European Journal of Radiology 54 (2005) 344–351

EJR
EUROPEAN JOURNAL OF RADIOLOGY

www.elsevier.com/locate/ejrad

Interrater reliability of sonographic examinations of orbital fractures

Jank Siegfried[a,*], Deibl Martina[b], Strobl Heinrich[a], Oberrauch Andreas[a], Nicasi Alessandro[a], Missmann Martin[a], Bodner Gerd[c]

[a] *Department of Oral and Maxillofacial Surgery, University of Innsbruck, Maximilianstr. 10, A-6020 Innsbruck, Australia*
[b] *Department of Biostatistics, University of Innsbruck, Austria*
[c] *Department of Radiology, University of Innsbruck, Austria*

Received 7 June 2004; received in revised form 12 June 2004; accepted 14 July 2004

Abstract

Objective: The purpose of this study is to determine whether there are statistically significant variations among different observers when examining fractures of the orbital walls.
Material and methods: From December 2003 to April 2004, 28 patients with clinically suspected orbital fractures were examined by ultrasound prospectively. The US images of the infra-orbital margins, the orbital floors, the medial and lateral orbital walls of each patient were reexamined by two independent investigators.
Results: Computed tomography revealed fractures of the orbital floor in 28 out of 31 patients (90.3%). The infra-orbital margins showed fractures of 14 of 31 patients (45.2%). The ultrasound examinations of the orbits by the three examiners presented satisfactory correlation regarding sensitivity and specificity. There were no significant differences between investigators. There was good agreement among the ultrasound examiners regarding the infra-orbital margins. This was not the case for the orbital floors.
Conclusions: If there are clear cut clinical findings ultrasound examination could represent an alternative to computed tomography. If the clinical findings were indeterminate, computed tomography was essential as implicated by this study. Accordingly, further evaluation of ultrasound examinations of fractures of the orbital margins and floors are necessary.

Keywords: Ultrasound; Orbit; Blow-out-fractures

1. Introduction

Fractures and dislocations are frequent consequences of traumas of the maxillo-facial area. The orbital floor is involved in nearly all cases if mid-facial injury has been sustained. Isolated fractures of the orbital floor, which are the consequences of direct injury to the eye are referred to as "blow-out" fractures. Computerized tomography (CT) in addition to conventional radiological examination still remains the "gold standard" in the literature [1–4]. The disadvantages of conventional radiological evaluations are that anatomical structures overlap [5–7]. This renders the interpretation difficult. The disadvantages of CT are the higher cost and radiation exposure of the eye. The ultrasound (US) examination of fractures of the orbital floor using a linear transducer did not achieve satisfactory diagnostic information, because the transducer did not fit the orbital margin sufficiently close [8–13]. The use of a curved array transducer produced images of satisfactory sensitivity and specificity [14–17]. These studies, however, were conducted by only one examiner. To provide information regarding the reliability of a specific method it is necessary to compare the findings of several examiners.

2. Objective

The aim of this study was to determine whether there were statistical differences among several examiners when evaluating fractures of the orbital walls by ultrasound (US).

* Corresponding author. Tel.: +43 512 504 24390, +43 699 11107628 (M).
E-mail address: siegfried.jank@gmx.de (J. Siegfried).

0720-048X/$ – see front matter © 2004 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ejrad.2004.07.009

3. Materials and methods

Beginning December 2003 through April 2004, 28 patients with the clinical impression of a fracture of the orbital floor and/or the infra-orbital margin were prospectively examined. Patients were included if one of the following clinical criteria were fulfilled:

1. hematoma,
2. diplopia,
3. reduced mobility of the eye,
4. hyposphagma and
5. exophthalmos or Enophthalmos.

With each patient an axial thin-layer CT with coronal reconstruction was done (GE® Multislice). The following CT parameters were used: Helical CT with 1.25 mm slices, reconstruction slice thickness 1.2 mm (MIP-Modus), pitch: 3.75 high quality (standard or bone). Only orbits with a clinically suspected fracture were investigated. The findings were evaluated by an experienced radiologist according to the following criteria:

1. fracture of the infra-orbital margin,
2. fracture of the orbital floor,
3. fracture of the medial orbital wall,
4. fracture of the lateral orbital wall.

