Chapter 2

Methods discussed:

Diagnosis of vertebral fractures on lateral chest X-ray: intra-observer agreement of semi-quantitative vertebral fracture assessment

Hanna C. van der Jagt-Willems, Barbara C. van Munster, Mariska Leeflang, Linda R. Tulner, Evelijn Y. Beuerle, Willem F. Lems

Accepted for publication in European Journal of Radiology
Abstract

Introduction: In clinical practice lateral images of the chest are performed for various reasons. As these lateral chest X ray show the vertebrae of the thoracic and thoraco-lumbar region, we wondered if these X-rays can be used for evaluation of vertebral fractures instead of separate thoracic spine X-rays.

Methods: to evaluate the agreement and intraobserver reliability of the semiquantitative method for vertebral fractures on the lateral chest X-ray (X-chest) in comparison to the lateral thoracic spine X-ray (X-Tspine), two observers scored vertebral fractures on X-Tspine and twice on X-chest, separately, blinded and in different time periods. Agreement and Cohens’ kappa were calculated for a diagnosis of any fracture on patient level and on vertebral body level. The study was done in patients visiting an outpatient geriatric day clinic, with a high prevalence of vertebral fractures.

Results: 109 patients were included. The intraobserver agreement for X-chest versus X-Tspine was 95-98% for the two levels of fracturing, with a Cohen’s kappa of 0.88-0.91. The intraobserver agreement and reliability of the re-test on the X-chest showed an agreement between 91-98% with a Cohen’s kappa of 0.81-0.93. More vertebrae were visible on the X-chest, mean 10.2, SD 0.66 versus mean 9.8, SD 0.73 on the X-Tspine (p<0.001).

Conclusion: The results show good agreement and intraobserver reliability on the X-chest compared to the X-Tspine for visualizing vertebral fractures. The results of this study emphasizes that the routinely performed X-chest is reliable in order to diagnose vertebral fractures.
Methods discussed

Introduction
Vertebral fractures are the most common fractures and are usually caused by severe osteoporosis (1). The prevalence of vertebral fractures rises with age and is between 25% in 80-year-old women in the community (2) and up to 50% for geriatric patients (3). Unlike non-vertebral fractures, it is estimated that only thirty percent of the vertebral fractures comes to medical attention and ten percent leads to hospitalisation (4). Most vertebral fractures remain undiagnosed due to their asymptomatic course (5). Even if vertebral fractures are visualized on radiography of the spine there is underdiagnosis and undertreatment (6,7). Diagnosis and thereby secondary prevention of further fractures with treatment is essential, since both symptomatic and asymptomatic vertebral fractures are associated with high morbidity and mortality (8-11). Prevalent vertebral fractures quadruple the risk for subsequent vertebral fractures and double the risk for non-vertebral fractures (12,13).
Vertebral fractures are usually diagnosed on lateral X-rays of the thoracic and lumbar spine, although there is no consensus regarding criteria for the radiographic diagnosis (14). In contrast to the diagnosis of a non-vertebral fracture, which is usually very easy (severe pain after a fall, deformity), vertebral fractures are much more difficult to diagnose. Another point is that non-vertebral fractures occur as a yes or no phenomena, while vertebral fractures may occur gradually. This means that a vertebral fracture evolves over time from height loss of the vertebrae of 20% to height loss of more than 40%. The height loss is calculated by comparison of ventral, middle and dorsal side of each vertebrae (15). The certainty of a diagnosis of vertebral fracture is, due to the origin of gradualism of the fractures, low when the fracture is only 20%, but obvious if the fracture has evolved to severe height loss of 40% or more.
In clinical practice radiologists usually use a visual method to diagnose vertebral fractures (16). However, this method has a low reproducibility and misses fractures in 50% of the fractures in different studies (17,6). In scientific research more quantitative methods are used, e.g. the semiquantitative method of Genant (15). This method is widely used and is used as surrogate gold standard (14,18). The method is validated on X-rays of the spine in comparison to quantitative morphometry (15). It has a high intraobserver agreement (93% for an unexperienced reader and 99% for an experienced reader) and interobserver agreement varies between 90-99% for experienced as well as relatively inexperienced readers (15-19) in the original study. When using the semiquantitative method, vertebrae are graded visually, without direct measurement in height on the ventral side, dorsal side and in the middle. The method scores each vertebra for type of fracture (wedge, biconcave or crush) and severity (according to the percentage height loss).
In many hospitals, X-rays of the chest (posterior-anterior and lateral) are used in patient care to visualize heart and lungs. Some studies have used the (lateral) X-chest for diagnosing vertebral fractures (5,17), although the X-chest is not yet qualified to diagnose vertebral fractures. The question arises whether the X-chest in clinical practice is both as useful and accurate for diagnosing vertebral fractures as the X-Tspine. If that is the case,
Chapter 2

