

Observer variation in grading sacroiliac radiographs might be a cause of 'sacroiliitis' reported in certain disease states

H YAZICI,¹ M TURUNÇ,¹ H ÖZDOĞAN,¹ S YURDAKUL,¹
A AKINCI,² AND C G BARNES³

From the Divisions of ¹Rheumatology and ²Radiology, Department of Medicine, Cerrahpaşa Medical Faculty, University of Istanbul; and the ³Department of Rheumatology, The London Hospital

SUMMARY Radiological sacroiliitis in Behçet's syndrome (BS) has been a subject of controversy. We have examined pelvic radiographs of 38 patients with BS and 28 age and sex matched controls which we reported previously, and also 17 with ankylosing spondylitis (AS), 27 with non-renal familial Mediterranean fever (FMF), and 33 with primary osteoarthritis (OA). Initially, five observers assessed radiographs on two different occasions according to the New York criteria for sacroiliitis in a blind protocol. Later, three of them examined the various possible abnormalities of the sacroiliac (SI) joints after training sessions. Although the inter- and intraobserver variation was quite high, all observers found the expected changes in patients with AS. The abnormalities detected in the other diseases were either mild, inconsistent, or both. Erosions were confined to patients with AS, and osteophytes and glenoid sulci to patients with OA. We conclude that high observer variation in interpreting a film of the anteroposterior (AP) view of the pelvis for sacroiliitis may be a major cause of reported 'sacroiliitis' in BS and FMF.

Key words: Behçet's syndrome, familial Mediterranean fever, New York criteria, ankylosing spondylitis, osteoarthritis, pelvic radiographs.

It has been reported that Behçet's syndrome (BS) may be associated with sacroiliitis.^{1,2} In the only controlled survey of sacroiliitis in BS we showed that sacroiliac joint changes were not more common than in controls.³ Controversy continues,⁴ however. This study tries to elucidate the cause(s) of this controversy. We have examined the same radiographs as in the first study³ with the following additions and modifications: we have increased the number and the kind of diseased controls, we have increased the number of observers, and finally, we made an attempt to analyse the various radiographic components of the SI joint separately.

Materials and methods

The diagnosis, number, and ages of the patients studied are shown in Table 1. Patients with ankylosing spondylitis (AS) and familial Mediterranean

fever (FMF) were regular attenders at the rheumatology clinic of the Cerrahpaşa medical faculty. Patients with FMF were selected for the absence of renal disease. Patients with osteoarthritis (OA) were drawn from an ongoing study on the distribution of primary OA, and represented patients with the involvement of the hands, hips, and/or knees.

Standard AP radiographs of the pelvis were available. The SI joints were read 'blind' to the diagnosis and age of the proband.

Five observers took part in the first phase of the study. Three of the observers were Turkish rheumatologists (HY, SY, HÖ), the fourth an English rheumatologist (CGB), and the fifth (AA) was a radiologist. CGB and HY were the more senior members. AA was a general radiologist with interest but no subspecialty training in bone and joint radiology. The observers did not have any preliminary discussion together but were told immediately before the films were read of the criteria to be used. They read the radiographs twice, on two successive days. The films were graded according to the New

Table 1 Patient population and availability of radiographs

Diagnosis	No of patients	Male:female	Age (years)*	Radiographs available	
				1st phase	2nd phase
Behçet's syndrome	38	32:6	34 (11)	35	29
Ankylosing spondylitis	17	12:5	30 (10)	15	10
Familial Mediterranean fever	27	18:9	29 (11)	25	22
Osteoarthritis	33	31:2	58 (9)	31	29
Controls†	28	24:4	30 (7)	22	16

*Values are mean (SD).

†Drawn from hospital staff.

York criteria⁵ on a 0–4 scale (0=normal, 1=suspicious, 2=possible or minimal, 3=moderate changes, and 4=ankylosis of the SI joint).