The CT findings were accepted as reference. Subsequently each patient was sonographically evaluated by an experienced examiner with a curved array transducer (Picker CS03®/7.5 MHz) (Figs. 1 and 2). This US examination of the orbits with clinically suspected fractures was done without knowledge of the CT findings. Only those orbits which had clinically suspected fractures were included in this study. The US images were reexamined independently by two investigators without prior knowledge of the CT findings. The US investigation and re-investigation of the images included the evaluation of the bony structures of the orbit, the soft-tissue

Fig. 1. Ultrasound investigation of the lateral orbital wall. The transducer is overlying the lateral orbital rim.

Fig. 2. Ultrasound investigation of the medial orbital wall. The transducer is overlying the medial orbital rim.

was not evaluated. Sensitivity, specificity, accuracy and positive predictive value (PPV) as well as negative predictive values (NPV) were determined with the aid of SPSS (SPSS® Inc.). To compare the individual investigators a chi-square test was performed. The result was interpreted as statistically significant if P had a value of <0.05. The interrater reliability/interobserver variation was calculated using the lambda-coefficient (λ) [18]. The value of λ ranges from 0 to 1, where 0 stands for no reliability and 1 for maximal reliability. λ-values below 0.4 represent poor reliability. λ-Values between 0.4 and 0.75 present good reliability. λ-Values above 0.75 are considered very good reliability [18].

4. Results

4.1. CT-findings (reference)

Twenty-two patients (78.6%) were male, six patients (21.4%) female. Maximum and minimum ages were 85 and 11 years, respectively. The mean age was 36 years. Three patients revealed clinical evidence of fractures in both orbits. Consequently a total of 31 orbits presented clinical evidence of a fracture and were examined sonographicly. CT revealed fractures of the floor in 28 of 31 orbits. In three orbits no fracture was demonstrated. In 14 of 31 orbits (45.2%) infra-orbital margin fractures were detected. In 17 cases no fracture of the infra-orbital margins could be found. In nine of 31 orbits (29.0%) fractures of the medial orbital walls were noted. In 15 orbits (48.5%) the lateral walls were fractured.

4.2. Ultrasound findings of the infraorbital margin (Figs. 3 and 4)

US examination (initial examination) of the infra-orbital margins showed 12 true-positive, 16 true-negative, one false positive and two false negative diagnoses (Table 1). This

Table 1
Results of the US examinations of the orbit on patients with a tentative diagnosis of a fracture by the original examiner and the two re-investigators. (Reference CT)

Localisation	Result	Examiner 1 (initial examination)	Re-investigator 1 (secondary reading)	Re-investigator 2 (secondary reading)
Infra-orbital margin	True-positive	12	11	12
	True-negative	16	16	16
	False-positive	1	2	2
	False-negative	2	2	1
Orbital floor	True-positive	26	26	26
	True-negative	3	2	3
	False-positive	0	1	0
	False-negative	2	2	2
Medial orbital wall	True-positive	8	9	9
	True-negative	20	19	19
	False-positive	2	2	1
	False-negative	1	1	2
Lateral orbital wall	True-positive	14	15	12
	True-negative	14	13	16
	False-positive	2	2	1
	False-negative	1	1	2

presents a sensitivity of 86%, a specificity of 94% and accuracy of 90%. PPV and NPV are 92 and 89%, respectively (Table 2, Figs. 3–10).

The findings of re-investigators 1 and 2 (secondary reading) are shown in Tables 1 and 2.

Comparing the results of the US examinations of the infra-orbital margin by the chi-square test one obtains between examiner and re-investigators 1 and 2 P-values of 0.688 and 1.0, respectively (Table 3). Accordingly there are no significant differences between the original examiner and the two re-investigators. Comparing the findings of the two re-investigators yields a P-value of 0.688 (Table 3). Accordingly there is no significant difference between the two re-investigators. Considering the reliability of findings with the aid of λ-values (Table 4) one obtains a λ value of 0.714 when comparing the CT-findings with the US-findings.