then there is no need to make an extra X-Tspine to investigate the presence of fractures in these elderly persons. The aim of this study was to evaluate the agreement and intra and interobserver reliability of the semiquantitative method on X-chest versus X-Tspine.

Methods
Setting and participants
The present cross-sectional study was conducted in the geriatric outpatient clinic at the Slotervaart Hospital in Amsterdam, a large teaching hospital. Between July and September 2011, 144 consecutive geriatric patients were recruited prospectively at their first visit. Patients were referred for several reasons, such as cognitive decline, falling, loss of weight or polypharmacy. There were no exclusion criteria, except for being unable to undergo radiology. The local medical ethics committee gave approval for the present study. All patients, or their caregivers gave their informed consent. 144 consecutive geriatric patients were recruited prospectively at their first visit and 126 gave their informed consent. Seventeen patients were excluded: two patients due to severe scoliosis and fifteen due to missing radiology of the X-chest.

Measurements
Patient’s comprehensive geriatric assessment was routinely performed including a posterior anterior and lateral X-chest. Patients included in the present study had an extra X-Tspine. For the X-Tspine the x-ray beam was centered over T7 and 80 kV was used, the distance focus to film was 110cm. For the lateral X-chest 125kV was used and a focus to film distance of 183cm.

Two independent investigators (HJ and EB) were trained to diagnose vertebral fractures with the semi-quantitative method of Genant (15). EB is a radiologist with more than ten years of experience, but without a specialty on vertebral fracture assessment, HJ is a clinician and researcher with 2 years of experience in vertebral fractures assessment on radiography.

Fractures were categorized by fracture type (wedge, biconcave or crush-fracture) and severity grade 1 - mild (20-25% height loss); grade 2 - moderate (25-40% height loss); or grade 3 - severe (> 40% height loss).

The two observers scored separately both X-rays for vertebral fractures. First, the X-Tspine was scored, blinded for name and age. Secondly, each investigator scored separately and blinded for name and age, the X-chest. The individual scores of the X-chest were compared to the X-Tspine. Thirdly, both observers scored in the same way the X-chest again to test the reproducibility of the method on this X-ray.

The number of vertebral bodies that were visible for evaluation from T4 to L2 on both X-rays was scored. This means that if more vertebrae were visible on X-chest, these extra vertebrae were left out in the analysis of the comparison between the X-Tspine and X-chest, but not in the comparison between the first and second reading of the X-chest.
Methods discussed

Analysis of vertebral bodies above T4 was not performed, as these are usually not visible on spine radiographs. Presence of vertebral fractures was scored on two levels: patient level and vertebral body level. At the patient level the presence or absence of a vertebral fracture of any grade was scored. At vertebral level the presence or absence of a fracture of any grade was scored per vertebra. Agreement was defined as full agreement on fracture type and fracture grade. As there is no gold standard for detecting vertebral fractures, we made an equal comparison between both X-rays, with the hypothesis that the X-chest can just as well be used in practice as the X-Tspine.

Statistics
The Statistical Package for the Social Sciences (SPSS version 15.0 for windows; SPSS Inc, Chicago IL) was used for statistical analyses. The intraobserver scores for the X-Tspine and the first reading of the X-chest and the intraobserver scores for the first reading of the X-chest and the second reading were compared on patient and vertebral body level for any fracture. Raw agreement scores and Cohen’s kappa were used. In addition we calculated an interobserver agreement and Cohen’s kappa. We used the WHO criteria that acceptable agreement should be in the range of 85% to 95% (20). For interpreting Cohens’ kappa, we used the Fleiss criteria: values of less than 0.40 reflect poor agreement, 0.40-0.75 would be a fair to good agreement, and values of 0.75 or above indicate a near perfect or excellent agreement (21).