After evaluation of the first phase of the study and when the results were known to the observers three (HY, SY, and AA) took part in the second phase. These observers had several sessions together to reach a consensus on various aspects of the radi-

ology of the SI joint (listed in Table 3). Grading according to the New York criteria⁵ was also included. Radiographs used in these training sessions were omitted from the subsequent reading. Each observer completed the total reading of the films, again blind to the disease category and age, separately in about two months. χ^2 Analyses and weighted κ statistics⁶ were used to analyse the

Table 2 Grades 0 and 1 (a) versus grades 2–4 (b); tabulation of findings in phase 1 according to each observer

Observers		Normals		FMF		BS		OA		AS	
		I	II	I	II	I	II	I	II	I	II
HY	a	44	40	50	50	68	65	60	51	10	9
	b	0	4	0	0	2	5	2	11	20	21
CGB	a	43	44	44	45	63	61	59	57	7	4
	b	1	0	6	5	7	9	3	5	23	26
SY	a	38	30	39	33	41	43	30	17	4	5
	b	6	14	11	17	29	27	32	45	26	25
HÖ	a	33	38	33	32	39	55	34	59	3	4
	b	11	6	17	18	31	15	28	3	27	26
AA	a	18	31	12	31	38	46	6	19	0	0
	b	26	13	38	19	32	24	56	43	30	30

The numbers refer to the number of sacroiliac joints; I and II=first and second readings.

Table 3 Grades 0–2 (a) versus grades 3 and 4 (b); tabulation of findings in phase 1

Observers		Normals		FMF		BS		OA		AS	
		I	II	I	II	I	II	I	II	I	II
HY	a	44	44	50	50	68	69	60	60	10	10
	b	0	0	0	0	2	1	2	2	20	20
CGB	a	43	44	44	45	63	61	59	57	7	4
	b	1	0	6	5	7	9	3	5	23	26
SY	a	44	42	44	42	63	65	57	59	5	5
	b	0	2	6	8	7	5	5	3	25	25
HÖ	a	43	42	40	38	59	65	49	59	6	4
	b	1	2	10	12	11	5	13	3	24	26
AA	a	29	40	20	39	44	58	37	55	2	1
	b	15	4	30	11	26	12	25	7	28	29

The numbers refer to the number of sacroiliac joints; I and II=first and second readings.

Table 4 Stepwise χ^2 analysis of the findings in phase 1

Observers		Normals		FMF		BS		OA		AS	
		I	II	I	II	I	II	I	II	I	II
HY	a	-	-	-	-	-	-	++	+	+++	+++
	b	-	-	-	-	-	-	-	-	+++	+++
CGB	a	-	-	-	-	-	+	-	-	+++	+++
	b	-	-	-	-	-	-	-	-	+++	+++
SY	a	-	-	-	-	++	-	+++	+++	+++	+++
	b	-	-	-	-	-	-	-	-	+++	+++
HÖ	a	-	-	-	-	-	-	-	-	+++	+++
	b	-	-	-	++	-	-	-	-	+++	+++
AA	a	-	-	+	-	+++	-	+++	+++	+++	+++
	b	-	-	+	-	-	-	-	-	+++	+++

See text for the statistical analyses. I and II= first and second readings.
 a=grades 0+1 versus grades 2+3+4; b=grades 0+1+2 versus grades 3+4.

findings. The right and left SI joints were graded and evaluated separately in all calculations.

Results

During the first phase of the study a number of films were excluded because of their poor quality; six films from the normal group, three from BS, two from AS, two from FMF, and two from OA groups. The number of radiographs evaluated in the second phase (normals 16, FMF 22, BS 29, OA 29, AS 10) were still less because of the omission of films used during the training sessions (Table 1).

