



Published in final edited form as:

*Ann Rheum Dis.* 2014 March ; 73(3): 544–550. doi:10.1136/annrheumdis-2012-202661.

## Quantitative measurement of syndesmophyte volume and height in ankylosing spondylitis using CT

Sovira Tan<sup>1</sup>, Jianhua Yao<sup>2</sup>, John A Flynn<sup>3</sup>, Lawrence Yao<sup>2</sup>, and Michael M Ward<sup>1</sup>

<sup>1</sup>National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland, USA

<sup>2</sup>Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Bethesda, Maryland, USA

<sup>3</sup>School of Medicine, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA

### Abstract

**Objective**—Syndesmophyte growth in ankylosing spondylitis can be difficult to measure using radiographs because of poor visualisation and semiquantitative scoring methods. We developed and tested the reliability and validity of a new computer-based method that fully quantifies syndesmophyte volumes and heights on CT scans.

**Methods**—In this developmental study, we performed lumbar spine CT scans on 38 patients and used our algorithm to compute syndesmophyte volume and height in four intervertebral disk spaces. To assess reliability, we compared results between two scans performed on the same day in nine patients. To assess validity, we compared computed measures to visual ratings of syndesmophyte volume and height on both CT scans and radiographs by two physician readers.

**Results**—Coefficients of variation for syndesmophyte volume and height, based on repeat scans, were 2.05% and 2.40%, respectively. Based on Bland–Altman analysis, an increase in syndesmophyte volume of more than 4% or in height of more than 0.20 mm represented a change greater than measurement error. Computed volumes and heights were strongly associated with physician ratings of syndesmophyte volume and height on visual examination of both the CT scans ( $p < 0.0001$ ) and plain radiographs ( $p < 0.002$ ). Syndesmophyte volumes correlated with the Schober test ( $r = -0.48$ ) and lateral thoracolumbar flexion ( $r = -0.60$ ).

**Conclusions**—This new CT-based method that fully quantifies syndesmophytes in three-dimensional space had excellent reliability and face and construct validity. Given its high

---

**Correspondence to** Dr Michael M Ward, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Building 10 CRC, Room 4-1339, Bethesda, MD 20892, USA; wardm1@mail.nih.gov.

**Contributors** MMW conceived the study. ST, JY, LY, JAF and MMW designed the study and ST and MMW performed the analysis. ST drafted the manuscript and all authors provided critical review and approval of the final version.

**Competing interests** None.

**Ethics approval** The study protocol was approved by the institutional review boards of the National Institutes of Health and Johns Hopkins Medical Center.

**Patient consent** Obtained.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** Primary data will be publicly available at the conclusion of the study.

precision, this method shows promise for longitudinal clinical studies of syndesmophyte development and growth.

---

Spinal fusion is the clinical and pathological hallmark of ankylosing spondylitis (AS). Spinal fusion results from the fusion of apophyseal joints and bridging syndesmophytes.<sup>1</sup> Because apophyseal joint fusion is difficult to detect, monitoring the progression of spinal fusion in AS has relied on radiographic examination of the development and growth of syndesmophytes, as, for example, in studies of biomarkers and genetic polymorphisms associated with radiographic severity<sup>2–5</sup> and of the link between inflammation and bone formation.<sup>67</sup> Studies of whether tumour necrosis factor alpha antagonists can slow spinal fusion used as the outcome the modified Stoke ankylosing spondylitis spinal score (mSASSS), a radiographic score heavily weighted by syndesmophytes.<sup>8–10</sup>

The visibility of syndesmophytes on radiographs may be influenced by patient positioning, radiographic technique and shadows from superimposed anatomical structures, particularly on lumbar studies. These limitations are a consequence of using a two-dimensional technique to assess a complex three-dimensional structure. The use of radiographs also requires that readers correctly apply semiquantitative scores of vertebral abnormalities. The ability of such scores to capture progression is dependent on the fineness of the rating categories (ie, precision). However, finer rating categories are typically associated with more measurement error. Both visualisation problems and semiquantitative scoring might limit the ability of radiographic scores to detect syndesmophyte progression. Even by the most sensitive radiographic measures, only 46.4% of patients with AS show progression over 2 years, and even fewer have progression larger than the smallest detectable difference.<sup>1112</sup> Testing interventions to slow spinal fusion would benefit from methods that can reliably detect changes in syndesmophytes over shorter periods.

