EXTENDED REPORT

Radiographic assessment of sacroiliitis by radiologists and rheumatologists: does training improve quality?

A van Tubergen, L Heuft-Dorenbosch, G Schulpen, R Landewé, R Wijers, D van der Heijde, J van Engelshoven, Sj van der Linden

....

See end of article for authors' affiliations
....

Correspondence to:
Professor Sj van der Linden, Department of Internal Medicine, Division of Rheumatology, University Hospital Maastricht, PO Box 5800, 6202 AZ Maastricht, The Netherlands; sli@sint.azm.nl

Accepted 2 September 2002
....

Objective: To assess performance of radiologists and rheumatologists in detecting sacroiliitis
Methods: 100 rheumatologists and 23 radiologists participated. One set of films was used for each assessment, another for training, and the third for confidence judgment. Films of HLA-B27+ patients with AS were used to assess sensitivity. For specificity films of healthy HLA-B27− relatives were included. Plain sacroiliac (SI) films with simultaneously taken computed tomographic scans (CTs) were used for confidence judgment. Three months after reading the training set, sensitivity and specificity assessments were repeated. Next, participants attended a workshop. They also rated 26 SI radiographs and 26 CTs for their trust in each judgment. Three months later final assessments were done.
Results: Sensitivity (84.3%/79.8%) and specificity (70.6%/74.7%) for radiologists and rheumatologists were comparable. Rheumatologists showed 6.3% decrease in sensitivity after self education (p=0.001), but 3.0% better specificity (p=0.008). The decrease in sensitivity reversed after the workshop. Difference in sensitivity three months after the workshop and baseline was only 0.5%. Sensitivity <50% occurred in 13% of participants. Only a few participants showed changes of >5% in both sensitivity and specificity. Intraobserver agreement for sacroiliitis grade 1 or 2 ranged from 65% to 100%. Sensitivity for CT (86%) was higher than for plain films (72%) (p<0.001) with the same specificity (84%). Confidence ratings for correctly diagnosing presence (7.7) or absence (8.3) of sacroiliitis were somewhat higher than incorrectly diagnosing the presence (6.6) or absence (7.4) of sacroiliitis (p<0.001).
Conclusion: Radiologists and rheumatologists show modest sensitivity and specificity for sacroiliitis and sizeable intraobserver variation. Overall, neither individual training nor workshops improved performance.

Ankylosing spondylitis (AS) is a chronic inflammatory disease, in which predominantly the axial skeleton is affected. Sacroiliitis is considered as the hallmark of the disease. Involvement of the sacroiliac (SI) joints is usually established by plain radiographs, or—to a lesser degree—by computed tomography scan (CT) or magnetic resonance imaging (MRI). Rheumatologists in daily practice mostly order plain radiographs for diagnosing sacroiliitis, and they often read these films themselves. Radiologists read and report radiographs of SI joints examinations requested by rheumatologists, general practitioners, or orthopaedic surgeons. Reading radiographs of the SI joints is considered difficult and the diagnosis of sacroiliitis is often missed or incorrectly established. Inter- and intraobserver variations have been reported to be large.[1–3] κ Statistics to express intraobserver variation in these studies ranged from 0.07 to 1.0[2 3] and interobserver variation from 0.19 to 0.79.[1–3]

We questioned whether the performance of radiologists and rheumatologists in reading SI radiographs differs, and whether this performance might be improved by offering training sessions. We provided individual self education with a training set of radiographs and as a second step organised workshops to read radiographs of the SI joints. Our study aimed at defining variations in sensitivity and specificity of detecting sacroiliitis for both radiologists and rheumatologists, and at investigating to what degree sensitivity and specificity would persistently change after training sessions. We also aimed at quantifying the intraobserver variability in reading these radiographs. Furthermore, we wanted to estimate the self confidence of radiologists and rheumatolo-

gists in determining the presence or absence of sacroiliitis, for both plain radiographs and CT of the SI joints.