Tables 2 and 3 show the results of the first part of the study for each observer. In Table 2 the results are shown as the number of SI joints interpreted as showing no change or suspicious changes (grades 0+1) combined versus minimal, moderate, and severe changes (grades 2-4) combined. In Table 3 grade 2 is included with grades 0 and 1 and the results given as grades 0+1+2 versus 3+4. Table 4 is a χ^2 analysis of Tables 2 and 3. Initially a 5x2 χ^2 analysis was performed for each observer for the separate readings in each disease category. After this the disease group that was thought to cause the significant χ^2 value was removed from the calculation (always leaving the normal group in) and a 4x2 χ^2 analysis was performed, and this was continued until the χ^2 value became non-significant ($p > 0.05$) for that particular degree of freedom. A score of +++ was given for a $p < 0.001$, ++ for a $p < 0.01$, and + for a $p < 0.05$.

It is to be noted that all observers in both readings found highly significant sacroiliitis among the patients with AS whichever way the results were analysed. Three of the five observers reported SI

joint changes in the patients with OA when the results were analysed as grades 0+1 versus 2+3+4. This significance disappeared when the minimal sacroiliitis group was included with grades 0 and 1. None of the observers found sacroiliac changes in BS in the analysis of grades 0+1+2 versus 3+4, whereas three of the five did in the analysis of grades 0+1 versus 2+3+4, but only in one of their readings. Finally, for the FMF group, only 2/5 observers found sacroiliac changes, and again this was observed in only one session for either observer.

Tables 5, 6, and 7 give the results of the readings of each of the three observers in the second phase of the study. The results are tabulated in two categories (a and b) as in the first phase for each disease and for each radiological feature. Table 8 shows χ^2 analyses of the same data. For each type of SI lesion and for the individual features of SI joint pathology, stepwise χ^2 analyses were performed as for phase 1. It is to be noted that all observers again recorded significant SI changes in AS. Only one observer considered that there were significant changes according to the New York criteria⁵ in a disease category other than AS (observer SY in OA). Erosions were not noted in significant numbers outside the AS group by any observer. Osteophytes and glenoid sulci, as noted by all three, were limited to radiographs of the OA group.

Table 9 gives the weighted κ analyses for intra-observer variation in the first phase of the study. Table 10 shows the interobserver variations between the second readings in the first phase and the readings in the second phase (according to the New York criteria) of the three observers (HY, SY, and AA) who took part in both phases. Finally, Table 11 shows the weighted κ values for interobserver

Table 5 Findings in phase 2 according to HY**†

Diagnosis	ER	SSS	ISS	LSSS	LISS	JN	OP	GS	CC	NYC									
	+	-	+	-	+	-	+	-	+	-									
Normals	a	2	30	3	29	1	31	1	31	2	30	-	-	-	12	20	4	28	
	b	1	31	1	31	1	31	1	31	1	31	3	29	1	31	1	31	3	29
FMF	a	1	43	1	43	1	43	1	43	3	41	-	-	-	-	11	33	1	43
	b	1	43	1	43	1	43	1	43	1	43	2	42	4	40	1	43	1	43
BS	a	1	57	2	56	1	57	1	57	5	53	-	-	-	-	11	47	6	52
	b	1	57	1	57	1	57	1	57	1	57	11	47	5	53	1	57	1	57
OA	a	2	56	9	49	7	51	3	55	1	57	10	48	-	-	5	53	11	47
	b	1	57	1	57	1	57	1	57	1	57	25	33	18	40	1	57	2	56
AS	a	11	9	9	11	9	11	5	15	2	18	6	14	-	-	1	19	14	6
	b	7	13	7	13	4	16	1	19	1	19	2	18	1	19	1	19	12	8

*Numbers refer to the number of SI joints.
 †For each aspect of the SI joint a scale of 0-4 was used where 0 indicated normality. 1 suspicious changes, 2-4 increasing degrees of severity. For osteophytes and glenoid sulci only three grades (0=absent, 1=suspicious, 2=present) were used.
 ER=erosions; SSS=sacral subchondral sclerosis; LISS=lateral sacral subchondral sclerosis (lateral implies extending well into the bone); LISS=lateral iliac subchondral sclerosis; JN=joint narrowing; OP=osteophytes; GS=glenoid sulci; CC=capsular calcification; NYC=according to New York criteria.
 a=grades 0+1 versus 2+3+4; b=grades 0+1+2 versus 3+4 (0+1 versus 2 for OP and GS).
 - =absent (i.e. grades 0+1 or 0+1+2); + =present (i.e. grades 2+3+4 or 3+4).