Accordingly there is good and very good reliability. Comparing the original examiner with the two re-investigators yields a λ-value of 0.636 (Table 4). This represents good reliability. Evaluation of the findings of the two re-investigators yields a λ-value of 0.625. This likewise is good reliability.

Table 2
Statistical results of the examiner and the two re-investigators of US examination of orbits with clinically suspecter fractures

Localisation		Examiner 1 (initial examination)	Re-investigator 1 (secondary reading)	Re-investigator 2 (secondary reading)
Infra-orbital rim	Sensitivity	0.86	0.85	0.92
	Specificity	0.94	0.89	0.89
	Accuracy	0.90	0.87	0.90
	PPV	0.92	0.85	0.86
	NPV	0.89	0.89	0.94
Orbital floor	Sensitivity	0.90	0.93	0.87
	Specificity	1.00	0.67	1.00
	Accuracy	0.94	0.90	0.90
	PPV	1.00	0.93	0.86
	NPV	0.60	0.50	0.94
Medial orbital wall	Sensitivity	0.89	0.90	0.82
	Specificity	0.91	0.90	0.95
	Accuracy	0.90	0.90	0.90
	PPV	0.80	0.82	0.90
	NPV	0.95	0.90	0.85
Lateral orbital wall	Sensitivity	0.93	0.94	0.86
	Specificity	0.88	0.87	0.94
	Accuracy	0.90	0.90	0.90
	PPV	0.88	0.88	0.92
	NPV	0.93	0.93	0.89

Reference: CT.

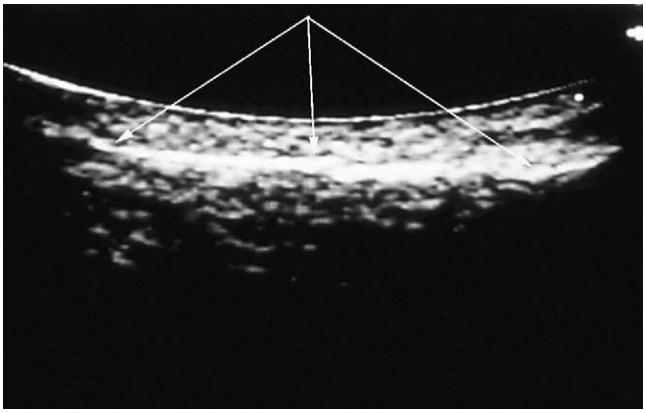

Fig. 3. Ultrasound image of the normal anatomy of the infraorbital rim. The arrows show the convex structure of the infraorbital rim.

4.3. Ultrasound findings of the orbital floor (Figs. 5 and 6)

The US examination (initial examination) of the orbital floors yielded 26 true positive, three true negative, one false negative and no false positive findings (Table 1). This calculates to a sensitivity of 90%, a specificity of 100% and accuracy of 94%. PPV and NPV were 100 and 60%, respectively (Table 2). The results of re-investigatiors 1 and 2 (secondary reading) are shown in Tables 1 and 2.

Analysis of the results of US examination of the infraorbital margin by the chi-square test comparing the examiner and re-investigators 1 and 2 revealed 2 P-values of 0.641 and 1.0, respectively (Table 3). This shows that there is no significant difference between examiner and re-investigators. If one compares the findings of the two re-investigators one obtains a P-value of 0.649 (Table 3), which shows that there is no significant difference of findings between the two re-investigators.

Comparing the CT-finding with the US findings (reliability of findings (λ-values) (Table 4) shows λ-values of 0.500 and 0.143, respectively. This represents poor reliability for re-investigator 2.

