Comparing morphometric X-ray absorptiometry and radiography in defining vertebral wedge fractures in patients with ankylosing spondylitis

D. Vosse¹, C. Heijckmann², R. Landewe¹, D. van der Heijde¹, S. van der Linden¹ and P. Geusens¹,³

Objective. To compare the level of agreement of quantitative morphometry of the vertebrae on lateral views of the spine using conventional X-ray and using a dual X-ray absorptiometry device (DXA) in determining the degree of wedging of vertebrae in patients with ankylosing spondylitis (AS).

Methods. Thirty patients with AS underwent DXA to acquire single-energy morphometric X-ray absorptiometry (MXA) scans and conventional lateral radiography (MRX) of the thoracic and lumbar spine. Vertebral anterior and posterior heights were measured and the anterior/posterior (AP)-ratio was calculated. We analysed the level of agreement for vertebral wedging between MRX and MXA on the patient level and on the vertebral level, using average AP-ratios per patient, and per vertebra, as well as dichotomized AP-ratios (above or below cut-off levels that are commonly used to identify fractures).

Results. Per-patient analysis showed good agreement between both methods in the whole spine [intraclass correlation coefficient (ICC) = 0.64], as well as in the thoracic (ICC = 0.66) and lumbar spine (ICC = 0.62) separately. Analysis on individual vertebrae showed differences in agreement dependent on which part of the spine was measured. The ICC on all vertebrae was 0.71, 0.76 in the lumbar and 0.43 in the thoracic vertebrae. If AP-ratios were translated into fractures (yes vs no) using different cut off levels for a fracture (AP-ratios 0.75, 0.80 or 0.85) between-method agreement became fair to good (κ 0.26–0.35 in the thoracic and 0.47–0.80 in the lumbar vertebrae). Differences in classifications were in both directions and in all vertebral fractures according to the Genant definition. In this study with a prevalence of 5% of vertebral fractures, the positive predicted value (PPV) was 39% and the negative predicted value (NPV) was 97%.

Conclusion. Although the agreement between MRX and MXA in measuring global vertebral wedging, expressed as (mean) AP-ratio, was good, the reliability of both measures to assess wedging at the vertebral level was highly variable, ranging from fair to very good agreement, dependent on the level. If fracture studies are performed with either of both the methods, the results of wedging at the individual vertebral level cannot be generalized to the other method, except for wedging <0.75 at the lumbar spine. However, as the NPV was high, DXA could be of clinical value to select patients for further evaluation by X-ray to assess vertebral fractures as a sign of bone failure.

Key words: Assessment vertebral deformities, Morphometric radiography (MRX), Morphometric X-ray absorptiometry (MXA), Ankylosing spondylitis, Fractures, Osteoporosis, Bone failure.

Introduction

Ankylosing spondylitis (AS) is a chronic inflammatory disease mainly affecting the axial skeleton and is characterized by ossification of the spinal joints and ligaments. Hyperkyphosis of the upper part of the spine is a frequent clinical problem in patients with ankylosing spondylitis (AS) [1–3]. In general, hyperkyphosis is associated with vertebral osteoporosis and it is increasingly recognized that osteoporosis is a problem in patients with AS [4–8]. Vertebral deformities are regarded as one of the hallmarks of bone failure. In previous studies, we showed that the degree of hyperkyphosis in patients with AS is independently related to two radiological measures of damage: (i) radiological damage as measured on X-rays by the modified Stoke Ankylosing Spondylitis Score (mSASSS) [9] that includes erosions, osteophytes and squaring and (ii) wedging of mainly thoracic vertebrae as measured by vertebral morphometry [10, 11].

