skeleton with MDCT with MPR. Secondly, we aimed to determine the benefit of MDCT with MPR in the evaluation of non-union. We hypothesized that an additional MDCT with MPR increases reliability and would benefit in the

Corresponding Author: Ydo V. Kleinlugtenbelt, Department of Orthopaedic and Trauma Surgery, Deventer ziekenhuis, The Netherlands
Email: ijdoklb@hotmail.com



THE ONLINE VERSION OF THIS ARTICLE
[ABJS.MUMS.AC.IR](http://abjs.mums.ac.ir)

assessment of non-union.

Materials and Methods

Study Design

Retrospectively, an online database of radiographic images and a brief clinical synopsis was established including patients who visited the Academic Medical Center (AMC) in Amsterdam, the Netherlands (Ethics approval was not required, because all patient data was anonymized). We conducted this study according to the Collaboration for Outcome Assessment in Surgical Trials (COAST) guidelines (7). The database was uploaded to a COAST website where participants could log in, view images, and assign ratings.

Study Participants

Our patient population consisted of 50 cases. Inclusion criteria were patients with fractures complicated either by delayed or non-union who had X-ray as well as a MDCT with MPR. All types of fixation were included. Reviewed fracture sites were in the appendicular skeleton.

Observers

The online panel reviewing the database consisted of 20 orthopaedic surgeons who were participants of the COAST research group. Most observers, 63%, had over 10 years experience in fracture care. 60% practiced in North America while 15% and 10% practiced in Europe and Australia, respectively.

Imaging protocol

X-Ray and MDCT with MPR were carried out on each patient using a current AMC radiological protocol. Slice thickness and pitch were varied depending on fracture site to ensure optimal imaging. MDCT scans of the femur were carried out using a thickness of 2.0 mm and a pitch of 0.875 while scans of the humerus used a 0.6 mm thickness and 0.950 pitch.

Scoring procedure

Data was presented to the panel in two separate rating sessions. During the first session, plain radiographs were evaluated. After two weeks, there was a second session evaluating MDCT imaging. MDCTs were converted to short video clips with sagittal and coronal views of fracture images.

The evaluation of non-union had to be scored on four categories: "healed", "bridging callus present", "persistent fracture line" and "surgery advised" for each case. Each observer was asked to rate each category on a Likert type 1-to-5 response scale, with a score of 1 corresponding to "strongly disagree", 2 to "disagree", 3 to "undecided", 4 to "agree" and 5 to "strongly agree".

Sample size calculation

In order to determine the number of subjects to be evaluated in this study, we calculated sample size using an estimated Intraclass Correlation Coefficient (ICC) (8). We expected increased reliability when adding a CT, so we estimated an ICC between 0.5 and 0.8. To obtain a 95% confidence interval with a confidence level of ± 0.10 ,

we needed 23 to 65 patients and more than 10 observers.

Statistical analysis

We present descriptive statistics of the study patients. Means and standard deviations (SD) are given for continuous data. Inter-observer reliability was evaluated using ICCs. Although other reliability studies used Kappa statistics, the ICC more accurately estimates the reliability of measurements made by different observers, or the same observer, on different occasions. Kappa statistics are less accurate if responses are skewed and are only appropriate for categorical data (7). The values were interpreted as described by Cicchetti (9). ICC values less than 0.40 indicate "poor agreement", values between 0.40 – 0.59 indicate "fair agreement", values between 0.60 – 0.74 indicate "good agreement", and values ranging from 0.75 – 1.00 indicate "excellent agreement".

To determine the influence of a MDCT on rater observations, the difference between the Likert scale scoring of each case caused by addition of MDCT to X-Ray was calculated for each individual rater. These differences were quantitatively determined for each of the four categories. A score of 0 means no change in rater observation of the case with the addition of MDCT. A positive change in the Likert score indicates a rater observing more healing, more bridging callus, persistent fracture line and stronger advisement of surgery respectively. A negative change in Likert score indicates the inverse. The frequency of each score (-4 to 4) was determined. We present the differences in these scores graphically for each category. To determine whether the observation on MDCT truly changes treatment plans, we calculated the average score of all observers per case for both X-ray and MDCT for the category "surgery advised". We divided the results in three groups; 1.00 – 2.50, 2.51 – 3.50, 3.51 – 5.00, which are corresponding to "nonoperative", "undecided", and "operative treatment", respectively. All statistics were performed using SPSS Version 22.

Results

Study Participants

50 patients with suspected non-unions were evaluated by the senior author (PK), with 18 cases being excluded due to incomplete data sets, leaving 32 cases available for online evaluation. Patient and case characteristics are shown in Table 1.