To address these limitations, we developed a computer algorithm that measures syndesmophytes on lumbar CT scans.<sup>1314</sup> Our goal was to develop a method that exploits the three-dimensional information of CT scans and assesses syndesmophytes along the entire vertebral rim, was fully quantitative, and measures both syndesmophyte volume and height. Compared to semiquantitative scores, a quantitative measure would have considerably better precision, that is, fineness of the measurement. With greater precision comes greater ability to distinguish different rates of progression. Automation would enhance the method's reliability by eliminating reader variation. The objective of this study was to evaluate the reliability and validity of syndesmophyte quantitation by this new method.

## METHODS

### Patients

We enrolled adults with AS in this prospective longitudinal study. Patients were a convenience sample of those seen at the National Institutes of Health or Johns Hopkins Medical Institutions. Inclusion criteria were age 18 years or older, diagnosis of AS by the modified New York criteria,<sup>15</sup> and a Bath ankylosing spondylitis radiology index (BASRI) lumbar spine score of 0, 1, 2, or 3 (ie, excluding patients with completely fused lumbar

spines).<sup>16</sup> To have representation of patients with different degrees of syndesmophytes, we enrolled at least five patients in each BASRI category. We excluded patients unable to provide informed consent, who were pregnant, or anticipated being unavailable for follow-up. The study protocol was approved by the institutional review boards of both centres, and all patients provided written informed consent.

Patients had a clinical evaluation, and completed questionnaires, including the Bath ankylosing spondylitis functional index (BASFI).<sup>17</sup> Lumbar motion was evaluated using the modified Schober test and lateral thoracolumbar flexion. Patients had anteroposterior and lateral radiographs of the lumbar spine and lumbar spine CT scans.

### **CT scanning**

Patients were scanned on a Philips Brilliance 64 (slice thickness 1.5 mm) or a GE Lightspeed Ultra scanner (slice thickness 1.25 mm). For both scanners, voltage and current parameters were 120 kVp and 300 mAs, respectively. Patients were scanned from T10 to L4, providing four intervertebral disk spaces (IDS) for processing: T11–T12, T12–L1, L1–L2, L2–L3. We focused on the thoracolumbar junction because syndesmophytes are most common in this region.<sup>18,19</sup> The estimated equivalent absorbed radiation dose from the CT scan was 8.01 mSv (0.801 rem).

### **Reliability study**

We invited a subsample of patients who had syndesmophytes at one or more vertebral levels to have two lumbar CT scans on the same day. After the first scan, patients stood up before lying down again for the second scan. This ensured that the variation between scans was in the range expected in a longitudinal study.

### **CT image analysis**

We developed a semi-automated computer algorithm to quantitate syndesmophyte volume and height.<sup>13,14</sup> The algorithm first identifies the complete vertebral bodies. Then, for each IDS, the algorithm detects syndesmophytes as any bone projecting from the periphery of the vertebral endplates, as voxels lying between the two planes of the endplates. We devised the algorithm to report the total volume of syndesmophyte at each IDS, regardless of whether they were contiguous. The algorithm also measured the height of the tallest syndesmophyte in each IDS. Measures in four IDS were added to provide volume and height measures per patient.

### **CT and radiography reading**

To test the validity of the computed measures, we used visual readings of both the lumbar CT scans and radiographs as reference standards. Reader 1, a rheumatologist (MMW), and reader 2, a musculoskeletal radiologist (LY), scored the baseline CT scans and radiographs independently and blinded to the computed measurements. Details of the scoring are provided in the legend to figure 2. As a second measure of validity, each IDS was rank-ordered by syndesmophyte volume by reader 1. The same reader also ranked patients by total syndesmophyte volume.

## Statistical analysis

We assessed reliability by the differences in syndesmophyte volume and height between paired scans, coefficients of variation (CV), and intraclass correlation coefficients (ICC). We evaluated agreement between paired measurements using Bland–Altman analysis, which plots the difference of each pair against their mean.<sup>20</sup> The CI around these differences provides an estimate of the variation expected from measurement error; differences greater than the 95% limits of agreement are taken to represent changes greater than expected by chance. Differences that vary with the mean of the measurements (ie, heteroskedasticity) may indicate bias but are also seen with ratio-scaled variables.

For validity, we compared the algorithms' volumes and heights with the physicians' semiquantitative readings. We used the stratified Kruskal–Wallis trend test to evaluate if volume and height measurements increased with the physicians' scores, accounting for non-independence of observations within patients.<sup>21</sup> For heights, we divided syndesmophyte height in millimetres by the height of the IDS, to be comparable to the physician's semiquantitative score. We tested agreement between the readers using the weighted kappa statistic.

We used Spearman correlations to test associations between physician ranking of syndesmophyte volume and rankings based on the algorithm's volume estimates, accounting for clustering of data within patients. We also correlated summed measures over four IDS for each patient with the duration of AS, Schober test, lateral thoracolumbar flexion (using the poorer of right and left-side values) and BASFI. We hypothesised that patients with more syndesmophytes would have a longer duration of AS, more limited spinal flexibility and more functional impairment. Spearman correlations were performed using SAS software (V. 9.2).