Several methods are available to establish vertebral deformities or fractures and to distinguish them from vertebral with a normal shape. They differ in the technique used [morphometric radiography (MRX) and morphometric X-ray absorptiometry (MXA)] and in the definition of fractures. Recently, Duboeuf et al. [12] reviewed these issues in the assessment of vertebral fractures. They discussed five scoring methods based on conventional morphometric radiography (MRX) and applied them on different densitometric devices (Hologic QRD, Hologic Delphi, Lunar Expert and Lunar Prodigy). One of the most common approaches is described by Genant et al. [13] in which a semiquantitative technique to assess vertebral deformities on conventional radiography is used plus the general interpretation of the image. This method uses fixed values of deformities (0.60, 0.75, 0.80). McCloskey et al. [14] developed a semi-automated technique taking into account the size of adjacent vertebrae in calculating fractures. Eastell et al. [15] defined fractures based on deviations of >3 SD as compared with a population-based database. The other two methods are the Melton approach, using reference radiographs of 200 healthy women aged 50 yr(s) and older, where vertebral heights are compared with the mean value of the control group for that vertebra and the Minne approach, dividing all individual heights (anterior, middle and posterior from T4 to L5) by the height of the fourth thoracic vertebrae, resulting in relative heights which are compared with the relative heights of a reference group [16, 17]. Rea et al. [18, 19] published the use of morphometric X-ray absorptiometry to describe vertebral deformities. All methods were tested in postmenopausal women with or without osteoporosis and could appropriately distinguish between normally and abnormally shaped vertebrae. These methods use a threshold of 15%...
[16, 17, 20] to 40% [13] height loss as fracture definition. In spite of the wealth of definitions and methods for morphometry, a gold standard in terms of technique (MRX vs MRA) and definition of fracture is still lacking [19]. Comparative studies using these techniques have not yet been performed in AS.

In the present study, we studied the agreement between two methods (MRX and MXA) for measuring vertebral wedge fractures in patients with AS.

**Patients and methods**

We invited 30 patients (52% male patients, 83% HLA-B27-positive) with AS visiting the outpatient Department of Rheumatology of the University Hospital Maastricht to take part in this cross-sectional study. They all met the modified New York criteria for AS [21]. Mean age of the patients was 52 yr(s) (±11); mean disease duration after diagnosis 17 yr(s) (±10). The patients are evaluated by a study nurse according to a standard protocol which included measurement of bone mineral density (BMD) (Hologic QDR 4500, NHANES-III reference group, Bedford, MA, USA). In addition, DXA was used to acquire MXA scans of the whole spine. These 30 patients also participated in the Outcome in AS International Study (OASIS) cohort, an international longitudinal, observational study on outcome in AS [22]. In this study, patients were followed according to a fixed protocol and data included lateral radiographs of the cervical, thoracic and lumbar spine. Both DXA scans and radiographs were performed in the same period. In the OASIS cohort, only a very small group of patients used steroids (6/215; 2.8%) at baseline. Only one of them participated in this study. The OASIS study was approved by the medical Ethical Committee of our hospital and informed consent was obtained for all patients.

On both MRX and MXA vertebral anterior and posterior heights were measured (Fig. 1A). For MXA, Hologic software using the single-energy absorptiometry was used to set markers on the vertebral corners by one observer (D.V.). Anterior (Ha) and posterior (Hp) height of the vertebrae was measured in millimetres using Hologic software (Fig. 1C). Anterior and posterior heights of the vertebrae were measured on lateral radiographs of the thoracic (T4–T12) and lumbar spine (L1–L4) in millimetres using a ruler by the same observer (DV) (Fig. 1B). Vertebral wedging was assessed on both methods by calculating the Ha/Hp ratio. Interpretation of the Ha/Hp-ratio on MRX was based on Bland and Altman [23].

To assess interobserver reliability of measuring anterior and posterior height on MXA, two readers measured the total spine of 30 AS patients. These patients were chosen by an independent observer who tried to include the entire spectrum of possible AS deformities. In a previous experiment using MRX, we have established that inter-observer reliability [ICC = 0.84 (95% CI: 0.74–0.96) for Ha/Hp ratios] was very good [11].