METHODS
Participants
All 154 Dutch rheumatologists, except for the six involved in the study, were invited to participate in this project. In total 117 agreed, but for a variety of reasons 17 did not start. At baseline, 100 (85% of the 117 consenting rheumatologists) participated, after self education 86 (74%), and after the workshop 75 (64%) (fig 1). In addition, a random sample of 30 consenting radiologists (out of the total number of 687 radiologists in the Netherlands) was taken from a list of members of the Dutch Radiology Society. The percentage of their daily work which consisted of viewing skeletal radiographs was as follows: for 3 radiologists <10%, 15 radiologists 10–30%, 10 radiologists 30–50%, and this was unknown for 2 radiologists. At baseline 23 (77% of the random sample of radiologists) participated, after self education 22 (73%), and after the workshop 8 (27%) (fig 1).

Radiographs of the sacroiliac joints
For this study, three different sets of radiographs of the SI joints were composed: a *scoring* set for the assessments, a *training* set for individual training, and a *confidence assessment* set to estimate the observer's perceived certainty in diagnosing

....

Rheumatologists

Radiologists
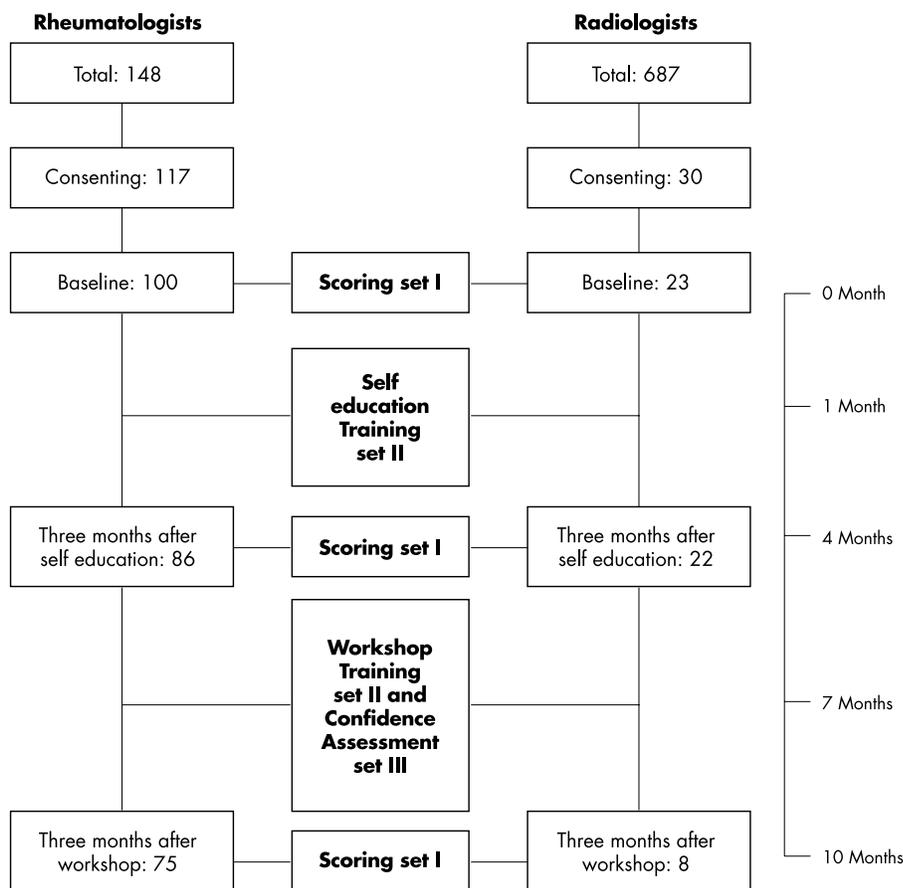
Figure 1 Flow chart of participants and examinations.

sacroiliitis by plain radiograph or CT of the SI joints. All radiographs were derived from a large Swiss family survey among 275 HLA-B27 positive patients with AS and 511 first degree relatives, who all completed questionnaires and underwent physical examination, HLA typing, and radiographic studies of the SI joints.[4] The radiographs were projected from an anteroposterior view, limited to the SI joints, and not including the hip joints. All radiographs had been scored twice "blindly" by each of four experts (two rheumatologists, one epidemiologist, and one radiologist). The mean score of the eight readings of each SI joint was rounded to the next whole figure. Only SI films of HLA-B27 positive patients with AS showing definite sacroiliitis (as defined by the expert panel) and fulfilling the modified New York criteria were included in this study to assess the observer's sensitivity. For the assessment of specificity we used the films of HLA-B27 negative first degree relatives of the HLA-B27 positive patients with AS. These relatives had no signs or symptoms suggestive of AS. In addition, a subset of radiographs with simultaneously taken CTs was selected. In the same way, the four experts had judged these CTs separately from the conventional plain SI films.

