# Observer variation for radiography, computed tomography, and magnetic resonance imaging of occult hip fractures

Collin, David; Dunker, Dennis; Gothlin, Jan H.; Geijer, Mats

*Published in:*
Acta Radiologica

*DOI:*
10.1258/ar.2011.110032

2011

# Observer variation for radiography, computed tomography, and magnetic resonance imaging of occult hip fractures

David Collin[1], Dennis Dunker[1], Jan H Göthlin[1] and Mats Geijer[2]

[1]Department of Radiology, Sahlgrenska University Hospital, Mölndal; [2]Center for Medical Imaging and Physiology, Skåne University Hospital, Lund University, Lund, Sweden

Correspondence to: David Collin. Email: david.collin@vgregion.se

## Abstract

**Background:** Conventional radiography is insufficient for diagnosis in a small but not unimportant number of hip fractures, and secondary imaging with computed tomography (CT) or magnetic resonance imaging (MRI) is warranted. There are no convincing observer variation studies performed for conventional radiography or CT in occult fractures, and no large materials for MRI.

**Purpose:** To assess observer variation in radiography, CT and MRI of suspected occult, non-displaced hip fractures, and to evaluate to what extent observer experience or patient age may influence observer performance.

**Material and Methods:** A total of 375 patients after hip trauma where radiography was followed by CT or MRI to evaluate a suspected occult hip fracture were collected retrospectively from two imaging centers. After scoring by three observers with varying degrees of radiologic experience, observer variation was assessed by using linear weighted kappa statistics.

**Results:** For radiography, agreements between the three observers were moderate to substantial for intracapsular fractures, with kappa values in the ranges of 0.56-0.66. Kappa values were substantial for extracapsular fractures, in the ranges of 0.69-0.72. With increasing professional experience, fewer fractures were classified as equivocal at radiography. For CT and MRI, observer agreements were similar and almost perfect, with kappa values in the ranges of 0.85-0.97 and 0.93-0.97.

**Conclusion:** There were almost perfect observer agreements for CT and MRI in diagnosing non-displaced, occult hip fractures. Observer agreements for radiography were moderate to substantial, and observer experience influenced agreement only at radiography.

## Keywords

Hip fracture diagnosis, observer variation, radiography, computed tomography, magnetic resonance imaging, quality assurance

Prompt and accurate diagnosis of hip fractures in the elderly is necessary to minimize complications such as increased co-morbidity and mortality (1). Also, from an ethical point of view it is necessary to efficiently take care of these frail patients. Furthermore, a fast and efficient management of hip fractures is cost-effective.

Even when conventional radiographs (CXR) of traumatic hips are technically good, interpretation may be difficult. The diagnostic competence varies between radiologists and hip fractures are overlooked (2). One obvious reason for this is perception errors. Other factors could be the experience level of the readers, patient age, or image interpretation under stressful conditions in the emergency room or by an on-call radiologist after office hours. The fracture may, however, simply be impossible to detect with CXR. Magnetic resonance imaging (MRI) has been reported as a good secondary examination with good interobserver agreement (2, 3) but no large study has been published. Despite its widespread use in musculoskeletal trauma computed tomography (CT) of hip fracture is not well documented. Perusal of the literature has not revealed any large material on the value of CT as secondary investigation.

The current, retrospective study was performed to assess to what extent interpreter experience and patient age influenced observer variation in diagnosing occult hip fractures with conventional radiography, computed tomography and magnetic resonance imaging.

## Material and Methods

All patients aged 60 years or older with clinically suspected hip fracture after trauma, with normal or equivocal radiography and subsequent CT or MRI, were retrospectively collected from Skåne University Hospital, Lund and Sahlgrenska University Hospital, Mölndal. The material comprises 375 patients, mean age 81 years (range 60 –107), male:female ratio 0.46. After initial radiography, 232 patients had been examined with CT and 170 with MRI during 2006 – 2008. Of these, 27 had been examined with both CT and MRI. All secondary imaging was done within 10 days of radiography, mean less than 24 h. The patients were identified using the radiology information systems (RIS; Sectra Imtec AB, Linköping, Sweden and in-house developed RIS, Sahlgrenska University Hospital, Gothenburg, Sweden) or picture archiving and communication systems (PACS; IDS7, Sectra Imtec AB, Linköping Sweden and Centricity 600, GE Healthcare, Milwaukee, WI, USA).

The selection of patients for further imaging was based on the fact that radiography had been judged equivocal or that radiography had been interpreted as normal despite the clinical findings. The decision to use CT or MRI was made by the radiologist or the orthopedic surgeon, independently or in agreement. The selection of imaging modality outside office hours was always CT as MRI was not available; during office hours either CT or MRI. One center showed a preference for CT, the other for MRI. There was no systematic selection of imaging modality for different patients or suspected fracture types.

