The interobserver reliability of the diagnosis and classification of scaphoid fractures using high-resolution peripheral quantitative CT

Citation for published version (APA):

Document status and date:
Published: 01/04/2020

DOI:

Document Version:
Publisher's PDF, also known as Version of record

Document license:
Taverne

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• Users may not further distribute the material or use it for any profit-making activity or commercial gain
• Users may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:
repository@maastrichtuniversity.nl
providing details and we will investigate your claim.

Download date: 17 Sep. 2023
The interobserver reliability of the diagnosis and classification of scaphoid fractures using high-resolution peripheral quantitative CT

Aims
Besides conventional radiographs, the use of MRI, CT, and bone scintigraphy is frequent in the diagnosis of a fracture of the scaphoid. However, which techniques give the best results remain unknown. The investigation of a new imaging technique initially requires an analysis of its precision. The primary aim of this study was to investigate the interobserver agreement of high-resolution peripheral quantitative CT (HR-pQCT) in the diagnosis of a scaphoid fracture. A secondary aim was to investigate the interobserver agreement for the presence of other fractures and for the classification of scaphoid fracture.

Methods
Two radiologists and two orthopaedic trauma surgeons evaluated HR-pQCT scans of 31 patients with a clinically-suspected scaphoid fracture. The observers were asked to determine the presence of a scaphoid or other fracture and to classify the scaphoid fracture based on the Herbert classification system. Fleiss kappa statistics were used to calculate the interobserver agreement for the diagnosis of a fracture. Intraclass correlation coefficients (ICCs) were used to assess the agreement for the classification of scaphoid fracture.

Results
A total of nine (29%) scaphoid fractures and 12 (39%) other fractures were diagnosed in 20 patients (65%) using HR-pQCT across the four observers. The interobserver agreement was 91% for the identification of a scaphoid fracture (95% confidence interval (CI) 0.76 to 1.00) and 80% for other fractures (95% CI 0.72 to 0.87). The mean ICC for the classification of a scaphoid fracture in the seven patients diagnosed with scaphoid fracture by all four observers was 73% (95% CI 0.42 to 0.94).

Conclusion
We conclude that the diagnosis of scaphoid and other fractures is reliable when using HR-pQCT in patients with a clinically-suspected fracture.

Cite this article: Bone Joint J 2020;102-B(4):478–484.
Flowchart showing the eligibility criteria and selection of patients with a clinically-suspected scaphoid fracture. *Concurrent fracture (Fx) of the ipsilateral lower/upper arm. ED, Emergency Department.

Diagram of the classification of scaphoid fracture according to Herbert and Fisher.33

with the current clinical diagnostic techniques. The primary aim was to assess the interobserver reliability of HR-pQCT scans of consecutive patients with a clinically-suspected scaphoid fracture using independent observers. A secondary aim was to assess the interobserver agreement for other fractures and for the classification of fractures of the scaphoid.

Methods
This prospective feasibility study was conducted between December 2017 and October 2018. It was approved by the Medical Ethics Committee ("SHOTGUN, Scaphoid fracture diagnosis with HR-pQCT", NL 62476.068.17), and was performed according to the principles of the Declaration of Helsinki and the Medical Research Involving Human Subjects Act (WMO).

All consecutive patients presenting to our emergency department with a suspected scaphoid fracture within one week of injury were screened for eligibility. Pregnant women and patients with a previous ipsilateral scaphoid fracture were excluded. All patients were initially treated with a forearm cast in conformity with current practice, until the time of reassessment and additional imaging. This involved 91 patients who consented and had a HR-pQCT scan of the affected forearm about ten days after presentation, as additional imaging at the time of further routine clinical evaluation. Patients with a scaphoid fracture were followed up to 26 weeks after trauma. An overview of the eligibility criteria and selection of patients is shown in Figure 1.