Results
144 participants were recruited, and 126 gave their informed consent. Seventeen patients were excluded: two patients due to severe scoliosis and fifteen due to missing radiology of the X-chest. 109 patients were included in this study. The mean age of the participants was 81 years (range 51-94), 58% female. On the X-Tspine in 109 patients 1065 vertebrae (mean 9.8, SD 0.73) were visible. The X-chest showed 1103 vertebrae (mean 10.2, SD 0.66) in total. The difference in number of vertebrae on the two X-rays was significant (p <0.001). In 47 of the 109 patients (43%) at least one additional vertebra (in the thoraco-lumbar region) was visible on the X-chest. In 3 cases the X-chest showed a vertebral fracture not visible on the X-Tspine, which led to a new diagnosis of a vertebral fracture in a patient. The vertebral fractures visible on the X-Tspine in this cohort were all visible on the X-chest. The observers agreed in diagnosis in 99 out of 109 patients (91%) of whom 43 with a vertebral fracture. The calculated interobserver agreement on the X-Tspine for the existence of at least one vertebral fracture of any grade was 91% matching a Cohen’s kappa of 0.81 (95% confidence interval (CI) 0.92-0.70).
Chapter 2

The intraobserver agreement and reliability

On patient level observer 1 scored on the X-Tspine 40% (43/109) patients with at least one vertebral fracture. On the X-chest observer 1 scored in the first reading 42% (45/109) and in the second reading 44% (47/109) patients with at least one vertebral fracture. Observer 2 scored at least one vertebral fracture in respectively 48% (52/109) of the cases on the X-Tspine, and 51% (56/109) and 48% (52/109) of the cases on the X-chest.

On vertebrae level, observer 1 found 69 fractures in 1065 vertebrae (6%) of the X-Tspine. On the X-chest 79 fractures were found in 1103 vertebrae (7%) by observer 1 in the first reading and 74 fractured vertebrae (6%) in the second reading. Observer 2 found respectively 96 (9%) fractures on the X-Tspine, 98 (9%) and 86 (8%) fractured vertebrae on the X-chest. In figure 1 the difference in number of fractures and classification by severity (grades) of the two observers on both X-rays is shown.

Table 1 shows the intraobserver scores of the first X-chest and the X-Tspine. Intraobserver agreement was 95-98% for patient- and vertebral level, with a Cohen’s kappa of 0.88-0.91. Table 2 shows the intraobservers agreement and reliability of the test re-test on the X-chest in 1103 vertebrae. Results show an agreement between 91-98% with a Cohen’s kappa of 0.81-0.93.

Figure 1: Total number of fractures and classification by severity of detected vertebral fractures by observer and by X-ray
Table 1: Intraobserver agreement on lateral thoracic spine X-ray versus lateral chest X-ray in 109 patients with 1065 vertebrae

<table>
<thead>
<tr>
<th>Observer</th>
<th>Level</th>
<th>X-Tspine*</th>
<th>X-chest**</th>
<th>Agreement</th>
<th>kappa (CI****)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Patient Vertebral</td>
<td>Fracture, n (%)</td>
<td>43 (40)</td>
<td>44 (40)</td>
<td>95%</td>
</tr>
<tr>
<td></td>
<td>Fracture, n (%)</td>
<td>69 (6)</td>
<td>77 (7)</td>
<td>98%</td>
<td>0.88 (0.82-0.94)</td>
</tr>
<tr>
<td>2</td>
<td>Patient Vertebral</td>
<td>Fracture, n (%)</td>
<td>52 (48)</td>
<td>54 (50)</td>
<td>95%</td>
</tr>
<tr>
<td></td>
<td>Fracture, n (%)</td>
<td>96 (9)</td>
<td>96 (9)</td>
<td>98%</td>
<td>0.91 (0.87-0.96)</td>
</tr>
</tbody>
</table>

* X-Tspine: Lateral X-ray of the thoracic spine  
** X-chest: First reading lateral chest X-ray  
*** CI: 95% confidence interval

Table 2: Intraobserver reliability on lateral chest X-ray in 109 patients with 1103 vertebrae