## RESULTS

### Patients

We enrolled 38 patients (31 men and seven women; mean age 46.1 years (SD 11.5); mean duration of AS 20.0 years (SD 11.8). BASRI lumbar spine scores were 0 in 21%, 1 in 13%, 2 in 26% and 3 in 40%. The mean Schober test was 3.3 cm (SD 1.2), mean lateral thoracolumbar flexion was 11.1 cm (SD 4.8), and mean BASFI was 25.7 (SD 23.0). The validity study included 152 IDS of 38 patients. The algorithm detected syndesmophytes in 99 IDS (65%). When viewed three-dimensionally, syndesmophytes exhibited great variety in shape and location (figure 1). Rows 1, 3 and 5 show three IDS with several syndesmophytes, both ascending and descending. Row 2 shows two thick syndesmophytes that appear bridged on plain film but not on CT. Row 4 shows a single thin ascending syndesmophyte. Row 6 shows a partially bridged IDS, which appears completely fused on plain film but shows potential for progression when examined by CT.

### Reliability study

Nine patients (eight men, mean age 54 years) were enrolled in the reliability study, providing data on 36 IDS. The absolute differences in syndesmophyte volume and height between

scans were small (table 1). Very high reliability was also reflected by the low CV values and high ICC.

In the Bland–Altman analysis, volume measures were heteroskedastic, with larger inter-scan differences for IDS with higher syndesmophyte volumes. Bland–Altman analysis was therefore performed on log-transformed values, and the 95% limits of agreement for volume were in terms of percentage<sup>22</sup> (see supplementary figure S1, Height measures exhibited no heteroskedasticity and limits of agreement were evaluated in the original scale (see supplementary figure S1, available online only). The 95% limits of agreement were narrow, indicating very good reliability. For individual IDS, a difference in volume of more than 4%, or an increase in height of more than 0.20 mm, represents changes beyond those expected by measurement error. For summed values of the four IDS in an individual patient, the limits of agreement were 3% for syndesmophyte volume and 0.39 mm for height.

### Validity study

Figure 2 shows the association of computed volumes and heights with the physicians' CT and radiography ratings. Based on visual readings of the CT scans, reader 1 categorised 31% of the 152 IDS as having a syndesmophyte volume score of 0, 26% with a volume score of 1, 21% with a volume score of 2 and 22% with a volume score of 3. Reader 2 categorised 24%, 35%, 23% and 18% of the 152 IDS in the four categories, respectively, with very good agreement between readers (weighted kappa 0.80;  $p < 0.0001$ ). Volumes computed by the algorithm were significantly associated with the readers' scores (figure 2A,  $p < 0.0001$ ). Of note, only two IDS (1.3%) with a visual rating of 0 had a computed syndesmophyte volume of greater than 0, indicating that the algorithm produced very few false positive results. Conversely, the algorithm detected a computed volume of 0 in only eight IDS (5.3%) that were given a score of 1 by both readers.

Results were similar for computed heights, with strong associations with visual readings of syndesmophyte height on the CT scans (figure 2B,  $p < 0.0001$ ). IDS read as not having syndesmophytes had computed heights concentrated at 0, while IDS with bridging had computed heights concentrated at 1.0 (as the proportion of syndesmophyte height to IDS height). Computed volumes and heights were also significantly associated with physicians' scores of plain radiographs ( $p < 0.002$  and  $p < 0.0001$ , respectively) (figure 2C,D).

The dispersions of computed volumes and heights in figure 2 were smaller for CT scores than for radiograph scores, and the associations were stronger with CT readings. These results reflect the improved detection of syndesmophytes by CT scan. The number of IDS with a score of 0 was much larger for readings of the radiographs than for readings of the CT. For example, reader 1 scored 31% of IDS as without syndesmophytes by CT, 47% without syndesmophytes on the anteroposterior and lateral radiographs, and 68% without syndesmophytes on reading only the lateral radiograph.

We also assessed validity by comparing the rank order of syndesmophyte volumes by the algorithm and the physician's reading. The correlations between the physician ranking and computer ranking were 0.95 for individual IDS and 0.93 for summed volumes of individual patients (figure 3). We also used these rankings to test our semiquantitative scoring system

of physician readings of syndesmophyte volume on the CT scans. The median ranks of IDS rated as 0, 1, 2, or 3 were 1, 16, 57.5 and 81.5 for reader 1, and 1, 11.5, 60 and 81 for reader 2.