The sample size was determined according to the requirements described by Donner and Rotondi in 2010. We aimed to achieve an interobserver agreement of $\kappa = 0.80$ according to Landis and Koch between four observers. Based on the expected detection of a fracture using HR-pQCT, we estimated $\pi$ at 0.30, which is slightly higher than the currently reported assessment of the incidence of 0.20. The minimal acceptable limit (lower 95% confidence interval (CI)) was prespecified as substantial agreement expressed by $\kappa = 0.60$, resulting in a sample size of 25 patients who needed to be examined. In order to ensure a sufficiently large cohort, we selected 31 patients using a computer-based random number generator (being one-third of our cohort) out of the 91 eligible patients. All 31 HR-pQCT scans were assessed by four independent observers.

All patients had a HR-pQCT scan using the second-generation HR-pQCT scanner (XtremeCT II; Scanco Medical AG, Zürich, Switzerland). The standard protocol of the distal radius, with an
isotropic voxel size of 0.061 mm, was adapted to three consecutive stacks (30.6 mm) to cover the entire scaphoid.28-30 Thus the surrounding bones, including the proximal carpal row and part of the distal carpal row, were displayed. This resulted in a radiation dose of 15 µSv, which is about 0.6% to 0.5% of the annual background radiation exposure for an individual in The Netherlands and USA.31,32 All scans were conducted with the wrist in a fibreglass cast with a detachable piece around the thumb, which was only used during the HR-pQCT procedure, for added stability. The forearm was placed in an anatomically shaped container to obtain a standardized position. Based on a scout view, the region of interest was determined and a reference line was placed on the rim at the articular surface of the distal radius. Motion-induced degradation of the images was graded according to the manufacturer’s protocol and the method described by Pialat et al.30 Scans with motion artefacts (grade 4 to 5) were repeated once. The HR-pQCT scans were exported in Digital Imaging and Communications in Medicine (DICOM) format. The images were reconstructed into transverse, coronal, and sagittal planes of the wrist. All HR-pQCT images and reconstructions were anonymized and uploaded into a local workstation.

The independent observers were two musculoskeletal radiologists (DL, SS) and two hand and wrist trauma surgeons (HJ, SK). They were asked to determine the presence of a scaphoid or other fracture of the distal radius or (meta)carpal, and to classify the scaphoid fracture, if present, according to the Herbert or Fisher classification system33 (Figure 2). They were aware of the fact that all patients were suspected of having a scaphoid fracture clinically, were blinded to the findings on conventional radiographs and CT, and were not aware of each other’s assessment. They did not have access to any clinical data.

**Statistical analysis.** The distribution of age was tested with Q-Q plots and the Kolmogorov-Smirnov test. Chi-squared and Fisher classification system33

Table I. The characteristics of the selected (n = 31) and non-selected (n = 60) patients.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Selected (n = 31)</th>
<th>Non-selected (n = 60)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age, yrs (IQR)*</td>
<td>48 (29 to 71)</td>
<td>52 (28 to 65)</td>
<td>0.768</td>
</tr>
<tr>
<td>Male sex, n (%)</td>
<td>16 (52)</td>
<td>29 (48)</td>
<td>0.767</td>
</tr>
<tr>
<td>Dominant hand affected, n (%)</td>
<td>14 (46)</td>
<td>32 (53)</td>
<td>0.460</td>
</tr>
</tbody>
</table>

*IQR, interquartile range.

Table II. Results of high-resolution peripheral quantitative CT fracture classification in 31 patients by four observers.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Observer</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>O</td>
<td>-</td>
<td>-</td>
<td>O</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>O</td>
<td>O</td>
<td>-</td>
<td>O</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>-</td>
<td>O</td>
<td>-</td>
<td>O</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>O</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>26</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>27</td>
<td></td>
<td>-</td>
<td>O</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>28</td>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>29</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>31</td>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

*Scaphoid fractures κ = 0.91 (95% confidence interval 0.76 to 1.00).
†Scaphoid and other fractures κ = 0.84 (95% confidence interval 0.73 to 0.94).