The New York scoring method for the SI joints was followed: 0=no abnormalities; 1=suspicious changes (no specific abnormalities); 2=minimal sacroiliitis (loss of definition at the edge of the SI joints, there is some sclerosis and perhaps minimal erosions, there may be some joint space narrowing); 3=moderate sacroiliitis (definite sclerosis on both sides of the joint, blurring and indistinct margins, and erosive changes with loss of joint space); 4=complete fusion or ankylosis of the SI joint (without any residual sclerosis).[5] According to the modified New York criteria at least grade 2 bilaterally or grade 3 or 4 unilaterally is necessary for the diagnosis of AS.[6] The gradings for the left and right SI joint were recoded into one final grading representing sacroiliitis according to the New York criteria.

The *assessment* or *scoring* set comprised 50 radiographs, of which 10 appeared twice (reversed—that is, the left joint was now marked as the right joint and vice versa): 12 (+3 repeats) radiographs with a final New York grading of "no sacroiliitis", 12 (+3 repeats) with a grading of "dubious abnormalities", 12 (+3 repeats) with bilateral definite sacroiliitis grade 2, and 4 films (+1 repeated) with a grading 3 or 4. For each radiograph, only the age and sex of the subject was provided. The mean age of the 16 patients with AS represented in the scoring set was 44.1 years; the mean disease duration was 14.9 years at the time the SI radiographs for this study were taken. The *training* set also comprised 50 radiographs (10 for each grading 0–4). For each radiograph, information on the grading was provided. The *confidence assessment* set was composed of 17 plain SI radiographs and 17 corresponding CTs with sacroiliitis and 9 SI/CT pairs without sacroiliitis. For these films no information on grading, age, sex, or clinical findings was provided. All radiographs in the *scoring* set and in the *confidence assessment* set appeared in a completely random order.

Assessments
In total three assessments took place with the scoring set (fig 1). At each of these three occasions all participants individually graded each SI joint of the 50 radiographs according to the New York criteria. Firstly, a baseline score of sensitivity and specificity was established with the scoring set. One month later, each participant received individually the training set in order to practice reading radiographs individually. Three months after this training by self education, the participants again received the scoring set to assess the presence of persisting effects of this training procedure.

Another three months later, the participants attended one of several workshops organised throughout the country (with a maximum of 20 participants per workshop), in which the

Table 1 Sensitivity and specificity of radiographs of the sacroiliac joints by rheumatologists and radiologists in comparison with a "gold standard". Results shown as percentages (SD)

	Rheumatologists			Radiologists		
	Mean (SD)	Median	Range	Mean (SD)	Median	Range
Sensitivity at baseline	80 (18)	81	31–100	84 (22)	88	25–100
Sensitivity after self education	74 (19)*	75	25–100	80 (19)	78	44–100
Sensitivity after workshop	79 (17)†	81	25–100	83 (17)	84	50–100
Specificity at baseline	75 (15)	75	38–100	71 (16)	71	46–100
Specificity after self education	78 (14)‡	79	29–100	70 (18)	73	38–96
Specificity after workshop	76 (13)	79	38–96	80 (16)	85	50–96

*Statistically significant decrease compared with baseline scores (p=0.001); †statistically significant increase compared with scores after self education (p=0.002); ‡statistically significant improvement compared with baseline scores (p=0.008) (all paired *t* test).
Differences between rheumatologists and radiologists were not statistically significant at any time.

full spectrum of normal and abnormal SI joints, and the grading of these joints according to the New York criteria, was shown and discussed with the participants. In this workshop the same training set as for individual self education was used. Finally, in the same workshop, the participants also judged another 26 radiographs of the confidence assessment set for the presence or absence of sacroiliitis (according to the modified New York criteria), and rated their self confidence about their "yes/no" judgment from 1 (maximally uncertain) to 10 (absolutely sure). The same exercise was done for the 26 CTs from the same patients. Half of the participants first assessed the plain SI joints, and the other half started with reading the 26 CT films. When all scoring data were completed all the radiographs and CTs were discussed referring to the judgments by the expert panel. Three months after this workshop, the final assessment took place with the *scoring* set as before.