**Statistical analysis**

We analysed agreement between MRX and MXA on two main levels: the patient level (n = 30) and at the vertebral level (n = 390). On the patient level we correlated wedging deformities measured on MRX and MXA on different levels with BMD. On the patient level as well as on the vertebral level we calculated intraclass correlation coefficient (ICC) on Ha/Hp ratio for the lumbar and thoracic spine separately, as well as for the total spine. Furthermore, we calculated Cohen’s k’ on dichotomized fracture scores with the following cut-off ratios for the Ha/Hp ratio for a fracture: 0.75, 0.80 and 0.85. The k’s were calculated for the lumbar and thoracic spine separately as well as for the total spine. In an attempt to improve the agreement between MRX and MRA in establishing a fracture, we assumed MRX as the dependent variable and used receiver operating characteristic (ROC) analysis to find out the optimal cut-off level for Ha/Hp-ratio in determining a fracture by MRA. Different cut-off levels (0.60 – 0.85) for MRX were tested, so that different combinations of cut-off levels for MRA and MRX were obtained. We analysed if wedging deformities of vertebrae assessed by MRX vs MXA were correlated to BMD. Furthermore, linear regression analysis was performed to investigate the independent contribution of mean thoracic wedging on MRX and MXA to explain variation in the dependent variable occupit-to-wall distance (OWD). Not normally distributed variables were first normalized by the Van der Waerden technique. Specificity, sensitivity, positive (PPV) and negative predictive values (NPV) were calculated.

**Results**

Complete MRX and MXA data were available in all 30 patients. Of 390 available vertebrae, 335 could be analysed. Most missing data (concerning 55 vertebrae) occurred in the T4–T6 region and were equally distributed across methods. Eight of the 55 non-eligible vertebrae occurred in the T10–T12 region and were preferentially occurring on MRX while eligible on MXA.

Inter-observer reliability for Ha/Hp ratios on MXA on vertebral level was very good [ICC = 0.84 (95% CI: 0.51–0.95)]. Table 1 shows the correlation between deformities in separate levels and in the total spine measured on MRX and MXA. Good correlations are found between methods on concurrent levels (thoracic 0.71; lumbar 0.75; total 0.72). This analysis also confirms our previous findings that overall hyperkyphosis is mainly explained in the thoracic spine (0.89 MRX and 0.82 MXA). We did not find any correlation between deformities and BMD.

Table 2 shows the agreement in measuring the absolute Ha/Hp-ratio on MRX vs MRA on a per-patient level and on a vertebral level. Per-patient analysis showed good agreement between both methods, in the total spine (ICC = 0.64) as well as in thoracic (ICC = 0.66) and lumbar (ICC = 0.62) parts separately. Analysis on separate vertebrae (n = 335) showed a different pattern of agreement dependent on the part of the spine that

**Table 1. Correlation between deformities in separate levels and in the total spine measured on MRX and MXA.**

<table>
<thead>
<tr>
<th>BMD</th>
<th>Mean wedging thoracic spine</th>
<th>Mean wedging lumbar spine</th>
<th>Mean wedge total spine</th>
<th>Mean ratio thoracic spine</th>
<th>Mean ratio lumbar spine</th>
<th>Mean ratio total spine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.12</td>
<td>-0.12</td>
<td>-0.20</td>
<td>-0.15</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>Mean wedging thoracic spine</td>
<td>1</td>
<td>-0.12</td>
<td>0.15</td>
<td>0.66**</td>
<td>0.61**</td>
<td></td>
</tr>
<tr>
<td>Mean wedging lumbar spine</td>
<td>1</td>
<td>0.26</td>
<td>0.58**</td>
<td>0.48**</td>
<td>0.72**</td>
<td></td>
</tr>
<tr>
<td>Mean wedge total spine</td>
<td>1</td>
<td>1</td>
<td>1.17</td>
<td>0.82**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ratio thoracic spine</td>
<td>1</td>
<td>1</td>
<td>1.17</td>
<td>0.82**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ratio lumbar spine</td>
<td>1</td>
<td>1</td>
<td>1.17</td>
<td>0.82**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ratio total spine</td>
<td>1</td>
<td>1</td>
<td>1.17</td>
<td>0.82**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Correlation is significant at P = 0.05.
**Correlation is significant at P = 0.01.