Observers

Of the 20 raters included, 19 completed the radiograph module, and 18 raters completed the MDCT module. One rater was excluded having only completed a third of the CT module, leaving 17 observers for analysis.

Interobserver Reliability

All four categories show a fair interobserver reliability when using plain radiographs. MDCT showed no significant improvement, the reliability was poor for the categories "bridging callus present" and "persistent fracture line", and fair for "healed" and "surgery advised" [Table 2].

Influence of MDCT on Evaluation of Non-Union

The mean difference (SD) between the Likert score on X-ray and MDCT per category is demonstrated in Table

Table 1. Patient Characteristics	*Number of Patients (%) n=32
Gender	
Male	(70) 22
Female	(30) 10
Location of Fracture	
Lower Extremity	(84) 27
Upper Extremity	(16) 5
Type of Fixation	
ORIF/ LISS/ Screw	(76) 24
PFN/ IM	(9) 3
External Fixation	(9) 3
Conservative	(6) 2
Age (Mean (SD))	(15.4) 44.6

*Although the database consisted of 50 cases, 18 were excluded because of incomplete data sets. ORIF (= Open Reduction Internal Fixation), LISS (= Less Invasive Stabilization System), PFN (= Proximal Femoral Nail), IM (Intra medullary). SD (Standard Deviation)

Category	Interobserver reliability with 95 % CI		Difference in Likert score	
	X ray	CT scan	Mean difference	SD
Healed	(0.37-0.64) 0.49	(0.61 – 0.33) 0.46	-0.15	1.265
Bridging callus	(0.30-0.58) 0.42	(0.49 – 0.23) 0.34	-0.03	1.276
Persistence of the fracture line	(0.35-0.63) 0.47	(0.53 – 0.26) 0.37	0.25	1.143
Surgery advised	(0.27-0.54) 0.49	(0.61 – 0.33) 0.45	0.39	1.215

2. The distribution per category is shown in Figures 1-4. Figure 2 (“bridging callus present”) shows a normal bell curve distribution. This demonstrates that there is no systematic change in observing the presence of bridging callus when using a MDCT. Figures 1, 3 and 4 show also a bell curve distribution, but these are slightly skewed. This is most pronounced in Figure 4 (“surgery advised”).

The amount of cases with nonoperative, undecided and operative treatment based on the average Likerts scale score of all observers are presented in Table 3. In none of the cases MDCT led to a change of management from nonoperative to operative treatment or vice versa. 18 out of 32 cases showed no change in treatment plans. In seven cases MDCT led to operative treatment while on

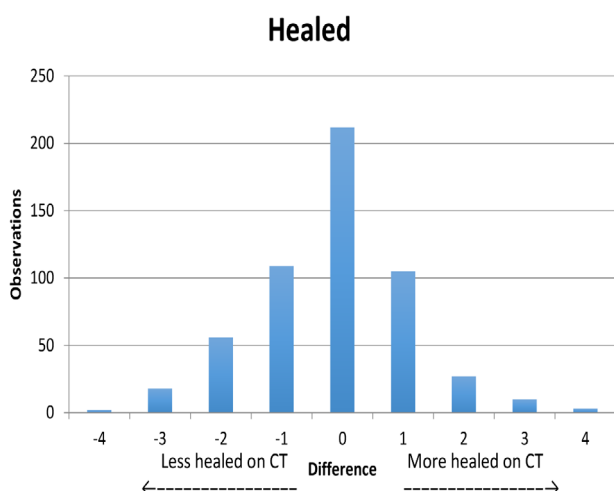


Figure 1. The distribution of the difference between the Likert score on X-ray and MDCT for the category “healed”.

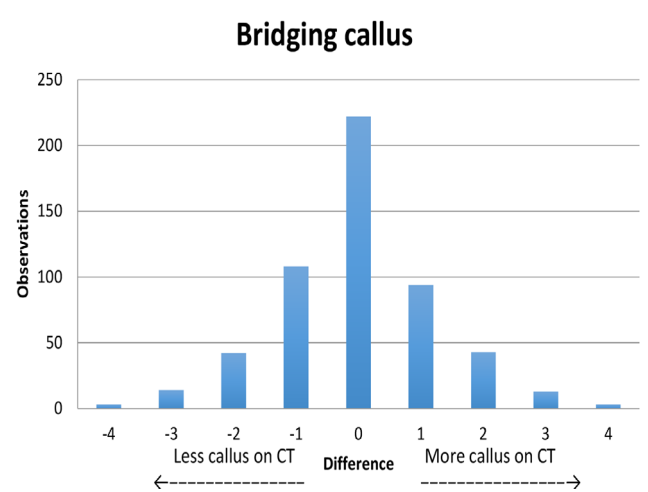


Figure 2. The distribution of the difference between the Likert score on X-ray and MDCT for the category “bridging callus”.