All radiograph sets were sent by post to the participants, who were requested to read the radiographs within the next two weeks. If necessary, a reminder was sent. After the last measurement, all participants received feedback on their personal scores together with the aggregated (centiles) results of all participants.

Statistical analysis
After completion of each *scoring* set, the final grading for each radiograph was dichotomised into the presence or absence of sacroiliitis. Grade 2 or more bilaterally, and grade 3 or 4 unilaterally was taken as the presence of sacroiliitis according to the New York criteria. The results from each participant were compared with the "gold standard" as defined by the expert panel. Sensitivity and specificity for each participant were calculated using 2×2 tables for each assessment period. The results of sensitivity and specificity at baseline were compared with the results after self education and after the workshop by paired *t* test.

The mean intraobserver agreement (concordance rate) was calculated by means of 2×2 tables for both radiologists and rheumatologists for each of the 10 films that were presented in duplicate (although in reversed right-left order). The κ statistic was not applied. Repeated films were included only once in the analysis of sensitivity and specificity.

To calculate sensitivity and specificity the judgments on the presence or absence of sacroiliitis for the 26 SI radiographs and CTs were compared with the standard defined by the experts. Differences in sensitivity and specificity between SI radiographs and CTs were calculated with paired *t* tests. The ratings that the participants had provided for their self confidence were subdivided into two groups: films *correctly* (according to the expert panel) diagnosed for the presence or absence of sacroiliitis and films *incorrectly* diagnosed for the presence or absence of sacroiliitis. This was done for plain radiographs and CTs separately. The differences between ratings on correctly versus incorrectly diagnosed sacroiliitis, and the differences in

ratings of radiographs versus CTs, were analysed by independent *t* tests.

RESULTS
Sensitivity and specificity
Table 1 shows the mean sensitivity and specificity scores for both rheumatologists and radiologists at baseline before any training, after individual self education, and after the workshop. In general, the radiologists showed somewhat higher sensitivity scores and lower specificity scores than the rheumatologists, but the differences were not statistically significant at any time. The scores for the radiologists did not significantly change after self education or the workshop, possibly partly reflecting the smaller numbers involved, especially at the last assessment. In contrast, the rheumatologists showed a statistically significant decrease in sensitivity after individual self education, but at the same time a statistically significant improvement in specificity. The decrease in sensitivity was reversed after the workshop, because at that time a statistically significant increase compared with the results after self education was observed. The differences between sensitivity three months after the workshop and sensitivity at baseline were not statistically significant.

Because the differences in sensitivity and specificity between rheumatologists and radiologists were not statistically significant at each of the three assessments, further analyses were performed with pooled data from both groups.

Table 2 presents the distribution in sensitivity and specificity of all participants at each assessment period. Clearly, a relatively large group of participants had difficulties in diagnosing sacroiliitis (for example, sensitivity of ≤50% at baseline for 13% of the participants). Participants with a low sensitivity or a low specificity score appeared to show high specificity or sensitivity scores respectively (table 3). Similarly, high sensitivity and high specificity scores were associated with low specificity and low sensitivity scores, respectively.

Change in sensitivity and specificity for individual participants
Minor changes in sensitivity or specificity at the *group* level do not preclude larger changes for *individual* participants. Therefore, to find out what kind of changes in sensitivity and specificity occurred after self education or after the workshop for individual participants, we recoded the change scores of each participant dichotomously by considering up to ±5% change in sensitivity or specificity as "no important change" and all other changes above or below this cut off point as a "relevant change". Table 4 shows the results of the profiles of the participants three months after self education and three months after the workshop. Relevant increases in both sensitivity and specificity occurred only in a minority of the participants. Most of the participants showed an increase in either

Table 2 Distribution of sensitivity and specificity of both rheumatologists and radiologists at baseline, after self education, and after the workshop are presented. Results shown as percentages of participants

	Distribution						
	≤50%	50–60%	60–70%	70–80%	80–90%	90–100%	Total
Sensitivity							
At baseline	13	2	11	11	23	40	100
After self education	17	8	16	13	20	26	100
After workshop	7	6	13	19	20	35	100
Specificity							
At baseline	8	12	16	25	21	17	100
After self education	7	8	12	27	29	17	100
After workshop	6	4	13	36	26	16	100

Table 3 The corresponding score of either sensitivity or specificity is represented when respectively specificity or sensitivity is low or high. Results are shown as mean percentages (standard deviation)

	If sensitivity ≤50, then specificity:	If specificity ≤50, then sensitivity:	If sensitivity ≥90, then specificity:	If specificity ≥90, then sensitivity:
Sensitivity				
At baseline		99 (2)		59 (17)
After self education		99 (2)		50 (13)
After workshop		99 (3)		61 (17)
Specificity				
At baseline	90 (6)		61 (11)	
After self education	91 (5)		58 (13)	
After workshop	88 (7)		63 (10)	

sensitivity or specificity without a relevant change in specificity or sensitivity, respectively.