Summed syndesmophyte volumes were larger in patients with longer durations of AS ( $r=0.35$ ;  $p=0.03$ ), and those with more limited Schober test ( $r=-0.48$ ;  $p=0.003$ ), more limited lateral thoracolumbar flexion ( $r=-0.60$ ;  $p<0.0001$ ), and higher BASFI ( $r=0.32$ ;  $p=0.05$ ). Similarly, the sum of syndesmophyte heights across the four IDS was correlated with the duration of AS ( $r=0.34$ ;  $p=0.04$ ), Schober test ( $r=-0.55$ ;  $p=0.0003$ ), lateral thoracolumbar flexion ( $r=-0.60$ ;  $p<0.0001$ ) and BASFI ( $r=0.38$ ;  $p=0.02$ ).

### Longitudinal study

To assess this method's ability to detect syndesmophyte growth over time, we compared syndesmophyte volume and height between baseline CT scans and those obtained 1 year later in two patients (figure 4). For patient 1, the T11/T12 IDS was completely fused and showed little change. T12/L1 and L1/L2 IDS showed appreciable increases in both syndesmophyte volume (128% and 55%, respectively) and height (0.92 mm and 0.56 mm). For patient 2, T11/T12 and T12/L1 had stable syndesmophytes, while L1/L2 and L2/L3 exhibited volume progression above the 95% limits of agreement. Neither patient had readily detectable changes on plain radiographs over 1 year.

## DISCUSSION

An accurate, reliable and precise method to measure syndesmophytes is needed for studies of the pathogenesis of spinal damage in AS and evaluation of the effects of treatments on the progression of spinal fusion. Studies based on radiographs have been useful, but poor visualisation and interpreter variation may limit both the accuracy and reliability of radiographic readings. To overcome these problems, we tested a new method for evaluating syndesmophytes using CT. The method is fully quantitative and semi-automated, and measures syndesmophyte volume and height in three dimensions.

High reliability may be difficult to achieve with very precise measures. However, our method is both precise and highly reliable. We tested reliability by scanning patients twice on the same day, replicating the variation likely to be encountered in a longitudinal study. The limits of agreement were quite small, indicating the method is capable of detecting relatively small increases in both volume (4% or more) and height (0.2 mm or more) as true changes. Having a measure that is both precise and highly reliable is important because it allows changes to be detected over shorter time periods and potentially with fewer patients.

We examined the validity of the method by comparing its measures with the readings of two physicians and observed very strong associations. Computed volumes and heights were higher in IDS that were also scored as more severely affected by the physician readers on the CT scans. There was little suggestion that the algorithm over-detected syndesmophytes. Because methods for visual ratings of syndesmophyte volume or height on CT scans have not been published, we developed measures for this study. These measures had face validity, and exhibited the expected associations with rank-ordered scans. The consistency of results

between readers also supports the validity of these associations. The association between rank orders of computed volumes and physician ratings of volume also provides evidence of the algorithm's validity using a method not dependent on the semiquantitative scores. We also assessed validity against readings of radiographs, despite anticipating that these associations would not be as strong as those based on readings of the CT scans. The algorithm's measures are based on the three-dimensional information available in a CT scan, and the CT readings also take account of this information. In contrast, volume estimation on two-dimensional radiographs is very difficult, and we used the mSASSS syndesmophyte score as a proxy. Computed volumes and heights were strongly associated with radiographic readings despite this limitation.

In the longitudinal study of two patients, the computer algorithm was able to detect increases in syndesmophyte volume in five of eight IDS over 1 year. In four of the IDS, the changes were also qualitatively visible in the CT surface reconstruction. These changes provide additional evidence of the method's validity.

Better visualisation with CT comes with the cost of increased radiation exposure. The average equivalent dose per patient was 8.1 mSv with the CT protocol we used, compared to 3.70 mSv from an anteroposterior and lateral lumbar spine radiograph, and 2.59 mSv from lateral films of the cervical and lumbar spine. With newer scanning technology, it will be possible to lower the radiation dose by 25%.<sup>2324</sup> We chose to develop this method using CT given the exquisite sensitivity of CT for calcification and bone formation. Quantifying syndesmophyte volume with MRI would be difficult, given the typical lack of magnetic resonance signal associated with calcification, and the similar and generally low signal of the annulus fibrosis, intervertebral ligaments and adjacent paravertebral musculature. The AS spine MRI-chronic (ASspiMRI-c) has been used to measure spinal damage based on MRI scans, but it assesses the number rather than the size of syndesmophytes per IDS, and features such as erosions contribute to the same score as syndesmophytes.<sup>25</sup> Its reliability was limited, and did not correlate with mSASSS.<sup>25</sup> Difficulty visualising syndesmophytes has also been noted in other MRI studies.<sup>2627</sup> MRI has not gained wide-spread use to measure spinal damage in AS.<sup>28</sup>