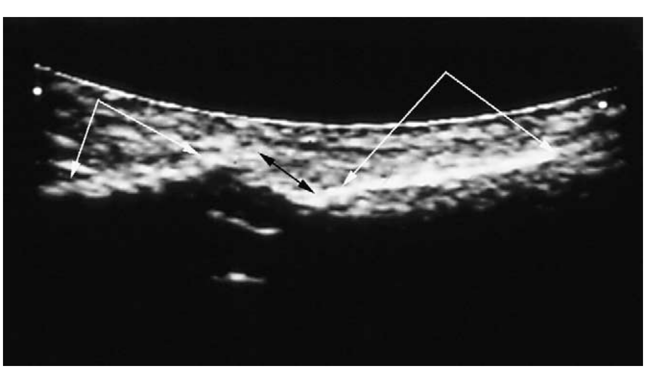

Fig. 4. Ultrasound image of a fracture of the infraorbital rim. The black double arrow shows the fractured area. The white arrows mark the unfractured region.

Fig. 5. Ultrasound image of the normal anatomy of the orbital floor. The double arrow shows the globe. The three arrows show the convex structure of the orbital floor.

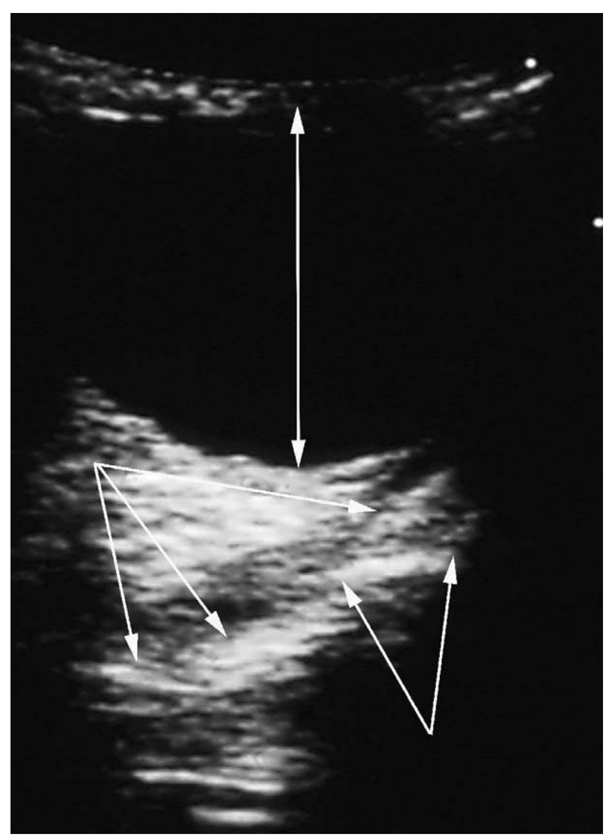

Fig. 6. Ultrasound image of an orbital floor fracture. The double arrow shows the eyeball. The lower arrows show on the dislocated fragment. The upper arrows show the unfractured area of the orbital floor.

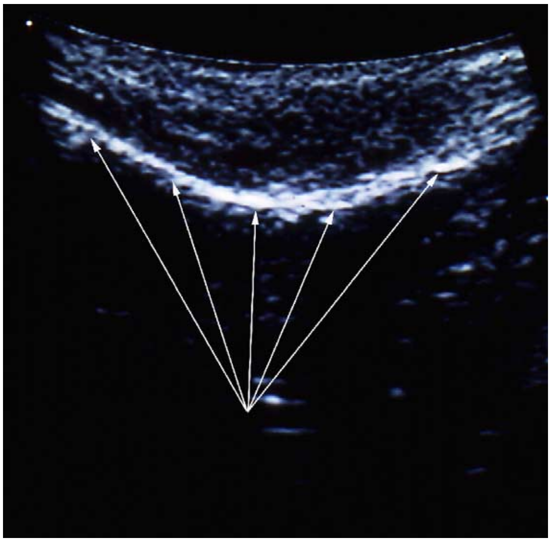

Fig. 7. Ultrasound image of the normal anatomy of the medial orbital wall. The arrows show the convex structure of the medial orbital wall. Compared to the lateral orbital wall the medial wall shows more convexity.