All imaging studies were evaluated by three radiologists with varying degree of experience; a resident, a specialist in general radiology and a specialist in musculoskeletal radiology. The observers only had access to information about the patient name, identification number and gender and worked independently of each other. The use of all PACS tools such as zoom, pan, window and level settings was allowed. The specialist in general radiology made a second review on a subset of images after one year for the purpose of intra-observer agreement calculation (radiography n=104, CT n=92, MRI n=104).

At radiography, all fractures were occult or non-displaced, thus belonging to Garden classification either Grade I or Grade II for cervical fractures, and to type 1 of the Evans/Jensen classification for trochanteric fractures. On radiographs, the fractures were categorized as either intracapsular (cervical neck) or extracapsular (lateral cervical neck/basicervical and trochanteric fractures). Separate calculations on interobserver agreements were made between the three observers for the intracapsular and extracapsular fracture types, respectively, where the radiologic findings were categorized as definite (complete or incomplete), equivocal, or no fracture. Intraobserver agreement for CXR, CT and MRI was calculated for one reviewer (DC). Observer agreement was calculated separately for each age decade group to determine whether age related changes such as osteoarthritis with attrition or osteoporosis influenced observer agreement.

Bi-rater kappa statistic with linear weighting was used (4, 5) to evaluate observer agreement. The linear weighted kappa measures the relative concordance between observers. The diagnosis suspicion of fracture was given less statistical weight than definite or no fractures in case of observer disagreement. Kappa ($\kappa$) values $<0$ represent less than chance agreement, 0.01-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and 0.80-0.99 almost perfect agreement (4).

## Results

*Intracapsular fractures*

Radiography

There was a substantial agreement between musculoskeletal specialist and general specialist ($\kappa = 0.66$; Table 1) but only good agreements with overlapping confidence intervals between musculoskeletal specialist and resident as well as between general specialist and resident ($\kappa = 0.56$ and 0.58, respectively; Tables 2 and 3). In patients older than 90 years with intracapsular fractures the agreement was lower than in the other age groups but the difference was not statistically significant (Table 4). The resident reported twice as many equivocal fractures as the two other observers. In summary, observer experience influenced agreement only in radiography of intracapsular fractures, and patient age did not influence any observer agreement.

Intraobserver agreement for the specialist in general radiology was good ($\kappa = 0.60$).

Computed tomography

The agreements between the observers were almost perfect, with $\kappa$ values 0.86 (musculoskeletal specialist - general specialist), 0.87 (musculoskeletal specialist - resident; Table 5), and 0.85 (general specialist – resident). Intraobserver agreement was also almost perfect ($\kappa = 0.94$).

Magnetic resonance imaging

There were almost perfect agreements between the observers, with κ values for musculoskeletal specialist - general specialist 0.95, musculoskeletal specialist - resident 0.97 (Table 6), and for general specialist - resident 0.95.

Intraobserver agreement was also almost perfect (κ = 0.99).

*Extracapsular fractures*

Radiography

The agreements on radiography of all extracapsular fractures were substantial, with κ values for musculoskeletal specialist - general specialist of 0.71, musculoskeletal specialist - resident 0.72, and for general specialist - resident 0.69. When avulsion fractures of the greater trochanter were excluded the agreements were reduced to moderate, with kappa values 0.52, 0.43, and 0.50, respectively.

Intraobserver agreement for extracapsular fractures was substantial (κ = 0.73).

Computed tomography

The agreements on all extracapsular fractures were almost perfect, with κ values for musculoskeletal specialist - general specialist 0.94, for musculoskeletal specialist - resident 0.97, and for general specialist - resident 0.91. When avulsion fractures of the greater trochanter were excluded the agreements were reduced to substantial, with κ values 0.59, 0.67, and 0.76, respectively.

Intraobserver agreement was almost perfect (κ = 0.95).

Magnetic resonance imaging

There were almost perfect agreements with κ values for musculoskeletal specialist - general specialist 0.94, musculoskeletal specialist - resident 0.97, and for general specialist - resident 0.93. Even when avulsion fractures of the greater trochanter were excluded the κ values were still almost perfect, with kappa values 0.94, 0.92, 0.91.

Intraobserver agreement was perfect (κ = 1.00).

## Discussion

To our knowledge, no large studies have previously been published on observer agreement of non-displaced or occult hip fractures for CXR, CT or MRI. A perusal of the literature has revealed no prior studies on observer agreement of CT of hip fracture, let alone occult hip fracture. Near-perfect observer agreement for MRI of suspected hip fracture has been reported previously (2) with κ = 0.847 for the 62 examined patients. As the current study comprises a much larger number of cases the statistics are most likely more robust. Previous studies are on small materials, address specific research questions, or include clearly displaced fractures.