The strengths of this study include the evaluation of both reliability and validity of this new method, with multiple measures to assess face and construct validity, and with two independent readers. The method does not require specialised or modified equipment, but uses data from high-resolution CT scanners currently used in routine clinical practice. The study is limited in not testing criterion validity, as physical measurements of syndesmophytes were not possible. However, for a longitudinal study, reliability is arguably more important. The study is also limited by the relatively small number of patients enrolled in the reliability study, but this number was sufficient for estimation of the limits of agreement. A 1-day repeat CT scan entails additional radiation exposure that advances knowledge but offers no clinical benefit. Because the algorithm only considers as a syndesmophyte bone that lies between the vertebral endplates, it underestimates the volume of syndesmophytes that originate below the vertebral rim. Despite this limitation, the method has excellent validity. The method can be readily applied to other regions of the spine, but scanning a more extensive region would entail greater radiation exposure.

For studies that monitor changes in syndesmophytes, the novel method we have presented has high reliability and is substantially more precise than methods based on radiographs. This method has promise for use in characterising the development and progression of syndesmophytes over time.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

The authors would like to thank Lori Guthrie, RN, and Amanda Bertram for assistance.

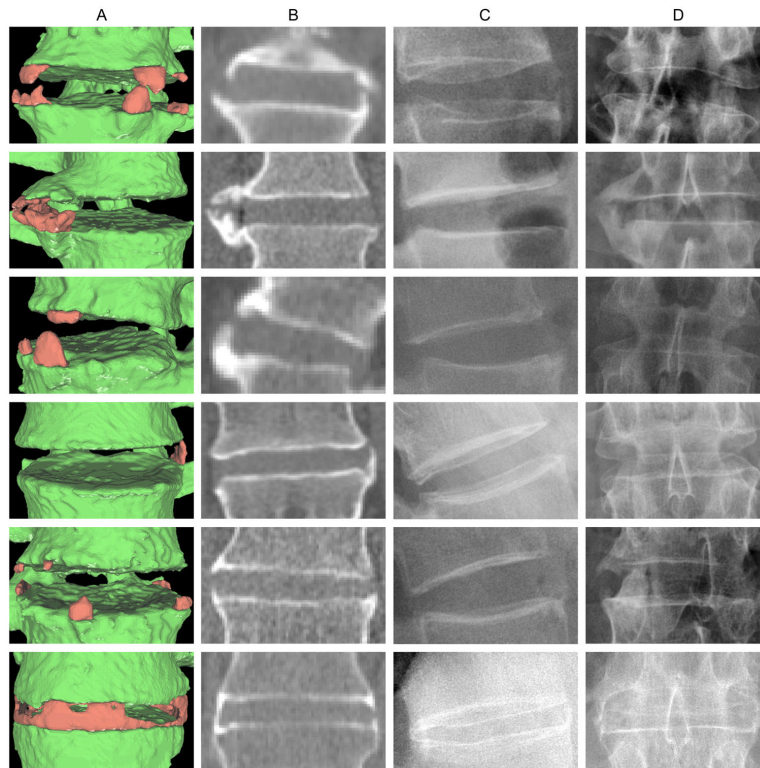
**Funding** This work was supported by the Intramural Research Program, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, and by the Clinical Center, National Institutes of Health, and the Johns Hopkins University School of Medicine General Clinical Research Center (grant number M01-RR00052 from the National Center for Research Resources/NIH).