**Table 2. Agreement between radiographic and DXA morphometry (absolute Ha/Hp ratios) at patient and at vertebral level.**

<table>
<thead>
<tr>
<th></th>
<th>Per patient</th>
<th>Per vertebra</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thoracic</td>
<td>0.66 (0.41–0.82)</td>
<td>0.43 (0.31–0.53)</td>
</tr>
<tr>
<td>Lumbar</td>
<td>0.62 (0.29–0.81)</td>
<td>0.76 (0.67–0.83)</td>
</tr>
<tr>
<td>Total spine</td>
<td>0.64 (0.57–0.70)</td>
<td>0.71 (0.47–0.85)</td>
</tr>
</tbody>
</table>
was examined. For the total spine (ICC = 0.71) as well as for the lumbar (ICC = 0.76) part, a good level of agreement was achieved. For the thoracic part the level of agreement (ICC = 0.43) was moderate. Bland–Altman analysis on the vertebral level showed that measurement error was independent of the magnitude of wedging (Fig. 2). Bland–Altman analysis on the patient level, however, showed that measurement error increased by decreasing Ha/Hp ratio (Fig. 3).

Subsequently, we dichotomized Ha/Hp ratios at different cut-off levels for the presence or absence of fracture (0.75; 0.80; 0.85) and we tested the agreement again. Agreement tended to improve by a more conservative cut-off level for fracture, more so in the lumbar spine than in the thoracic spine (Table 3). For moderate wedging in the lumbar spine (cut-off: 0.75) there was a good agreement (κ = 0.80), but for less severe deformations only fair-to-moderate agreement between both methods was found. On separate vertebrae the differences in classification were significant. Differences were in both directions: fractures measured on MXA but not on MRX and vice versa. Using the Genant cut-off level of 0.80 for Ha/Hp-ratio, 28 fractures were identified in 335 separate vertebrae by either method, only seven matched (25%) (Table 4). All identified fractures, both discrepant and matching, were localized in the T6–L1 region.

We further examined the influence of different cut-off levels using MRX at different cut-off levels as the dependent variable and testing an optimal cut-off level for MRA by ROC analysis. However, none of the found combinations arrived at higher levels of agreement (data not shown).

Finally, parallel to a previous study we looked at the correlation between morphometry and hyperkyphosis (expressed as OWD) in MXA. As expected we found a correlation between mean thoracic wedging measured on MRX and MXA and OWD (P = 0.51 MRX vs 0.39 in MXA). However, in a linear regression model
contribution of mean thoracic wedging measured on MXA did not reveal independent contribution to explaining hyperkyphosis (standardized $\beta = -0.488$ [MRX] vs standardized $\beta = -0.138$ [MXA]; $R^2 = 0.255$). If we assumed MRX to be the gold standard, with a prevalence of 5% of wedged fractures in this study, sensitivity was 41%, specificity 97%, PPV 39% and NPV 97%.

**Discussion**

In this study, we compared assessment of MRX on lateral radiographs of the spine and MXA on single-energy X-ray absorptiometry scans in identifying vertebral deformities and fractures. We used Genant’s method by calculating Ha/Hp ratios to describe vertebral deformities and defined different cut-off levels for vertebral fractures [13, 16, 17, 20]. If the mean Ha/Hp ratio per patient is calculated per method and both methods are compared, substantial agreement was achieved. However, the focus of interest is the separate vertebral level and here agreement was only fair.

A similar level of between-method agreement as in this study is described in the literature. Ferrar et al. [24] described an appropriate assessment of vertebral fractures on MXA scans, considering spinal radiographs as gold standard. Rea et al. [19] reported a direct comparison of visual MXA and MRX in 24 patients. They found that the prevalence of new fractures could not be detected as precisely on MXA as compared with MRX, especially in subjects with vertebral deformities. In our study, we did not use a gold standard, but we looked at direct agreement for Ha/Hp ratios assessed by both MRX and MXA. Though the interobserver ICC for both methods was good, the agreement between methods in identifying a vertebral fracture was not. Using a Ha/Hp ratio $\leq 0.80$ as cut-off point seven fractures were scored in both methods, on MXA 11 another fractures were identified and again another 10 on MRX. All identified fractures, both discrepant and matching, were localized in the thoraco-lumbar spine (T6–LI). Fifty-five individual vertebrae were not analysable, but this did not interfere with our results as they were all localized in the upper thoracic region and no discrepancies were found there.

We analysed if agreement between both MRX and MXA was better in more severe deformities, but no improvement in agreement was seen, except for moderate deformities in the lumbar spine.