The participants with an improvement in sensitivity after self education compared with baseline of >5%, showed a mean (SD) sensitivity at baseline of 63 (20) (n=23), and those with a change of >5% after the workshop a mean baseline sensitivity of 64 (16) (n=23). Similarly, an improvement in specificity of >5% after self education and after the workshop was associated with lower specificity scores at baseline of 60 (11) (n=33) and 63 (10) (n=22), respectively, than the mean specificity scores of the whole group as shown before (table 1). Furthermore, the participants with a decrease of >5% in sensitivity after self education and the workshop compared with baseline, showed higher baseline scores in sensitivity (84 (14) (n=57) and 86 (15) (n=33), respectively), and those with a decrease of >5% in specificity showed higher baseline scores in specificity (77 (16) (n=28) and 84 (14) (n=23), respectively).

Intraobserver variation

Table 5 presents the intraobserver variation for both rheumatologists and radiologists for each of the 10 radiographs—together representing the spectrum of sacroiliitis—that were used for this assessment. No major differences in agreement between each of the repeated radiographs were found between rheumatologists and radiologists. The most extreme grades 0 and 4 showed the highest agreement. Sacroiliitis grades 1 and 2 showed substantial variation represented by lower agreement rates.

Plain radiograph compared with computed tomography

During the workshop each participant had to read independently in random order 26 radiographs and 26 CTs of the same patients for the presence or absence of sacroiliitis according to the New York criteria. In addition, for every judgment the participants also had to provide a rating on a 1–10 scale for their perceived self confidence for each of these diagnostic

Table 4 Profiles of both rheumatologists and radiologists of change after the interventions. Results are given as percentages of participants

	After self education compared with baseline scores (n=108)	After workshop compared with baseline scores (n=84)	After workshop compared with scores after self education (n=84)
Sensitivity and specificity remained equal	13	16	13
Sensitivity and specificity improved	1	4	4
Sensitivity improved, specificity decreased *or* remained equal	21	27	39
Specificity improved, sensitivity decreased *or* remained equal	31	26	21
Both sensitivity and specificity decreased *or* either decreased and the other remained equal	34	27	23
Total	100	100	100

"Equal" was defined as a change of up to 5%, "improved" as an increase of more than +5%, and "decreased" as a change of more than −5%.

Table 5 Intraobserver agreement for both rheumatologists and radiologists for each of 10 duplicated radiographs of sacroiliac joints showing different degrees of sacroiliitis as defined by an expert panel. Dubious abnormalities show lower concordance. The results are shown as mean percentages of agreement

Grading (sacroiliitis)	Rheumatologists			Radiologists		
	Baseline	After self education	After workshop	Baseline	After self education	After workshop
0 (no)	99	100	100	100	100	100
0 (no)	96	99	97	100	100	100
0 (no)	98	95	99	83	91	100
1 (dubious)	82	83	83	87	91	88
1 (dubious)	83	76	70	78	82	88
1 (dubious)	65	66	70	78	82	63
2 (yes)	70	76	76	83	52	75
2 (yes)	74	76	70	83	82	63
2 (yes)	83	80	87	100	86	88
4 (yes)	99	94	99	96	100	100

decisions. Table 6 presents the results of sensitivity and specificity, and the confidence ratings of the participants for both plain SI radiographs and CTs. The sensitivity score for CTs was significantly higher than for plain radiographs, whereas no difference in specificity was found.

DISCUSSION

The presence of sacroiliitis is mandatory for the diagnosis of AS. The SI joints are unilaterally or bilaterally affected with mild to severe inflammation, which may eventually lead to partial or complete ankylosis.[7] The recognition of sacroiliitis is, however, often considered as difficult and requires experience. In this study the performance of rheumatologists and radiologists in detecting sacroiliitis has been evaluated. Three features of this nationwide study—in which more than 50% of all Dutch rheumatologists and a small sample (4.4% at baseline; 1.2% at completion) of radiologists participated—are striking.