## REFERENCES

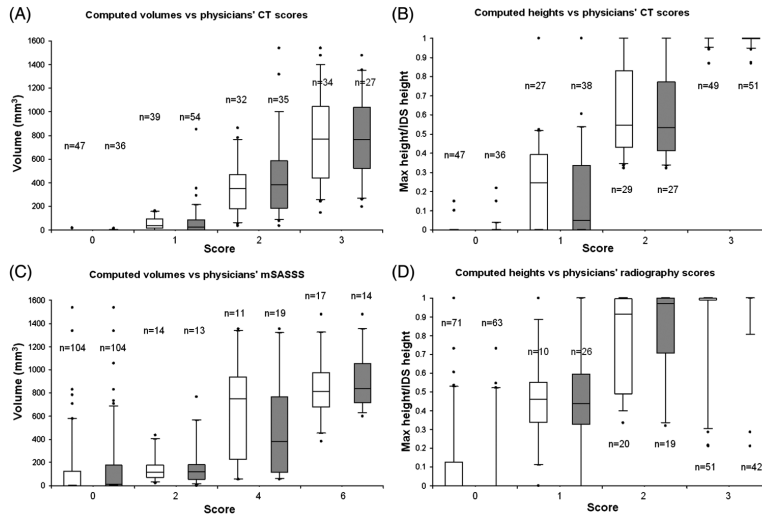
- Braun J, Sieper J. Ankylosing spondylitis. *Lancet*. 2007; 369:1379–90. [PubMed: 17448825]
- Haroon N, Maksymowych WP, Rahman P, et al. Radiographic severity of ankylosing spondylitis is associated with polymorphism of the large multifunctional peptidase 2 gene in the Spondyloarthritis Research Consortium of Canada cohort. *Arthritis Rheum*. 2012; 64:1119–26. [PubMed: 22034108]
- Appel H, Ruiz-Heiland G, Listing J, et al. Altered skeletal expression of sclerostin and its link to radiographic progression in ankylosing spondylitis. *Arthritis Rheum*. 2009; 60:3257–62. [PubMed: 19877044]
- Daoussis D, Liossis SN, Solomou EE, et al. Evidence that Dkk-1 is dysfunctional in ankylosing spondylitis. *Arthritis Rheum*. 2010; 62:150–8. [PubMed: 20039407]
- Maksymowych WP, Landewé R, Conner-Spady B, et al. Serum matrix metalloproteinase 3 is an independent predictor of structural damage progression in patients with ankylosing spondylitis. *Arthritis Rheum*. 2007; 56:1846–53. [PubMed: 17530713]
- Maksymowych WP, Chiowchanwisawakit P, Clare T, et al. Inflammatory lesions of the spine on magnetic resonance imaging predict the development of new syndesmophytes in ankylosing spondylitis: evidence of a relationship between inflammation and new bone formation. *Arthritis Rheum*. 2009; 60:93–102. [PubMed: 19116919]
- Baraliakos X, Listing J, Rudwaleit M, et al. The relationship between inflammation and new bone formation in patients with ankylosing spondylitis. *Arthritis Res Ther*. 2008; 10:R104. [PubMed: 18761747]
- van der Heijde D, Landewé R, Baraliakos X, et al. Radiographic findings following two years of infliximab therapy in patients with ankylosing spondylitis. *Arthritis Rheum*. 2008; 58:3063–70. [PubMed: 18821688]
- van der Heijde D, Landewé R, Einstein S, et al. Radiographic progression of ankylosing spondylitis after up to two years of treatment with etanercept. *Arthritis Rheum*. 2008; 58:1324–31. [PubMed: 18438853]
- van der Heijde D, Salonen D, Weissman BN, et al. Assessment of radiographic progression in the spines of patients with ankylosing spondylitis treated with adalimumab for up to 2 years. *Arthritis Res Ther*. 2009; 11:R127. [PubMed: 19703304]
- Wanders AJ, Landewé RB, Spoorenberg A, et al. What is the most appropriate radiologic scoring method for ankylosing spondylitis? A comparison of the available methods based on the Outcome Measures in Rheumatology Clinical Trials filter. *Arthritis Rheum*. 2004; 50:2622–32. [PubMed: 15334477]