The current study showed almost perfect observer agreements for CT and MRI, while the agreements between the three observers were good to substantial for CXR.

Basicervical fractures are not clearly defined (6) but were for the purposes of the current study classified as extracapsular. Since intracapsular and extracapsular hip fractures undergo different types of surgery it is important to demonstrate possible differences in observer agreement. This was done only for radiography, where observer agreement was less for intracapsular than for extracapsular fractures, indicating greater difficulties in diagnosing cervical than trochanteric fractures.

Avulsion fractures of the tip of the greater trochanter are according to MRI almost always combined with intertrochanteric or pertrochanteric fractures, at least when bone bruise is regarded as diagnostic (7). At radiography, the intertrochanteric part of the extracapsular group was evaluated separately, with differences in agreements between the two subgroups trochanteric fractures with and without avulsions. The observer agreements were reduced from substantial to moderate when the avulsions from the tip of the greater trochanter were excluded, since there were disagreements on the presence of a trochanteric extension.

The resident classified twice as many intracapsular fractures at CXR as equivocal as the other observers, apparently due to lack of experience. Influence of experience could otherwise not be demonstrated with kappa statistics. Its results depend on sample size, weighting, number of categories, prevalence of lesions, and bias (8). The translation of intervals of κ values into the terms such as slight, fair, moderate, substantial, and almost perfect (4) is ’inevitably arbitrary‘ (8). From common experience, it is evident that radiologic experience matters, and that a ’substantial‘ observer agreement is not sufficient to reliably express true fracture diagnosis agreement. In most cases, an experienced senior radiologist has no difficulty in diagnosing or ruling out a hip fracture where a resident may hesitate.

There were almost perfect agreements on intra- and extracapsular fractures on CT with almost no equivocal fractures. When avulsion fractures of the greater trochanter were excluded from the extracapsular fracture group the agreements were lower because of the disagreements in diagnosis of the intertrochanteric extension.

The agreements between the observers on both intra- and extracapsular fractures on MRI were statistically very close to perfect, in agreement with previous reports (2). The slight disagreements in the current study were caused by differences in the interpretation of bone bruise.

The high intraobserver agreement κ values for CT and MRI indicate a high consistency of diagnosis with those modalities. On the other hand, the low kappa values for radiography reasonably are the result of both the inherent difficulties in hip fracture diagnosis with radiography, and also a positive learning curve during the study.

Patients younger than 60 years were excluded since they were too few to constitute a group in which statistically significant results could be anticipated. Hip fractures in younger and middle-aged persons are typically not linked to fall trauma but rather to underlying factors such as congenital malformations or malignancy, or to high-energy trauma.

The limitations of this study lie mainly in its retrospective nature with only the original images or scan sequences available and thus no possibilities to perform reconstructions at the review.  For radiography and MRI this is probably of no great consequence, whereas for CT a prospective study with more standardized acquisition parameters and the possibility for the observers to do their own MPR and 3D reconstructions might have improved the agreements for CT even further.

In conclusion, the present study has demonstrated a robust diagnostic capability with concordant results and almost perfect observer agreement for CT as well as for MRI in the diagnosis of non-displaced hip fractures, with an advantage for MRI. Observer agreements for radiography were moderate to substantial. There was no significant influence from patient age. Observer experience influenced agreement only at radiography.

## References


1. Borgström F, Sobocki P, Ström O, et al. The societal burden of osteoporosis in Sweden. *Bone* 2007;**40**:1602-9
2. Dominguez S, Liu P, Roberts C, et al. Prevalence of traumatic hip and pelvic fractures in patients with suspected hip fracture and negative initial standard radiographs--a study of emergency department patients. *Acad Emerg Med* 2005;**12**:366-9
3. Verbeeten KM, Hermann KL, Hasselqvist M, et al. The advantages of MRI in the detection of occult hip fractures. *Eur Radiol* 2005;**15**:165-9
4. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159-74
5. Lowry R. Kappa as a measure of concordance in categorical sorting. http://faculty.vassar.edu/lowry/kappa.html. Last accessed on November 12, 2010
6. Saarenpää I, Partanen J, Jalovaara P. Basicervical fracture--a rare type of hip fracture. *Arch Orthop Trauma Surg* 2002;**122**:69-72
7. Feldman F, Staron RB. MRI of seemingly isolated greater trochanteric fractures. *Am J Roentgenol* 2004;**183**:323-9
8. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;**85**:257-68

## Tables

**Table 1.**

| **Radiography of intracapsular fractures** | | | | | |
|---|---|---|---|---|---|
| | | Specialist in musculoskeletal radiology | | | |
| | | Fracture | Equivocal | Negative | Total |
| Specialist in general radiology | Fracture | 39 | 11 | 4 | 54 |
| | Equivocal | 7 | 6 | 19 | 32 |
| | Negative | 9 | 13 | 267 | 289 |
| | Total | 55 | 30 | 290 | 375 |
| Linear weighted kappa (95% CI): 0.66 (0.58-0.74) SE 0.04 | | | | | |

There was substantial observer agreement for intracapsular fractures on radiography between the specialist in musculoskeletal radiology and the specialist in general radiology for 375 patients. They were initially reported either normal or equivocal, and CT or MRI followed after radiography due to remaining suspicion of occult hip fracture.