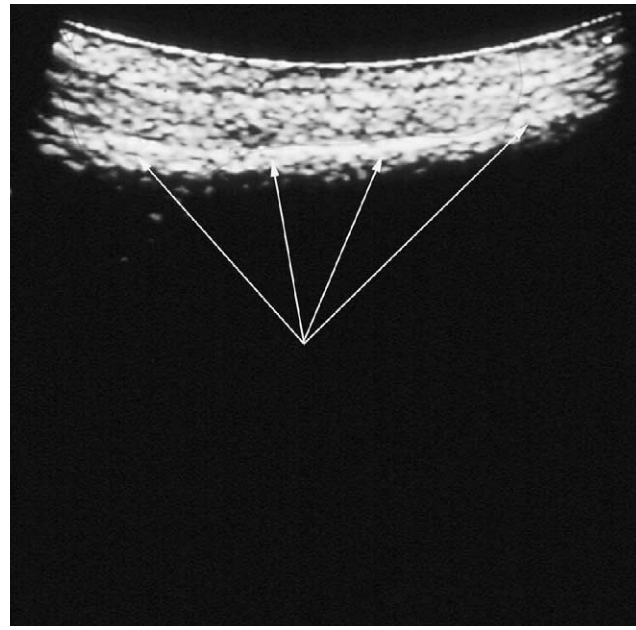

Fig. 9. Ultrasound image of the normal anatomy of the lateral orbital wall. The arrows show the convex structure of the lateral orbital wall.

Comparing the examiner with the two re-investigators one obtains λ-values of 0.200 and 0.091, respectively (Table 4). This represents poor reliability.

Comparing the findings of the two re-investigators show a λ-value of 0.091, indicating poor reliability.

4.4. *Ultrasound findings of the medial orbital wall (Figs. 7 and 8)*

Sonographic exploration (initial examination) of the medial orbital wall by investigator 1 revealed eight true-positive, 20 true negative, two false positive and one false negative findings (Table 1). This calculates to a sensitivity of 89%, specificity of 91% and accuracy of 90%. PPV and NPV are 80 and 95%, respectively (Table 2). The results of re-investigators 1 and 2 (secondary reading) are shown in Tables 1 and 2.

Comparison of US findings of the examiner and the two re-investigators by chi-square test yielded a P-value of 1.000, indicating no significant difference (Table 3). Calculating the reliability of findings (Table 4) by CT and US showed λ-values of 0.739 and 0.714, respectively. This represents good reliability. Comparing the two US findings yielded a P-value of 0.571 and 0.583, respectively. This represents good reliability (Table 4).

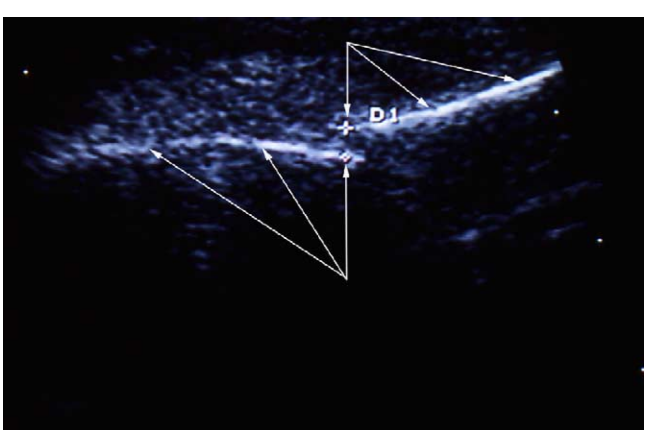

Fig. 8. Ultrasound image of a medial orbital wall fracture. The dislocated fracture is marked with a cross. The upper and lower arrows show the unfractured area of the medial orbital wall.