12. Spoorenberg A, de Vlam K, van der Linden S, et al. Radiological scoring methods in ankylosing spondylitis. Reliability and change over 1 and 2 years. *J Rheumatol*. 2004; 31:125–32. [PubMed: 14705231]
13. Tan S, Yao J, Ward MM, et al. Computer aided evaluation of ankylosing spondylitis using high-resolution CT. *IEEE Trans Med Imaging*. 2008; 27:1252–67. [PubMed: 18779065]
14. Tan S, Yao J, Yao L, et al. Improved precision of syndesmophyte measurement for the evaluation of ankylosing spondylitis using CT: a phantom and patient study. *Phys Med Biol*. 2012; 57:4683–704. [PubMed: 22750760]
15. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum*. 1984; 27:361–8. [PubMed: 6231933]
16. MacKay K, Mack C, Brophy S, et al. The Bath Ankylosing Spondylitis Radiology Index (BASRI): a new, validated approach to disease assessment. *Arthritis Rheum*. 1998; 41:2263–70. [PubMed: 9870884]
17. Calin A, Garrett S, Whitelock H, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. *J Rheumatol*. 1994; 21:2281–5. [PubMed: 7699629]
18. Vinje O, Dale K, Møller P. Radiographic evaluation of patients with Bechterew's syndrome (ankylosing spondylitis) and their first-degree relatives. Findings in the spine and sacro-iliac joints and relations to non-radiographic findings. *Scand J Rheumatol*. 1985; 14:119–32. [PubMed: 3873698]
19. Spencer DG, Park WM, Dick HM, et al. Radiological manifestations in 200 patients with ankylosing spondylitis: correlation with clinical features and HLA B27. *J Rheumatol*. 1979; 6:305–15. [PubMed: 490525]
20. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 1:307–10. [PubMed: 2868172]
21. Hothorn T, Hornik K, van de Wiel MA, et al. Implementing a Class of Permutation tests: the coin Package. *J Stat Software*. 2008; 28:1–23.
22. Euser AM, Dekker FW, le Cessie S. A practical approach to Bland–Altman plots and variation coefficients for log transformed variables. *J Clin Epidemiol*. 2008; 61:978–82. [PubMed: 18468854]
23. May MS, Wüst W, Brand M, et al. Dose reduction in abdominal computed tomography: intraindividual comparison of image quality of full-dose standard and half-dose iterative reconstructions with dual-source computed tomography. *Invest Radiol*. 2011; 46:465–70. [PubMed: 21467948]
24. Kalra MK, Woisetschläger M, Dahlström N, et al. Radiation dose reduction with Sinogram Affirmed Iterative Reconstruction technique for abdominal computed tomography. *J Comput Assist Tomogr*. 2012; 36:339–46. [PubMed: 22592621]
25. Braun J, Baraliakos X, Golder W, et al. Analysing chronic spinal changes in ankylosing spondylitis: a systematic comparison of conventional x rays with magnetic resonance imaging using established and new scoring systems. *Ann Rheum Dis*. 2004; 63:1046–55. [PubMed: 15066862]
26. Chiowchanwisawakit P, Østergaard M, Pedersen SJ, et al. Validation of definitions for structural lesions detected by magnetic resonance imaging in the spine of patients with spondyloarthritis. *J Rheumatol*. 2009; 36(Suppl. 84):39–47.
27. Madsen KB, Jurik AG. MRI grading method for active and chronic spinal changes in spondyloarthritis. *Clin Radiol*. 2010; 65:6–14. [PubMed: 20103415]
28. Baraliakos X, van der Heijde D, Braun J, et al. OMERACT magnetic resonance imaging initiative on structural and inflammatory lesions in ankylosing spondylitis— report of a special interest group at OMERACT 10 on sacroiliac joint and spinal lesions. *J Rheumatol*. 2011; 38:2051–4. [PubMed: 21885516]

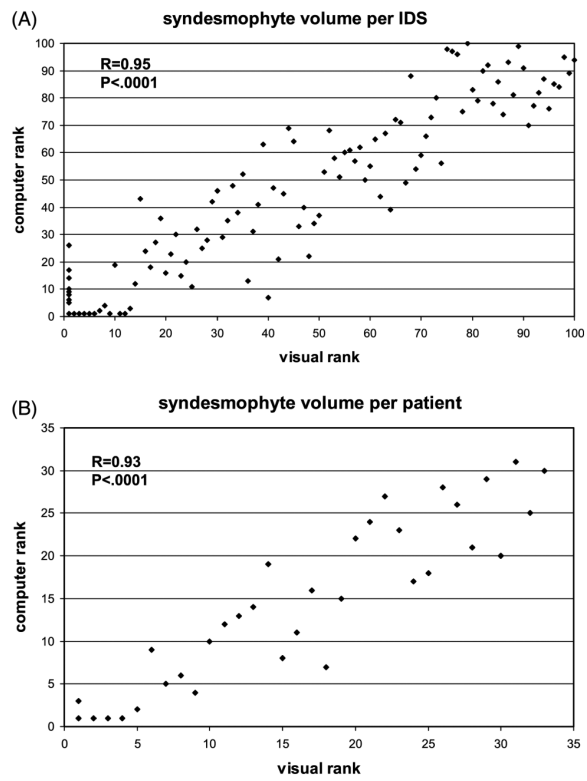


**Figure 1.** Examples of syndesmophytes detected by the algorithm. Column A shows three-dimensional surface reconstructions of the CT image with syndesmophytes in red and vertebral bodies in green. Column B shows a single slice from the CT scan. Columns C and D show the corresponding lateral and anteroposterior radiographs, respectively.