Table 6 Findings in phase 2 according to SY*

Diagnosis	ER	SSS	ISS	LSSS	LISS	JN	OP	GS	CC	NYC									
	+	-	+	-	+	-	+	-	+	-									
Normals	a	1	31	10	22	8	24	2	30	1	31	5	27	-	-	20	12	11	21
	b	1	31	1	31	1	31	1	31	1	31	1	31	5	27	7	25	2	30
FMF	a	4	40	16	28	18	26	12	32	2	42	3	41	-	-	37	7	23	21
	b	2	42	2	42	2	42	1	43	1	43	2	42	1	43	15	29	6	38
BS	a	1	57	20	38	14	44	12	46	5	53	2	56	-	-	46	12	20	38
	b	1	57	4	54	2	56	1	57	1	57	1	57	2	56	31	27	4	54
OA	a	8	50	45	13	32	26	11	47	4	54	10	48	-	-	37	21	47	11
	b	4	54	2	56	1	57	1	57	1	57	1	57	12	46	41	17	2	56
AS	a	16	4	10	10	14	6	19	1	8	12	11	9	-	-	9	11	19	1
	b	14	6	14	6	7	13	12	8	6	14	9	11	1	19	5	15	3	17

*See footnotes to Table 5.

Table 7 Findings in phase 2 according to AA*

Diagnosis	ER	SSS	ISS	LSSS	LSSS	JN	OP	GS	CC	NYC						
Normals	a	1	31	9	23	5	27	1	31	1	31	1	24	8	24	
	b	1	31	3	29	3	29	1	31	1	31	1	31	1	31	3
FMF	a	3	41	21	23	12	32	13	31	3	41	3	41	24	20	18
	b	1	43	4	40	1	43	1	43	1	43	1	43	1	43	7
BS	a	8	50	22	36	11	47	4	54	1	57	5	53	22	36	17
	b	1	57	1	57	1	57	1	57	1	57	1	57	1	57	6
OA	a	10	48	29	29	24	34	6	52	1	57	10	48	14	44	28
	b	2	56	5	53	1	57	3	55	1	57	1	57	1	57	1
AS	a	19	1	18	2	10	10	13	7	1	19	6	14	6	14	19
	b	12	8	16	4	2	18	4	16	1	19	3	17	5	19	3

*See footnotes to Table 5.

variation for the different features of SI joint abnormalities and for the New York criteria in the second phase. It is to be noted that although most of these weighted κ values are better than chance agreements, only a few are 'substantial' if one uses similar guidelines in interpreting a 'weighted κ ' as one does for κ values.⁶

Discussion

Our results show that only patients with ankylosing spondylitis have appreciable radiological SI joint changes when compared with normal controls. The changes observed in the SI joints in other diseases were either mild, inconsistent, or both.

Our results in patients with BS confirm previous findings by ourselves and others.^{7, 8} In neither phase of the study were consistently significant SI changes in BS observed.

We were surprised to find similar negative results in patients with FMF, which has also been reported to be associated with sacroiliitis in several studies.⁹⁻¹¹ Three points need to be considered, however: (a) none of the earlier studies were performed blind, or were controlled; (b) one of the studies¹¹ was conducted in children, where it is difficult to interpret radiographic sacroiliitis¹²; and (c) FMF frequently may be complicated by renal failure, which in turn can produce SI joint changes.¹² None of these studies quote how many patients had renal failure. On the other hand, we specifically excluded patients with renal involvement.