**Results**

The 31 patients had a median age of 48 years (interquartile range (IQR) 29 to 71) and 16 (52%) were male. This was similar to the non-selected patients (n = 60) (Table I). We decided to include all scans, irrespective of the presence of motion artefacts, as this is comparable with using an imaging technique in a clinical situation. Moreover, the number of stacks (2/93) with grade 4 motion artefacts (grade 4) was low and there were no grade 5 stacks. Nine patients (29%) were diagnosed with a scaphoid fracture by at least three observers (Table II), resulting in an interobserver agreement for the presence of a scaphoid fracture of κ = 0.91 (95% CI 0.76 to 1.00). Figures 3a and b show an example of HR-pQCT images displaying a scaphoid fracture detected by
The interobserver reliability of the diagnosis and classification of scaphoid fractures

Fig. 3

a) Transverse high-resolution peripheral CT (HR-pQCT) section of a 48-year-old female patient with a scaphoid fracture (arrow) as diagnosed by all observers. b) Sagittal HR-pQCT section of a 48-year-old female patient with a scaphoid fracture (arrow) as diagnosed by all observers.

Fig. 4

a) Transverse high-resolution peripheral CT (HR-pQCT) section of a 18-year-old female patient with a scaphoid fracture (arrow) as diagnosed by three observers. b) Sagittal HR-pQCT section of a 18-year-old female patient with a scaphoid fracture (arrow) as diagnosed by three observers.

Table III. Herbert and Fisher\textsuperscript{33} classification of scaphoid fracture with high-resolution peripheral quantitative CT in seven patients by the four observers.*

<table>
<thead>
<tr>
<th>Patients</th>
<th>Observers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td>1</td>
<td>A1</td>
</tr>
<tr>
<td>2</td>
<td>A1</td>
</tr>
<tr>
<td>3</td>
<td>B1</td>
</tr>
<tr>
<td>4</td>
<td>A2</td>
</tr>
<tr>
<td>5</td>
<td>B1</td>
</tr>
<tr>
<td>6</td>
<td>A1</td>
</tr>
<tr>
<td>7</td>
<td>A1</td>
</tr>
<tr>
<td>Total</td>
<td>A1</td>
</tr>
</tbody>
</table>

*Intraclass correlation coefficient = 0.73 (95% confidence interval 0.42 to 0.94).

All observers. Figures 4a and b show an example of images with a scaphoid fracture detected by three observers. Other fractures such as distal radial and carpal fractures were diagnosed in between seven (23%) and 11 (35%) patients by the observers (Table II). Diagnosing other fractures resulted in an interobserver agreement of $\kappa = 0.80$ (95% CI 0.72 to 0.87).

Seven patients (23%) were diagnosed with a scaphoid fracture by all four observers (Table III). The ICC for the classification of a scaphoid fracture in these seven patients was 0.73 (95% CI 0.42 to 0.94), indicating substantial agreement. Figure 5 shows HR-pQCT images with a scaphoid fracture detected by all observers but classified differently according to the Herbert and Fisher\textsuperscript{33} classification system.

Discussion
We found a 90% agreement between four independent observers for the diagnosis of fracture using HR-pQCT in 31 patients with a clinically-suspected scaphoid fracture. Previous studies assessing interobserver variability in patients with a suspected
Scaphoid fractures have been conducted for various imaging techniques. The κ-values for radiographs between two and six weeks after injury ranged from 0.14 to 0.37, representing slight to fair agreement between observers. These rates reflect the poor sensitivity of conventional radiological assessment of scaphoid fractures. This, together with the large number of false positives, implies that follow-up radiographs cannot be considered as a reference in the diagnosis of scaphoid fractures.