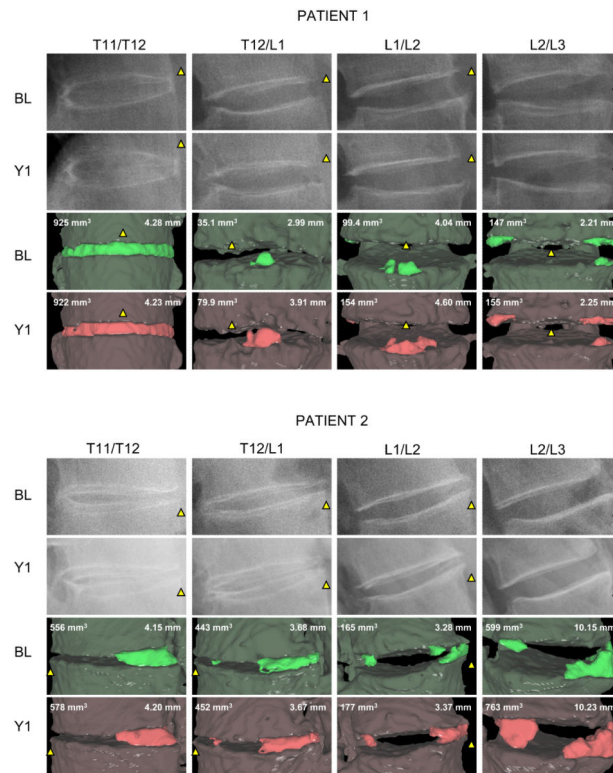


**Figure 2.**

Boxplots of computed syndesmophyte volume and height by physicians' scores (white for reader 1; grey for reader 2). (A) Computed volumes versus physicians' CT volume ratings. (B) Computed heights versus physicians' CT height ratings. (C) Computed volumes versus modified Stoke ankylosing spondylitis spinal score (mSASSS) scores. (D) Computed heights versus physicians' radiography height ratings. N is the number of intervertebral disc spaces. For syndesmophyte volume of the CT scans, we scored each intervertebral disc spaces (IDS) as 0=no syndesmophyte; 1=small isolated syndesmophytes involving less than a quarter of the vertebral rim and no bridging; 2=syndesmophyte involving more than a quarter of the vertebral rim or focal bridging; 3=bridging involving more than a quarter of the vertebral rim. For syndesmophyte heights, 0=no syndesmophyte; 1=tallest syndesmophyte less than half of the IDS height; 2=tallest syndesmophyte more than half of the IDS height but not bridging; 3=bridging. Lateral radiographs were scored based on the mSASSS as 0=no syndesmophyte; 2=syndesmophyte but not complete bridging; 3=bridging. We scored T11 lower to L3 upper (eight anterior corners per patient). An individual IDS could therefore be scored 0 if neither corner had a syndesmophyte, 2 if only one corner had a syndesmophyte, 4 if both corners had non-bridging syndesmophytes, or 6 if bridging was present. We scored antero-posterior and lateral radiographs for syndesmophyte height as 0=no syndesmophyte; 1=syndesmophyte less than half of IDS height; 2=syndesmophyte more than half of the IDS height but not bridging; 3=bridging.



**Figure 3.** Correlation between the computer and physician ranking of syndesmophyte volumes for individual intervertebral disc (IDS) spaces (A) and patients (B).



**Figure 4.** Syndesmophyte evolution over 1 year in two patients. The columns show the four intervertebral disc spaces. Lateral radiographs at baseline (BL) and year 1 (Y1) are shown in the first two rows. The following rows show the three-dimensional surface views of the syndesmophytes (bright green and red) segmented from the CT scans at baseline and year 1. The computed syndesmophyte volume (mm<sup>3</sup>) and maximal height (mm) are indicated. To aid orientation, the yellow triangles indicate the anterior aspect of the intervertebral disc space.

**Table 1**

Reliability of computed syndesmophyte volume and height measures\*

	Mean±SD			ICC	CV (%)	95% limits of agreement
	1st scan	2nd scan	Interscan difference			
Volume per IDS (n=36)	349±480 mm <sup>3</sup>	351±483 mm <sup>3</sup>	5.51±8.63 mm <sup>3</sup>	0.99	2.05	(-4.0 to 4.0) %
Volume per patient (n=9)	1396±1564 mm <sup>3</sup>	1404±1564 mm <sup>3</sup>	18.3±19.6 mm <sup>3</sup>	0.99	1.31	(-3.0 to 3.0) %
Height per IDS (n=36)	3.48±2.60 mm	3.49±2.62 mm	0.074±0.093 mm	0.99	2.40	(-0.18 to 0.20) mm
Height per patient (n=9)	13.9±7.01 mm	13.9±6.99 mm	0.13±0.13 mm	0.99	0.927	(-0.38 to 0.39) mm

CV, coefficient of variation; ICC, intraclass correlation coefficient; IDS, intervertebral disk spaces. The 95% limits of agreement are obtained from Bland–Altman analysis.

\* Interscan differences are in absolute value.