Our inconsistent, and generally negative, results of SI joint involvement in OA are most probably related to the insensitivity of the New York criteria for osteoarthritic changes. These were specifically designed for AS and do not include osteophytes. All three observers in the second phase of this study reported significant osteophytosis in patients with OA only. Glenoid sulci appear to be another feature of OA in sacroiliac joints, which may be related to aging.

There were considerable intra- and interobserver variations in our study (Tables 9, 10 and 11), with AA, the only radiologist in the study, having the highest degrees of intraobserver discordance (Table 9). Moreover, his scores generally leaned towards higher grades of sacroiliitis when compared with those of other observers (Tables 2, 3, 5, 6, and 7). It should be stressed, however, that even with these pronounced inter- and intraobserver variations and a systemic bias of reading higher degrees of joint changes by AA, the changes of AS were apparent to all (Tables 4 and 8). One would have expected higher concordance (both intra- and inter-) in reading films from normals and those with AS. This

Table 8 Stepwise χ^2 analysis of the findings in phase 2*†

		Normals			FMF			BS			OA			AS		
		HY	SY	AA	HY	SY	AA	HY	SY	AA	HY	SY	AA	HY	SY	AA
ER	a b													+++	+++	+++
SSS	a b										+	+++		+++	+++	+++
ISS	a b								++				++	+++		
LSSS	a b													+++	+++	+++
LISS	a b														+++	+++
JN	a b													+	+++	+++
OP	a b										+++	+		+++		
GS	a b										+++	+++	+++			
CC	a b								+							+
NYC	a b												+++	+++	+++	+++

*See text for statistical analysis.

†See footnotes to Table 5.

indeed was the experience of another group of investigators when reading films of HLA-B27 positive subjects.¹³ Although this was not true for

Table 9 The weighted κ analyses for intraobserver variation in phase 1

Diagnosis	Observers				
	HY	SY	HÖ	AA	CGB
Normals	0.66	0.59	0.40	0.24	0.20
FMF	0.19	0.49	0.46	0.24	0.57
BS	0.50	0.46	0.49	0.62	0.60
OA	0.51	0.37	0.09	0.24	0.48
AS	0.49	0.83	0.59	0.07	0.54
Total	0.67	0.61	0.47	0.48	0.65

Table 10 The weighted κ analyses for interobserver variation in phase 1 and phase 2

Observers	Phase 1	Phase 2
HY-SY	0.32	0.19
HY-AA	0.32	0.36
AA-SY	0.44	0.41

intraobserver variation in the first phase, this effect could be seen in the interobserver variation during the second phase when the highest weighted κ values were recorded while reading films of normal subjects. Interestingly, previous training sessions, with subsequent exclusion of films used in those sessions during the formal readings, did not seem to improve interobserver variability (Table 10). We are unaware of previous studies on this issue. We conclude that there are no consistent or appreciable SI joint changes in the diseases we have studied other than in AS when criteria specifically designed for the recognition of sacroiliitis in this disease are used.

The lack of sensitivity and specificity and the great observer variability in interpreting SI joints on a standard AP radiograph of the pelvis preclude any definite conclusions being reached. We agree with a previous statement that 'the results also call into question the validity of radiological "sacroiliitis" as distinct from AS.¹³

We acknowledge with thanks the secretarial assistance of Emine (Evcimen) Öztürk and Yusuf Yazici.