Interobserver agreement for MRI assessment of scaphoid fracture by four and five observers in cohorts of 79 and 64 patients, with a similar percentage of scaphoid fractures among those suspected clinically as in our study, was κ = 0.67 and κ = 0.44 respectively, representing moderate to substantial agreement. Beeres et al, in a study in which bone scans of both wrists of 100 patients with a clinically-suspected scaphoid fracture on at least one side were analyzed by three observers, found substantial agreement, κ = 0.61 to 0.80, for the diagnosis of a scaphoid fracture and for the diagnosis of other fractures. A study addressing the interobserver agreement of CT assessment of four observers in a cohort of 150 patients with a clinically-suspected scaphoid fracture, reported a κ-value of 0.51 (moderate agreement). In a study by Adey et al, with a comparable size of cohort to ours, 30 CT scans were assessed by eight observers. Although they used a CT with lower resolution than Beeres et al, they found a κ-value of 0.66. As a higher number of observers should reduce the sample size of patients needed to assure a specific κ-value, our results with only four observers are promising.

Only limited data are available about the classification of a scaphoid fracture. Beeres et al and de Zwart et al assessed the interobserver variability of the location of a scaphoid fracture (proximal, middle, distal) on MRI and CT for four observers and found κ-values of 0.57 and 0.48 respectively. Interobserver agreement for the classification of a scaphoid fracture by four observers in our study was substantial (κ 0.73, 95% CI 0.42 to 0.94). Although we have more than three groups, the number in each of the five groups is limited, which is an important limitation of our study. A higher number of scans should be assessed to evaluate the reliability of classification of a fracture between several observers.

The rates of agreement in our study exceed those of previous studies, suggesting that using HR-pQCT images may be reliable for the diagnosis and classification of a scaphoid fracture in patients with a clinically-suspected fracture. This might be explained by the considerable experience with other imaging techniques of the observers in our study. Moreover, as resolution increases it is conceivable that the distinction between vascular structures, motion artefacts and fractures will become clearer. Therefore, using conventional CT with higher resolution than the CT scanners used in the studies by Beeres et al and de Zwart et al might also increase interobserver agreement. However, the limited data available in the literature do not prove this hypothesis yet. Although interobserver agreement is only one aspect of diagnosis, it reveals the reliability of a technique. Another limitation of our study is that we did not determine intraobserver reliability. We did not incorporate this in our study as the agreement between different observers is more relevant for this topic and the implementation in clinical practice.

Novel techniques such as HR-pQCT should be incorporated into further studies addressing the accuracy and precision of the assessment of scaphoid fractures. Although no clear definition of a fracture on HR-pQCT is available, we found that diagnosis based on expert opinion appears to be reliable. Assessment in this study was made by both radiologists and orthopaedic trauma surgeons making extrapolation into clinical practice possible. The use of a uniform scanning protocol and DICOM viewers in this study assured that implementation and assessment in a clinical setting is achievable. The HR-pQCT scanner was designed to measure bone density and to quantify the 3D microarchitecture of distal radial and distal tibial bone (both one stack of approximately 1 cm). In order to scan the scaphoid, a protocol was developed comprising three stacks, instead of one, to capture the entire bone. The total procedure time for the scaphoid is about 30 minutes, including positioning the patient and scanning. The processing and analysis takes about 30 minutes and appropriate computational hardware and software are required. This further increases the costs. At present, the HR-pQCT is mainly used in research settings and it is not ready to be used in clinical practice due to logistical considerations including the time and costs involved. The costs of...
HR-pQCT scans are not yet reimbursed by healthcare insurers. We estimate that the costs of a HR-pQCT scan are about three-times higher than those of a routine CT scan. It is clearly likely that standard protocols for scanning the scaphoid will be developed in the near future, and that processing and analysis time will be reduced.

In conclusion, the diagnosis of scaphoid and other fractures using HR-pQCT is reliable in patients with a clinically-suspected fracture. Further research should compare the identification of a scaphoid fracture using HR-pQCT with currently used techniques in order to explore the potential of this promising new technique.

Take home message - This study shows that HR-pQCT, a novel imaging technique, is reliable for scaphoid fracture diagnosis and classification in clinical practice.

References