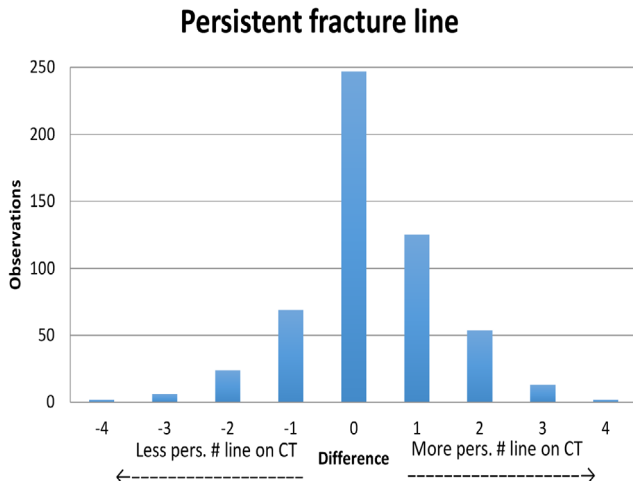


Figure 3. The distribution of the difference between the Likert score on X-ray and MDCT for the category "persistent fracture line".

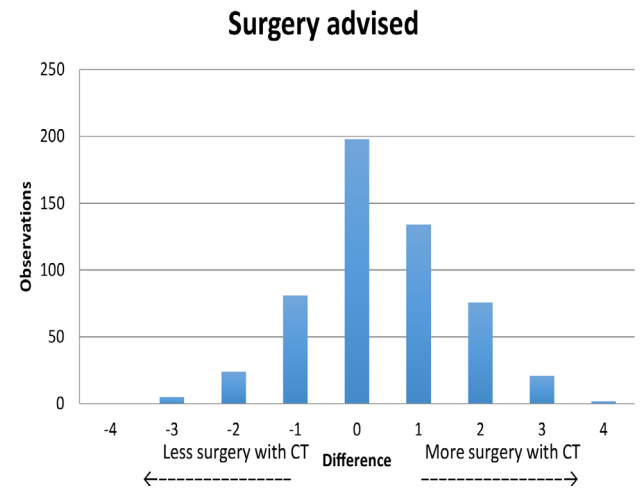


Figure 4. The distribution of the difference between the Likert score on X-ray and MDCT for the category "surgery advised".

Table 3. The average Likert scale score of all observers on the category "surgery advised" per case. Comparing X-ray and MDCT

		MDCT scan			Total
		Nonoperative (1.00 – 2.50)	Undecided (2.51 – 3.50)	Operative (3.51 – 5.00)	
X-ray	Nonoperative (1.00 – 2.50)	6 (60%)	4 (40%)	0 (0%)	10
	Undecided (2.51 – 3.50)	2 (11%)	9 (50%)	7 (39%)	18
	Operative (3.51 – 5.00)	0 (0%)	1 (25%)	3 (75%)	4
Total		8	14	10	32

X-ray the treatment plan was undecided.

Discussion

Key findings

Contradictory to our initial hypothesis, in this reliability study we found that MDCT as a diagnostic tool did not have a greater interobserver reliability than X-Ray in detecting non-union. MDCT scan showed a slight change in the evaluation of non-union. However, with careful analysis of graphs in Figures 1-4, it is doubtful whether a change of -0.03 -0.15 and 0.25 for the categories "bridging callus", "healed" and "persistence of fracture line", respectively, on a Likert scale from 1-5 is clinically relevant. This does not correspond with a systematic change in the evaluation of non-union that X-Ray interpretation already made. On the other hand, for daily practice, a possible change in management is most relevant. Figure 4 (category "surgery advised") demonstrated a more pronounced skew than the other figures, which could be due to a more invasive approach when using a MDCT. However, when individual cases were considered, management did not change from non-operative to operative treatment in any case. Furthermore, in case the management was undecided based on X-Ray, a MDCT led more often to operative (39%) than non-operative treatment (11%). Thus, especially in equivocal cases, MDCT led to a more invasive approach than X-Ray. In these equivocal cases, a MDCT seems to be justified. However, when a non-operative management is chosen based on plain radiographs, a MDCT has no benefits in this

study. Therefore, a risk-benefit decision must be made at the level of the individual patient and should involve balancing the highly context-dependent benefits of imaging against the patient-specific cumulative oncogenic risk (4).