Firstly, the diagnosis of radiographic sacroiliitis by radiologists and rheumatologists was comparable. Secondly, sensitivity and specificity scores were relatively moderate: 15–25% of the radiographs were incorrectly classified as if sacroiliitis was present (false positives), and 20–30% of the radiographs were incorrectly classified as if sacroiliitis was absent (false negatives) (table 1). A high sensitivity score was associated with a low specificity score (table 3), and an increase in sensitivity was often accompanied with decreased specificity and vice versa (table 4). Thirdly, improvement in both sensitivity and specificity that will persist for at least three months after a training session appeared to be difficult to achieve. It is worrying that both sensitivity and specificity decreased in a large group of participants. Thus, the individual training sessions

and the workshops as provided cannot be regarded as effective in promoting the performance of "blindly" diagnosing the presence or absence of radiographic sacroiliitis. Although individual compliance was not assessed, non-compliance overall cannot explain the apparent lack of effects of the workshop which was attended by the majority (75%) of the participants.

It is difficult to explain these observations. It seems that an improvement or a decrease in sensitivity (or specificity) after a training session was associated with correspondingly lower or higher sensitivity (or specificity) scores at baseline as compared with the mean score of the group at baseline (table 3). This effect may be attributed to regression to the mean or a floor-ceiling effect. Possibly, after training sessions, the attitude towards interpreting radiographs might have changed. Participants with initially low sensitivity scores might now have considered every spot or blurring at SI joints as aberrant, thereby improving the sensitivity score, but at the cost of specificity. Conversely, participants with initially low specificity scores might now have considered every spot or blurring at the SI joint more cautiously, at the cost of sensitivity. However, the participants were not informed about their sensitivity and specificity scores during the study period. Clearly, even after training it remains difficult to distinguish between the normal and abnormal. Possibly, the same intervention should not have been offered to every participant. It might have been better to have assessed sensitivity and specificity first and then provide different targeted interventions to those participants with low sensitivity (and high specificity) and those with high sensitivity (and low specificity). This might be an area for future research.

The relative roles of plain radiographs, CT, and MRI in the radiographic diagnosis of sacroiliitis remain a matter of

Table 6 Sensitivity, specificity, and mean confidence ratings of participants in defining presence or absence of sacroiliitis on radiographs compared with computed tomography (CT) of the sacroiliac joints. Results are shown as means (standard deviation).

	Radiograph	CT
Sensitivity	72.0 (15.1)*	85.6 (14.1)*
Specificity	84.5 (13.8)	83.5 (17.0)
Mean rating, presence of sacroiliitis correctly diagnosed	7.7 (1.0)	7.8 (1.1)
Mean rating, presence of sacroiliitis incorrectly diagnosed	6.6 (1.2)	6.7 (1.3)
Mean rating, absence of sacroiliitis correctly diagnosed	8.3 (1.0)	8.0 (1.1)
Mean rating, absence of sacroiliitis incorrectly diagnosed	7.4 (1.4)	7.0 (1.6)

Sensitivity and specificity range from 0 to 100, the range of the ratings is from 1 to 10.
*p<0.001 with paired *t* test. The differences in ratings between incorrectly and correctly diagnosed presence or absence of sacroiliitis were all significant (p<0.001).
The presence or absence of sacroiliitis were all significant (p<0.001). The differences in ratings for radiographs and CTs were not statistically significant.

debate. The high sensitivity of CT and MRI is well known. Several studies have reported that CT and MRI are better than plain radiographs in detecting early sacroiliitis.[8–13] However, because of the cost and other limitations to resources it is not always possible to use these techniques in the routine diagnosis of sacroiliitis.[14 15] Therefore, plain SI radiographs remain mostly the initial diagnostic tool. CT or MRI may be particularly helpful as an additional diagnostic aid in the early stages of sacroiliitis (when plain radiographs may be negative) if there is a high probability of sacroiliitis, or conventional radiographs are inconclusive.