<table>
<thead>
<tr>
<th>Observer</th>
<th>Level</th>
<th>1e reading</th>
<th>2e reading</th>
<th>Agreement</th>
<th>kappa (CI*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Patient Vertebral</td>
<td>Fracture, n (%)</td>
<td>45 (42)</td>
<td>47 (44)</td>
<td>91%</td>
</tr>
<tr>
<td></td>
<td>Fracture, n (%)</td>
<td>79 (7)</td>
<td>74 (6)</td>
<td>98%</td>
<td>0.81 (0.74-0.88)</td>
</tr>
<tr>
<td>2</td>
<td>Patient Vertebral</td>
<td>Fracture, n (%)</td>
<td>56 (51)</td>
<td>52 (48)</td>
<td>96%</td>
</tr>
<tr>
<td></td>
<td>Fracture, n (%)</td>
<td>98 (9)</td>
<td>86 (8)</td>
<td>98%</td>
<td>0.88 (0.83-0.93)</td>
</tr>
</tbody>
</table>

*CI: 95% confidence interval

Conclusion

This study shows that lateral chest X-rays can be used to diagnose vertebral fractures. For both observers the agreement and Cohens’ kappa for the X-chest versus the X-Tspine was high (respectively 95-98% and 0.88-0.91). In addition the intraobserver reliability for test re-test on the X-chest was high as well (agreement of 91-98%, kappa 0.81-0.93). This indicates that the results in our study are good in terms of agreement and are excellent in terms of Cohens’ kappa. Our findings are in line with the reported intraobserver reliability and agreement in comparable studies on vertebral fracture assessment on DXA scans versus lateral spine X-ray (22) and on CT scan versus lateral spine X-ray (23). In addition, the X-chest visualized more vertebrae, which led to a new diagnosis in 3 patients, who would have missed the diagnosis if only a lateral X-Tspine was made. No patients had a diagnosis of vertebral fractures on the lateral X-Tspine, but not on the lateral X-chest. The clinical implication of these results is that a X-chest can be used for diagnosing vertebral fractures of the spine in the mid thoracic (vertebrae T6-T9) and lumbar-thoracic region (vertebrae T10-L2). This is important because X-chest’s are performed routinely in many hospitals. However, a X-chest alone cannot replace a lumbar spine X-ray. It is shown in a cohort of 303 geriatric patients that 21% of the patients with a vertebral fracture, this fracture was located in the (lower) lumbar spine and was not visible on the X-chest (3). Guidelines of western countries postulate that a prevalent vertebral fracture needs treatment (24,25) because of the high risk of subsequent fractures. Vertebral fractures tend to occur earlier in life than hip fracture (26). The results of this study will give the
Chapter 2

opportunity to diagnose vertebral fractures and start treatment to prevent individuals from subsequent fractures, without the burden of an extra X-ray. Moreover, the very high prevalence of vertebral fractures in this cohort of patients recruited on a geriatric outpatient clinic shows the importance of the diagnosis: many subsequent fractures in this population can be prevented when medication is started. Generally a 40-50% reduction in risk is reached for subsequent vertebral fractures, and 20-30% risk reduction is shown for subsequent non-vertebral fractures (13).

We observed difference in prevalence and severity of the vertebral fractures between the observers, which led to a higher prevalence of vertebral fractures in this cohort according to observer 2. However, the interobserver agreement on patient level in this study was still high (91%) with a matching kappa of 0.81 what is classified as excellent. The difference is due to the experience and background of the observers and same variation is shown in other studies (15,22,23).

A limitation of the study is that we designed this study by using a semiquantitative method to diagnose vertebral fractures. However, in clinical practice often a simple visual method is used, which we did not test. Therefore, the results cannot right away be extrapolated to all clinical settings, but it can be expected that the results will be the same for the normal visual method. Still, it is important for all radiologists that vertebral fractures diagnosed on a X-chest do not need further exploration with additional radiography of the (thoracic) spine. Another limitation is that we were not able to calculate sensitivity and specificity of both methods and to compare these with each other. Such analyses would have required a gold standard, which was not available.

In conclusion, the lateral chest X-ray can just as well be used to diagnose vertebral fractures in practice as the X-ray of the thoracic spine. In this study vertebral fractures of the thoraco-lumbar region were more often visible on the lateral chest X-ray. In respect to high percentage of underdiagnosis of vertebral fractures on X-ray, the lateral chest X-ray should be used to diagnose vertebral fractures. Radiologists and physicians are requested to look carefully for vertebral fractures on every lateral chest X-ray.
References

Chapter 2