Strengths and limitations

Our study was strengthened by the generalizability of results due to its international nature. The web-based rating sessions were not hampered by geographical boundaries and illustrate possibilities for international research collaboration. Metal hardware did not impair the CT evaluation of fracture sites. 17 of the 20 raters were successfully able to complete the session and awarded a score for each video. This is therefore an effective way to present data to participants from different countries.

On the other hand, our study had several drawbacks. Firstly, we anticipated including 50 patients. Unfortunately, sufficient images were only available for 32 cases. Secondly, participating raters were more difficult to recruit than anticipated. Although we made several attempts to have all raters completing the modules, three were too busy in their clinical practice to complete the full set. Another limitation in this study is that the sample size was underestimated for the interobserver reliability. A possible reason MDCT imaging showed poorer agreement than expected may be due to the conversion to video. In this study, MDCT images were converted into pre-recorded video format, not allowing the opportunity for the observer to adjust the

pace and rotational angle, which would be possible in a standard clinical setting. Besides, during the second scoring session the MDCT images were evaluated without the plain radiographs, which did not represent the real life clinical setting. Furthermore, MDCT's are often discussed with the radiologist which could possibly improve the interobserver reliability. Radiographic evaluation was not hampered by these limitations and so may have more accurately mimicked real life situations. Finally, in a reliability study, it is ideal to have a homogeneous group of cases to investigate. In our study, we had both patients with and without fracture fixation, which led to some heterogeneity. The observation with MDCT could be favorable for the group with fixation because evaluation with plain radiographs could be more impaired by metal hardware.

Previous literature

No previous studies have directly compared the reliability of plain radiography and CT in the evaluation of non-union. The majority of the reliability studies focus on plain radiography alone in the evaluation of non-union. In general, they found a fair to moderate inter-observer reliability, which is consistent with our results (10,11). However, agreement improved to almost excellent agreement when bone specific radiographic union scores for tibia and hip fractures were used to determine union (10,11).

The use of CT scanning to investigate the inter-observer reliability for the evaluation of non-union has mainly focused on scaphoid fractures (12-14). Comparable results in all three studies were found with moderate to good agreement. Bhattacharyya et al. examined the evaluation of tibial fracture union by CT scan and determined an ICC of 0.89, which even indicates excellent agreement (15). These studies indicate that using CT scan has high inter-observer reliability, better than the inter-observer reliability of plain radiography described by previous literature. However, these studies did not compare inter-observer reliability of CT and plain radiography for the same cases, so it is difficult to draw conclusions from these studies.

The influence of a CT scan in the evaluation of non-union and its management has not been clearly determined yet. The only available study that examined if the addition of CT with MPR to plain radiograph would aid in the evaluation of non-union was done in 1988 (2). They found that additional CT scan changed the evaluation of non-union and led to an increase in surgical treatment in a series of cases in which previous radiographs were equivocal. Although, this study was limited to two observers and 19 cases, and there was no quantification of the degree of non-union. Comparable results are seen in our study as "undecided" management led four times more to operative treatment than nonoperative treatment. Koller et al. used the CT scan as a reference standard and stated that plain radiographs were not accurate to determine non-union in odontoid fractures (16). This is a very specific spine fracture, which is not comparable with appendicular fractures we included in our study.

Implications for future research

Ideally, data collection needs to be conducted

prospectively. Our findings may help in sample size calculations for future studies. The results of the present study may serve for hypothesis generation in future research.

Further evaluation must be done to determine the necessity of MDCT with MPR in addition to X-Ray, as our results indicate that there was no tangible difference made in evaluation of non-union, but clinical management changed in equivocal cases with the addition of MDCT with MPR imaging to the standard X-Ray. Furthermore, we should investigate the actual benefits of changing the management when using an additional MDCT using patient specific outcome scores.

In this study, the interobserver reliability of MDCT scan is not greater than conventional radiographs for determining non-union. However, a MDCT scan did lead to a more invasive approach especially in equivocal cases. This should be taken into consideration by clinicians before implementing the common practice of requisitioning CT scans to aid in X-Ray evaluation. Future sufficiently powered prospective studies are needed.

Ydo V. Kleinlugtenbelt MD

Department of Orthopaedic and Trauma Surgery, Joint Research Onze Lieve Vrouwe Gasthuis, Amsterdam, the Netherlands.

Division of Orthopaedic Surgery, McMaster University, Canada.