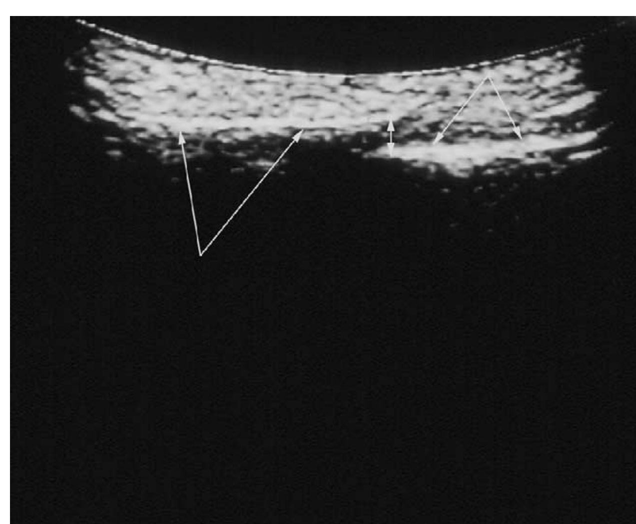

Fig. 10. Ultrasound image of a lateral orbital wall fracture. The double arrow shows the dislocated fracture. The upper and lower arrows show the unfractured area of the lateral orbital wall.

Table 3
Statistical comparison (P-values) the results of the examiner and the two re-investigators of US examination of the orbit with the aid of Chi-square test

Localisation		Re-investigator 1 (secondary reading)	Re-investigator 2 (secondary reading)
Infra-orbital rim	Examiner 1	0.688	1.000
	Re-investigator 1		0.688
Orbital floor	Examiner 1	0.641	1.000
	Re-investigator 1		0.649
Medial orbital wall	Examiner 1	1.000	1.000
	Re-investigator 1		1.000
Lateral orbital wall	Examiner 1	1.000	1.000
	Re-investigator 1		1.000

4.5. Ultrasound findings of the lateral orbital wall (Figs. 9 and 10)

Regarding the US examination (initial examination) of the lateral orbital wall investigator 1 achieved 14 true positive, 14 true negative, two false positive and one false negative finding (Table 1). This represents sensitivity of 93%, specificity of 87% and accuracy of 90%. PPV and NPV are 88 and 93%, respectively (Table 2). The findings of the re-investigations 1 and 2 (secondary reading) are shown in Tables 1 and 2.

Comparison of these results with chi-square test (Table 3) yields P-values of 1.000, which indicate no significant difference between examiner and re-investigators. The λ-values of the comparison of CT and US findings are 0.643, 0.778 and 0.765 respectively (Table 4), indicating good and very good reliability. The λ-values comparing the two US findings are 0.613 and 0.765, respectively. This represent likewise good to very good reliability.

5. Discussion

Earlier reports of sonographic examinations of the orbit did not differentiate among the orbital walls. They yielded unsatisfactory results because of poor fitting of the linear transducer to the curved orbital margins [8–13]. More recent investigations, using a curved array (sector) transducer [14–17], achieved sensitivities and specificities of 79–94 and 90–100%, respectively. The results of this investigation and our investigators showed that with the exception of NPV all statistical results were above 85%. These values are acceptable and correlate with published work. The results of the two re-investigators compare poorly with the original examiner. This is particularly due to the low sensitivity of 67% of re-investigator 1. Re-investigator 2, however, achieved satisfactory results with threshold value reaching 85%, except for one low NPV. It has to be discussed critically that only patients with clinically suspected orbital fractures have been included into the study, but the results show that there are many patients with true-negative findings. This is a sign, that not every clinically suspected fracture was confirmed by CT. It has to be remembered that the original examiner has the advantage, that the transducer can be moved and adjusted for more precise imaging. The re-investigators were confined to the static images. On the one hand, the re-investigators performing the secondary readings have less information and it is much more difficult to perform a secondary reading than doing the initial examination. On the other hand the secondary readings are not influenced by clinical findings. Of course, the definite diagnosis should be a combination of clinical and imaging diagnosis. It should further be noted that during this investigation relatively few false positive and false negative findings occurred and that even one single false diagnosis influenced the statistical results. Further, the initial US examination is not truly blinded, because the investigator can see clinical signs like hyposphagma or hematoma. Only the evaluation of the secondary readings could be defined as truly blinded, because no information about the