Table 11 Weighted κ values for interobserver variation in the second phase*

Diagnosis	Observers	ER	SSS	ISS	LSSS	LISS	JN	OP	GS	CC	NYC
Normals	HY-AA	0.50	0.33	0.43	0.18	0	0.47	0.50	0	0.30	0.55
	HY-SY	0.55	0.34	0.42	0.04	0	0.42	0.83	0	0.30	0.59
	AA-SY	0.48	0.76	0.64	0.01	0.04	0.42	0.78	0.06	0.15	0.71
FMF	HY-AA	0.19	0.20	0.05	0	0	0.04	0.39	0.67	0.29	0.09
	HY-SY	0.11	0.84	0.05	0	0	0.11	0.48	0.39	0.20	0
	AA-SY	0.41	0.32	0.33	0.35	0.35	0.27	0.40	0.25	0.07	0.24
BS	HY-AA	0.15	0.14	0.14	0.02	0	0.20	0.39	0.36	0.10	0.20
	HY-SY	0.20	0.12	0.12	0.06	0	0.12	0.31	0.16	0.07	0.29
	AA-SY	0.06	0.35	0.26	0.04	0.05	0.27	0.14	0.22	0.06	0.31
OA	HY-AA	0.36	0.20	0.31	0.28	0.18	0.40	0.21	0.69	0.06	0.29
	HY-SY	0.38	0.42	0.33	0.12	0.07	0.24	0.35	0.31	0.02	0.12
	AA-SY	0.42	0.44	0.53	0.58	0.20	0.24	0.21	0.36	0.09	0.30
AS	HY-AA	0.30	0.24	0.27	0.23	0.30	0.34	0.01	0	0	0.25
	HY-SY	0.70	0.20	0.39	0.19	0.07	0.14	0.24	0	0	0.21
	AA-SY	0.30	0.73	0.15	0.20	0.13	0.24	0.30	0.50	0.37	0.40

*See footnotes to Table 5.

References

- Dilşen N, Koniçe M, Övül C. Arthritic patterns in Behçet's disease. In: Dilşen N, Koniçe M, Övül C, eds. *Behçet's disease*. Amsterdam, Oxford: Excerpta Medica, 1979; 145-55.
- Moll J M H, Haslock I, Macrae I F, Wright V. Associations between ankylosing spondylitis, psoriatic arthritis, Reiter's disease, the intestinal arthropathies and Behçet's syndrome. *Medicine (Baltimore)* 1974; **53**: 343-64.
- Yazici H, Tuzlaci M, Yurdakul S. A controlled survey of sacroiliitis in Behçet's disease. *Ann Rheum Dis* 1981; **40**: 558-9.
- Dilşen N, Koniçe M, Aral M. Why Behçet's disease (BD) should be accepted as a seronegative arthritis. In: Lehner T, Barnes C G, eds. *Recent advances in Behçet's disease*. London: Royal Society of Medicine, 1986. (International Congress and Symposium Series; No 103.)
- Bennett P H, Wood P H N. *Population studies of the rheumatic diseases*. Excerpta Medical Foundation, 1966: 456. (International Congress Series 148.)
- Kramer M S, Feinstein A R. Clinical biostatistics LIV. The biostatistics of concordance. *Clin Pharmacol Ther* 1981; **29**: 111-23.
- Chamberlain M A. Behçet's syndrome in 32 patients in Yorkshire. *Ann Rheum Dis* 1977; **36**: 491-9.
- Shimizu T, Ehrlich G E, Inaba G, Hayashi K. Behçet's disease (Behçet's syndrome). *Semin Arthritis Rheum* 1979; **8**: 223-60.
- Shahin N, Sohar E, Dalith F. Roentgenologic findings in familial Mediterranean fever. *AJR* 1960; **84**: 269-74.
- Bradey P, Wlof S M. Radiographic changes in the sacroiliac joints in familial Mediterranean fever. *Radiology* 1975; **114**: 331-3.
- Lehman T J A, Harson V, Kornreich, Peters R S, Schwabe A D. HLA B27 negative sacroiliitis: a manifestation of familial Mediterranean fever in childhood. *Pediatrics* 1978; **61**: 423-6.
- Bellamy N, Park W, Rooney P J. What do we know about the sacroiliac joint. *Semin Arthritis Rheum* 1983; **12**: 282-313.
- Hollingsworth P N, Cheak P S, Dawkins R L, Owen E T, Calin A, Wood P H N. Observer variation in grading sacroiliac radiographs in HLA B27 positive individuals. *J Rheumatol* 1983; **10**: 247-54.