Department of Orthopaedic and Trauma Surgery, Deventer ziekenhuis, Deventer, the Netherlands

Vanessa A.B. Scholtes PhD

Christian Amaechi MD

Rudolf W. Poolman MD PhD

Department of Orthopaedic and Trauma Surgery, Joint Research Onze Lieve Vrouwe Gasthuis, Amsterdam, the Netherlands

Jay Toor MD

Mohit Bhandari MD PhD

Division of Orthopaedic Surgery, McMaster University, Canada

Mario Maas, MD PhD

Department of Radiology, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands

Peter Kloen MD PhD

Department of Orthopaedic and Trauma Surgery, Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands

References

1. Krestan CR, Noske H, Vasilevska V, Weber M, Schueller G, Imhof H, et al. MDCT versus digital radiography in the evaluation of bone healing in orthopedic patients. *AJR Am J Roentgenol.* 2006; 186(6):1754-60.
2. Kuhlman JE, Fishman EK, Magid D, Scott WW Jr, Brooker AF, Siegelman SS. Fracture nonunion: CT assessment with multiplanar reconstruction. *Radiology.* 1988; 167(2):483-8.
3. Bhandari M, Guyatt GH, Swiontkowski MF, Tornetta P 3rd, Sprague S, Schemitsch EH. A lack of consensus in the assessment of fracture healing among orthopaedic surgeons. *J Orthop Trauma.* 2002; 16(8):562-6.
4. Sodickson A, Baeyens PF, Andriole KP, Prevedello LM, Nawfel RD, Hanson R, et al. Recurrent CT, cumulative radiation exposure, and associated radiation-induced cancer risks from CT of adults. *Radiology.* 2009; 251(1):175-84.
5. Brenner DJ, Hall EJ. Computed tomography--an increasing source of radiation exposure. *N Engl J Med.* 2007; 357(22):2277-84.
6. Ohashi K, El-Khoury GY, Bennett DL, Restrepo JM, Berbaum KS. Orthopedic hardware complications diagnosed with multi-detector row CT. *Radiology.* 2005; 237(2):570-7.
7. Karanicolas PJ, Bhandari M, Kreder H, Moroni A, Richardson M, Walter SD, et al. Evaluating agreement: conducting a reliability study. *J Bone Joint Surg Am.* 2009; 91(Suppl 3):99-106.
8. Giraudeau B, Mary JY. Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Stat Med.* 2001; 20(21):3205-14.
9. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess.* 1994; 6(4):284-290.
10. Bhandari M, Chiavaras MM, Parasu N, Choudur H, Ayeni O, Chakraverty R, et al. Radiographic union score for hip substantially improves agreement between surgeons and radiologists. *BMC Musculoskelet Disord.* 2013; 14(1):1.
11. Whelan DB, Bhandari M, Stephen D, Kreder H, McKee MD, Zdero R, et al. Development of the radiographic union score for tibial fractures for the assessment of tibial fracture healing after intramedullary fixation. *J Trauma.* 2010; 68(3):629-32.
12. Buijze GA, Wijffels MM, Guitton TG, Grewal R, van Dijk CN, Ring D, et al. Interobserver reliability of computed tomography to diagnose scaphoid waist fracture union. *J Hand Surg Am.* 2012; 7(2):250-4.
13. Grewal R, Frakash U, Osman S, McMurtry RY. A quantitative definition of scaphoid union: determining the inter-rater reliability of two techniques. *J Orthop Surg Res.* 2013; 8(1):28-33.
14. Hannemann PF, Brouwers L, van der Zee D, Stadler A, Gottgens KW, Weijers R, et al. Multiplanar reconstruction computed tomography for diagnosis of scaphoid waist fracture union: a prospective cohort analysis of accuracy and precision. *Skeletal Radiol.* 2013; 42(10):1377-82.
15. Bhattacharyya T, Bouchard KA, Phadke A, Meigs JB, Kassarian A, Salamipour H. The accuracy of computed tomography for the diagnosis of tibial nonunion. *J Bone Joint Surg Am.* 2006; 88(4):692-7.
16. Koller H, Kolb K, Zenner J, Reynolds J, Dvorak M, Acosta F, et al. Study on accuracy and interobserver reliability of the assessment of odontoid fracture union using plain radiographs or CT scans. *Eur Spine J.* 2009; 18(11):1659-68.

* **COAST Group:** Bob Zura, Sumito Kawamura, Brett D. Crist, David Ring, Gregory J. Della Rocca, Robert J. Feibel, Kyle J. Jeray, Prof. Richard Page, Paul E. Levin, B McCormack, Rodrigo Pesantez, Charalampos Zalavras, M Prayson, Peter R.G. Brink, Andrew H. Schmidt, Christopher Allan, Jeffrey A Greenberg, A Barquet.