Table 4
Reliability of results (λ-values) of US examination of patients with clinically suspected orbital fractures

Localisation		Examiner 1 (initial examination)	Re-investigator 1 (secondary reading)	Re-investigator 2 (secondary reading)
Infra-orbital rim	CT	0.714	0.778	0.778
	Examiner 1	–	0.636	0.636
	Re-investigator 1	–	–	0.625
Orbital floor	CT	0.500	0.143	0.500
	Examiner 1	–	0.200	0.091
	Re-investigator 1	–	–	0.091
Medial orbital wall	CT	0.739	0.739	0.714
	Examiner 1	–	0.571	0.583
	Re-investigator 1	–	–	0.583
Lateral orbital wall	CT	0.724	0.643	0.778
	Examiner 1	–	0.765	0.613
	Re-investigator 1	–	–	0.613

clinical findings and/or CT results were given to the re-investigators.

This study shows that with uncertain clinical findings CT is still absolutely required. Further prospective clinical studies are needed to eliminate the risk of unnecessary surgery because of a false positive orbital fracture diagnosis. Comparison of the examiner with the re-investigators and between the re-investigators by chi-square analysis shows that the sonographic findings do not significantly vary. Chi-square analysis indicates only, that there are no differences of findings when reviewing all findings. It does not indicate, whether the mistakes occur at the same type of patients.

Such statements can be made with the aid of λ-values. There is good to very good reliability regarding the medial and lateral orbital walls. Likewise there is good to very good interobserver reliability regarding the orbital floor. This indicates that mistakes depend on the type of patients.

The sonographic examination of the orbital floor is a different matter.

Comparison among examiner and re-investigators show poor interobserver reliability. In the current study two secondary readings of images were compared between each other and the initial examination. The findings of the initial clinical examination are very important and the comparison between the re-investigators should be only interpreted as a sign for the reproducibility and possibility of re-evaluation of the method. Further investigations with two initial examiners have to be performed to get more information about the method and interobserver reliability.

This permits interpretation that mistakes during US examinations of the orbital floor do not depend on the type of patient but on the examiner. Each individual examiner or re-investigator showed acceptable values, but they do not occur with the same patient. Better calibration of examiners among each other is required regarding the examination of the floor of the orbit. Especially with severe orbital trauma or injuries to the skull and central nervous system CT remains the standard of care, because intra-cranial injuries and compressions of the optic nerve cannot be adequately evaluated by US.

Furthermore, it has to be noted that orbital floor fractures and combinations of orbital floor fractures and fractures of the medial orbital wall are the most common fractures in the orbital region. Therefore, the results of this study concerning the orbital floor, the infraorbital margin and the medial orbital wall are clinically interesting. Lateral orbital wall fractures are usually severe and clinically obvious. In most of the cases with a fracture of the lateral orbital wall, a CT investigation is absolutely necessary because of the severity of the injury. Therefore, an US investigation of lateral orbital wall fractures will be unnecessary, because of the necessity of a CT. For the preoperative planning, an imaging diagnosis is absolutely required. In the case of uncomplicated blow-out fractures or zygoma fractures a CT investigation is not necessary, because conventional radiography contains all the information which is required for the preoperative planning, while in more severe fractures a CT investigation has to be done, because the orbital volume has to be determined preoperatively, which could be only performed by CT.

The question remains whether US is an alternative to CT in the case of "blow-out" fractures or uncomplicated zygomatic/orbital floor fractures? Regarding the indication for surgical intervention the clinical examination is particularly important [19,20]. If there is a clear cut clinical picture US evaluation could represent an alternative to CT or conventional radiography. Finally, it can be stated that US examination of the orbit in patients with clinical evidence of a fracture of the infra-orbital margin and/or the medial and/or lateral orbital walls has good to very good reliability. The reliability of the evaluation of the orbital floor remains unsatisfactory, while sensitivity, specificity and accuracy remain acceptable. In this regard further studies and improvement of technologies are needed to provide acceptable results using US methodology.