In a previous study, we showed that hyperkyphosis in patients with AS has two major determinants: radiological damage and thoracic wedging [11]. For the purpose of this study we analysed whether wedging as assessed by MRX and MRA both correlated with hyperkyphosis (OWD). Both methods indeed contributed to OWD, but in MRX the results were somewhat better than in MRA. As both methods are highly correlated to one another and MXA does not adjust extra information, using both methods to explain hyperkyphosis seems not necessary.

To assess the radiological damage in our AS patients mSASSS is used. In addition, to evaluate bone failure, bone density can be used. The presence of vertebral fractures also reflects bone failure [25]. In spite of the lack in agreement in the two methods, this study indicates there is a role for MXA in assessing the presence of vertebral fractures i.e. bone failure. Indeed, the high NPV indicates that in patients without any vertebral fracture on MXA, we do not need X-ray to confirm the absence of fracture in cross-sectional studies. How far, this is true, in the clinical situation of acute back pain, remains to be studied.

One might argue that obtaining agreement among methods will be more problematic in AS patients as compared with non-AS patients, since AS patients may have spinal syndesmophytes and vertebral corner lesions that may jeopardize an appropriate assessment of the vertebral height. We were, however, able to demonstrate that two trained readers reached a high level of agreement in
measuring vertebral height in both MRX and MXA. In both experiments, we asked an independent observer to select cases in the training sets of both methods and tried to include the entire spectrum of possible AS deformities so that interference by syndesmophytes and other lesions is probably of minor importance.

Most authors find an acceptable performance of the different methods they used to identify vertebral fractures. Most of them used visual assessment of spinal radiographs as gold standard. In his review, Duboeuf et al. [12] mentioned reasonable to good sensitivity, specificity and x-scores. Nevertheless, our results were less convincing. Both methods identify vertebral deformities, but it seems they mainly detect different ones. One explanation for the discrepancy could be the poor image quality of MXA of radiographs in the thoracic region. Especially in the upper parts image quality can be insufficient. However, this can only be part of the explanation since all fractures were identified in the mid- and lower parts of the spine and here image quality is good. In 1998, Chappard et al. [26] published an evaluation of spine morphometry obtained by both MRX and MXA. They found in primary osteoporotic patients the same poor agreement for the thoracic level as we did, which underscores the argument that it is not primarily the presence of syndesmophytes that jeopardized agreement. Due to a lack of resolution, rib interposition and motion of the diaphragm, the image quality is insufficient. Furthermore, Ferrar et al. [27] stated that densitometric morphometry is less accurate on mild deformities compared with severe deformities. One could argue that imaging quality, especially in MXA, can also be influenced by BMD. Low BMD could cause vague images of vertebræ or, more specific, of vertebral corners. This problem is probably less prominent in MRX, as we know that low BMD cannot adequately be diagnosed on plain X-rays. So we examined the possible correlation between MRA, MRX and BMD, but no correlation was found. Low BMD can therefore not entirely explain the found differences.

Our study showed that the two methods we used to identify vertebral deformities in AS patients are not exchangeable at the level of T6–L1. We agree with other authors that a gold standard in vertebral morphometry would be recommendable [25]. As in rheumatoid arthritis, we need to standardize the method(s) which we will use to assess vertebral deformities in patients with AS. With the new, potentially disease-modifying drugs, we must be able to monitor radiographic progression of the disease in all its aspects, including deformities.

In summary, although the agreement between MRX and MXA in measuring global vertebral deformity expressed as (mean) AP-ratio is acceptable, both methods do not reliably assess whether or not there is a fracture. The results obtained by one method do not comply with the results obtained by another technique. Consensus on which methods should be used and how they should be applied to assess vertebral deformities in patients with AS is warranted. However, for the assessment of vertebral deformities as a sign of bone failure MRA is an accurate method to use.

**Disclosure statement:** The authors have declared no conflicts of interest.

**References**


**Rheumatology key messages**

- Measuring vertebral deformity on MRX and MXA expressed as AP-ratios is reliable if performed with trained readers.
- Agreement between methods in assessing vertebral fractures is insufficient; a gold standard in vertebral morphometry is warranted.
- MRX and MXA should be used in clinical trials to establish vertebral bone quality and deformity.