Owing to the difficulties in interpreting plain radiographs of SI joints, large inter- and intraobserver variations have been reported.[1–3] In our study, intraobserver variability was expressed as the percentage of agreement for each of 10 radiographs that appeared twice in the *scoring* set. Clearly, concordance is highest if SI joints are definitely normal (grade 0) or definitely abnormal (grade 4). The use of κ statistics would not have been useful in this situation because of the high levels of expected agreement. The amount of intraobserver variation was comparable for both rheumatologists and radiologists. Most variation was seen in grades 1 and 2. However, the diagnostic—and possibly therapeutic—consequences of such seemingly small differences in grading of SI joints are most important. Patients with grade 2 sacroiliitis bilaterally will usually be diagnosed and treated as having AS, whereas people with grade 1 will normally not be considered as having an inflammatory rheumatic disease. Especially in these cases, CT or MRI may be helpful.[16] It should be noted, however, that there are clear differences in properties—and therefore also in appropriateness of their application—among plain SI films, CT, and MRI. Plain radiographs provide an image where all sections are added to each other, whereas CT and MRI give information in slices. Furthermore, plain films and CT can assess mainly bone and bone destruction, whereas MRI can assess cartilage and inflammation in the acute stage. It should also be realised that AS might sometimes occur in the absence of radiographic sacroiliitis.[17]

Another aim of our study was to assess the degree of confidence of rheumatologists and radiologists in determining the presence or absence of sacroiliitis on plain radiographs and CTs of the same patients. The ratings for the *correctly* diagnosed presence or absence of sacroiliitis were on average higher than the *incorrectly* diagnosed presence or absence of sacroiliitis (p<0.001). Although the participants felt less certain about their judgments of those radiographs and CTs which they misdiagnosed than those which they correctly diagnosed, the ratings for the incorrectly diagnosed radiographs remained, somewhat surprisingly, relatively high (mean 6.6 versus 7.4 on a 0–10 scale) (table 6). Clearly, the use of CTs compared with the use of radiographs did not increase the self confidence of the participants. However, many rheumatologists felt they did not to have sufficient experience in reading CTs of SI joints and, therefore, these results might improve after training. On the other hand, the number of radiologists who participated in this part of the study is too small to generalise the findings.

The prevalence of definite sacroiliitis in the *scoring* set was 40%. This high a priori likelihood was unknown to the participants. It is unrealistically high in daily practice of radiologists, but on the other hand, diagnostic gain is at its highest level if the pretest probability is about 50%. Therefore, for rheumatologists this prevalence would indicate proficiency in making use of diagnostic tools. If a large number of normal (grade 0) SI films had been included in the *scoring* set this would have inflated the specificity artificially without clearly predictable effects on the sensitivity of diagnosing radiographic sacroiliitis.

Finally, except for data on the age and sex of the patient in the *scoring* set, no clinical findings from the patient's history or physical examination were provided. Therefore, only radio-

graphs were presented to radiologists and rheumatologists in order to evaluate their performance in detecting sacroiliitis. This radiological diagnosis is an indispensable condition for the diagnosis of AS. In daily practice, however, rheumatologists mostly take into consideration the clinical information of the patient before they come to a final judgment. Rheumatologists may decide to re-evaluate the patient at a later time, or to refer the patient for additional CT or MRI. Recently, a study has assessed the real performances of (Dutch) rheumatologists in daily practice, visited by patients incognito.[18] In particular, a female patient mimicking symptoms suggestive of AS and referred by her general practitioner visited a total of 25 rheumatologists. She brought with her a radiograph from another hospital clearly showing bilateral sacroiliitis. After history taking and physical examination, in which nearly all rheumatologists performed spinal mobility tests, more than 50% of the rheumatologists proposed additional radiographic imaging.[18] Evidently, a large group of rheumatologists felt uncertain about interpreting radiographs. Unfortunately, this study does not seem to contribute towards increasing their performance. It should again be emphasised that our study assessed sensitivity, specificity, and observer variation in reading films of SI joints, but did not take into consideration the effects on these characteristics of any clinical information. Such clinical data might already be known before reading the films or may be provided to the observer afterwards. The final effect of such additional information on the precision of establishing sacroiliitis as an indispensable condition for the diagnosis of AS is not yet known.

In conclusion, longlasting improvements in the performance of diagnosing sacroiliitis seem difficult to achieve, at least through self education using a training set of SI films or through uniform workshops. No statistically significant differences in sensitivity, specificity, and intraobserver variation of reading radiographs of SI joints were found between the radiologists and rheumatologists. Currently, CT of SI joints as compared with plain SI radiographs does not improve self confidence in diagnosing sacroiliitis.