References

[1] Ball Jr JB. Direct oblique sagittal CT of orbital wall fractures. AJR Am J Roentgenol 1987;148:601–8.

[2] Benateau H, Chevalier E, Hamon M, et al. The three dimensiononal spiral scanner and volume rendering technique: importance in craniofacial traumatology and reconstructive surgery. Rev Stomatol Chir Maxillofac 2002;103:233–8.

[3] Daly BD, Russell JL, Davidson MJ, Lamb JT. Thin section computed tomography in the evaluation of nasoethmoidal trauma. Clin Radiol 1990;41:272–5.

[4] Imhof H, Czerny C, Dirisamer A. Head and neck imaging with MDCT. Eur J Radiol 2003;45:23–31.

[5] Berado N, Leban SG, Williams FA. A comparison of radiographic treatment for evaluation of the orbit. J Oral Maxillofac Surg 1998;46:844–9.

[6] Dorbisz H, Voegeli E, Hardt N. Conventional radiology and computed tomography in facial fractures. Rontgenblatter 1993;36:428–33.

[7] Iinuma T, Ishio K, Yoshinami H, Kuriyama J, Hirota Y. Orbital wall fractures: a comparison of computed tomography and conventional views. Nippon Jibiinkoka Kaiho 1993;96:175–81.

[8] Forrest CR, Lata AC, Marcuzzi DW, Bailey MH. The role of orbital ultrasound in the diagnosis of orbital fractures. Plast Reconstr Surg 1993;92:28–34.

[9] Jenkins CM, Thuau H. Ultrasound imaging assessment of fractures of the orbital floor. Clin Radiol 1997;52:708–11.

[10] Klinger M, Danter J, Siegert R. Ultrasound diagnosis of orbital floor fractures: an alternative to computerized tomography. Laringorhinootologie 1996;75:242–4.

[11] Lata AC, Marcuzzi DW, Forrest CR. Comparison of real-time ultrasonography and coronal computed tomography in the diagnosis of orbital fractures. Can Assoc Radiol J 1993;44:371–6.

[12] McCann PJ, Brocklebank LM, Ayoub AF. Assessment of zygomatico-orbital complex fractures using ultrasonography. Br J Oral Maxillofac Surg 2000;38:525–9.

[13] Takashi H, Manders E, Nagamoto K, Saggers G. Ultrasonic observation of facial bone fractures. J Oral Maxillofac Surg 1996;54:776–8.

[14] Jank S, Strobl H, Emshoff R, et al. Bildgebende Diagnostik von medialen und lateralen Orbitawandfrakturen. Ultraschall versus CT. Mund Kiefer Gesichtschir 2003;4:208–13.

[15] Jank S, Emshoff R, Etzelsdorfer M, Strobl H, Nicasi A, Norer B. Effectiveness of ultrasonography in determining medial and lateral orbital wall fractures with a curved array transducer. J Oral Maxillofac Surg 2004;62:451–5.

[16] Jank S, Emshoff R, Etzelsdorfer M, Strobl H, Nicasi A, Norer B. The diagnostic value of ultrasonography in the detection of orbital floor fractures with a curved array transducer. Int J Oral Maxillofac Surg 2004;33:13–8.
[17] Jank S, Emshoff R, Etzelsdorfer M, Strobl H, Nicasi A, Norer B. Ultrasound versus computed tomography in the imaging of orbital floor fractures. J Oral Maxillofac Surg 2004;62:150–4.
[18] Fleiss J. Reliability of measurements. In: The Design and Analysis of Clinical Experiments. New York: John Wiley & Sons; 1986.
[19] Jank S, Schuchter B, Strobl H, et al. Post-traumatic ocular motility in orbital floor fractures. Mund Kiefer Gesichtschir 2003;7:19–24.
[20] Jank S, Schuchter B, Emshoff R, et al. Clinical signs of orbital wall fractures as a function of anatomic location. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2003;96:149–53.