**Table 2.**

| **Radiography of intracapsular fractures** | | | | | |
|---|---|---|---|---|---|
| | | Specialist in musculoskeletal radiology | | | |
| | | Fracture | Equivocal | Negative | Total |
| Resident | Fracture | 28 | 8 | 5 | 41 |
| | Equivocal | 19 | 15 | 42 | 76 |
| | Negative | 8 | 7 | 243 | 258 |
| | Total | 55 | 30 | 290 | 375 |
| Linear weighted kappa (95% CI): 0.56 (0.48-0.64) SE 0.04 | | | | | |

There was moderate observer agreement for intracapsular fractures on radiography between the specialist in musculoskeletal radiology and the resident. The resident categorized more than twice the patients having a suspected fracture (76) than the specialist in musculoskeletal radiology (30). There was complete disagreement, representing false negatives/positives, in 3.5% (5 + 8 of 375 patients).

**Tables 3.**

| **Radiography of intracapsular fractures** | | | | | |
|---|---|---|---|---|---|
| | | Specialist in general radiology | | | |
| | | Fracture | Equivocal | Negative | Total |
| Resident | Fracture | 30 | 5 | 6 | 41 |
| | Equivocal | 19 | 16 | 41 | 76 |
| | Negative | 5 | 11 | 242 | 258 |
| | Total | 54 | 32 | 289 | 375 |
| Linear weighted kappa (95% CI): 0.58 (0.50-0.66) SE 0.04 | | | | | |

There was moderate observer agreement for intracapsular fractures on radiography between the specialist in general radiology and the resident. The resident categorized more than twice the patients as having a suspected fracture (76) than the specialist in general radiology (32). There was complete disagreement, representing false negatives/positives, in 2.9% (5 + 6 of 375 patients).

**Table 4.**

| Radiography of intracapsular fractures, age over 90 | | | | | |
|---|---|---|---|---|---|
| | | Specialist in musculoskeletal radiology | | | |
| | | Fracture | Equivocal | Negative | Total |
| Specialist in general radiology | Fracture | 7 | 3 | 2 | 12 |
| | Equivocal | 0 | 1 | 5 | 6 |
| | Negative | 2 | 6 | 60 | 68 |
| | Total | 9 | 10 | 67 | 86 |
| Linear weighted kappa (95% CI): 0.54 (0.33-0.74) SE 0.10 | | | | | |

There was moderate observer agreement (κ = 0.54, SE 0.10) for intracapsular fractures on radiography between the specialist in musculoskeletal radiology and the specialist in general radiology for 86 patients over age 90. The standard error is high.

**Table 5.**

| **CT of intracapsular fractures** | | | | | |
|---|---|---|---|---|---|
| | | Specialist in musculoskeletal radiology | | | |
| | | Fracture | Equivocal | Negative | Total |
| Resident | Fracture | 36 | 1 | 0 | 37 |
| | Equivocal | 5 | 4 | 2 | 11 |
| | Negative | 4 | 3 | 177 | 184 |
| | Total | 45 | 8 | 179 | 232 |
| Linear weighted kappa (95% CI): 0,87 (0,80-0,94) SE 0,03 | | | | | |

There was almost perfect observer agreement for intracapsular fractures on CT between the specialist in musculoskeletal radiology and the resident for 232 patients with initially reported normal or equivocal radiography followed by CT due to remaining suspicion of occult hip fracture.

**Table 6.**

| **MRI of intracapsular fractures** | | | | | |
|---|---|---|---|---|---|
| | | Specialist in musculoskeletal radiology | | | |
| | | Fracture | Suspect | Negative | Sum |
| Resident | Fracture | 38 | 0 | 1 | 39 |
| | Suspect | 0 | 0 | 0 | 0 |
| | Negative | 1 | 0 | 130 | 131 |
| | Sum | 39 | 0 | 131 | 170 |
| Linear weighted kappa (95% CI): 0,97 (0,92-1,00) SE 0,02 | | | | | |

There was almost perfect observer agreement for intracapsular fractures on MRI between the specialist in musculoskeletal radiology and the resident for 170 patients with initially reported normal or equivocal radiography followed by MRI due to remaining suspicion of occult hip fracture.