. . . .

Authors' affiliations
A van Tubergen, L Heuft-Dorenbosch, G Schulpen, R Landewé, R Wijers, D van der Heijde, J van Engelshoven, Sj van der Linden, University Hospital Maastricht, The Netherlands

REFERENCES

1 **Hollingsworth PN**, Cheah PS, Dawkins RL,Owen ET, Calin A, Wood PH. Observer variation in grading sacroiliac radiographs in HLA-B27 positive individuals. J Rheumatol 1983;10:247–54.
2 **Yazici H**, Turunc M, Ozdogan H, Yurdakul S, Akinci A, Barnes CG. Observer variation in grading sacroiliac radiographs might be a cause of 'sacroiliitis' reported in certain disease states. Ann Rheum Dis 1987;46:139–45.
3 **Taylor HG**, Wardle T, Beswick EJ, Dawes PT. The relationship of clinical and laboratory measurements to radiological change in ankylosing spondylitis. Br J Rheumatol 1991;30:330–5.
4 **van der Linden S**, Khan MA, Rentsch HU,Gerber N, Cats A, Valkenburg HA, *et al*. Chest pain without radiographic sacroiliitis in relatives of patients with ankylosing spondylitis. J Rheumatol 1988;15:836–9.
5 **Dale K**. Radiographic gradings of sacroiliitis in Bechterew's syndrome and allied disorders. Scand J Rheumatol Suppl 1979;32:92–7.
6 **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.
7 **Braun J**, Bollow M, Sieper J. Radiologic diagnosis and pathology of the spondyloarthropathies. Rheum Dis Clin North Am 1998;24:697–735.
8 **Ryan LM**, Carrera GF, Lightfoot RW, Hoffman RG, Kozin F. The radiographic diagnosis of sacroiliitis. A comparison of different views with computed tomograms of the sacroiliac joint. Arthritis Rheum 1983;26:760–3.

9 **Docherty P**, Mitchell MJ, MacMillan L,Mosher D, Barnes DC, Hanly JG. Magnetic resonance imaging in the detection of sacroiliitis. J Rheumatol 1992;19:393–401.

10 **Battafarano DF**, West SG, Rak KM, Fortenbery EJ, Chantelois AE. Comparison of bone scan, computed tomography, and magnetic resonance imaging in the diagnosis of active sacroiliitis. Semin Arthritis Rheum 1993;23:161–76.

11 **Braun J**, Bollow M, Eggens U,Konig H, Distler A, Sieper J. Use of dynamic magnetic resonance imaging with fast imaging in the detection of early and advanced sacroiliitis in spondylarthropathy patients. Arthritis Rheum 1994;37:1039–45.

12 **Blum U**, Buitrago Tellez C, Mundinger A, Krause T, Laubenberger J, Vaith P, et al. Magnetic resonance imaging (MRI) for detection of active sacroiliitis—a prospective study comparing conventional radiography, scintigraphy, and contrast enhanced MRI. J Rheumatol 1996;23:2107–15.

13 **Yu W**, Feng F, Dion E,Yang H, Jiang M, Genant HK. Comparison of radiography, computed tomography and magnetic resonance imaging in the detection of sacroiliitis accompanying ankylosing spondylitis. Skeletal Radiol 1998;27:311–20.

14 **Murphey MD**, Wetzel LH, Bramble JM,Levine E, Simpson KM, Lindsley HB. Sacroiliitis: MR imaging findings. Radiology 1991;180:239–44.

15 **Fenton P**. Magnetic resonance imaging of the sacroiliac joints: worth the cost? J Rheumatol 1996;23:2020–1.

16 **Braun J**, Sieper J, Bollow M. Imaging of sacroiliitis. Clin Rheumatol 2000;19:51–7.

17 **Khan MA**, van der Linden SM, Kushner I, Valkenburg HA, Cats A. Spondylitic disease without radiologic evidence of sacroiliitis in relatives of HLA-B27 positive ankylosing spondylitis patients. Arthritis Rheum 1985;28:40–3.

18 **Gorter S**, van der Linden S, Brauer J, van der Heijde D, Houben H, Rethans JJ, et al. Rheumatologists' performance in daily practice. Arthritis Rheum 2001;45:16–27.