

ORIGINAL ARTICLE

Diagnostic accuracy of inflammatory back pain for axial spondyloarthritis in rheumatological care

Denis Poddubnyy,¹ Johanna Callhoff,² Inge Spiller,¹ Joachim Listing,² Juergen Braun,³ Joachim Sieper,¹ Martin Rudwaleit⁴

To cite: Poddubnyy D, Callhoff J, Spiller I, *et al.* Diagnostic accuracy of inflammatory back pain for axial spondyloarthritis in rheumatological care. *RMD Open* 2018;**4**:e000825. doi:10.1136/rmdopen-2018-000825

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/rmdopen-2018-000825>).

Received 22 September 2018
Revised 28 October 2018
Accepted 29 October 2018

ABSTRACT

Objective Inflammatory back pain (IBP), the key symptom of axial spondyloarthritis (axSpA), including ankylosing spondylitis, has been proposed as a screening test for patients presenting with chronic back pain in primary care. The diagnostic accuracy of IBP in the rheumatology setting is unknown.

Methods Six rheumatology centres, representing secondary and tertiary rheumatology care, included routinely referred patients with consecutive chronic back pain with suspicion of axSpA. IBP (diagnostic test) was assessed in each centre by an independent (blinded) rheumatologist; a second (unblinded) rheumatologist made the diagnosis (axSpA or no-axSpA), which served as reference standard.

Results Of 461 routinely referred patients, 403 received a final diagnosis. IBP was present in 67.3%, and 44.6% (180/403) were diagnosed as axSpA. The sensitivity of IBP according to various definitions (global judgement, Calin, Berlin, Assessment of SpondyloArthritis international Society criteria for IBP) was 74.4%–81.1% and comparable to published figures, whereas the specificity was unexpectedly low (25.1%–43.9%). The resulting positive likelihood ratios (LR+) were 1.1–1.4 and without major differences between sets of IBP criteria. The presence of IBP according to various definitions increased the probability of axSpA by 2.5%–8.4% only (from 44.6% to 47.1%–53.0%).

Conclusions The diagnostic utility of IBP in the rheumatology setting was smaller than expected. However, this was counterbalanced by a high prevalence of IBP among referred patients, demonstrating the effective usage of IBP in primary care as selection parameter for referral to rheumatology. Notably, this study illustrates potential shifts in specificity and LR+ of diagnostic tests if these tests are used to select patients for referral.

INTRODUCTION

Inflammatory back pain (IBP) is the key clinical symptom of axial spondyloarthritis (axSpA), including non-radiographic axSpA (nr-axSpA) and ankylosing spondylitis (AS, r-axSpA), and is often present at disease onset. IBP describes a spectrum of symptoms

Key messages

What is already known about this subject?

- Inflammatory back pain (IBP), the key symptom in axial SpA, is used for diagnostic purposes.
- Despite, its diagnostic accuracy in the rheumatology setting according to high level standards (e.g. STARD, Standards for Reporting of Diagnostic Accuracy) is unknown.

What does this study add?

- In the rheumatology setting IBP was a common finding among referred patients suggesting the effective use of IBP as selection parameter for referral in primary care.
- As a result, the specificity of IBP was lower and the diagnostic gain of IBP was smaller than expected.
- In the rheumatology setting no striking differences among defined IBP criteria (Calin, Berlin, ASAS) were found, yet the Calin criteria were least specific.

How might this impact on clinical practice?

- The study shows global judgement on the presence or absence of IBP as done in routine rheumatological care is influenced by knowledge of other SpA features.
- Rheumatologists should realize that diagnostic test characteristics such as specificity and likelihood ratios vary if the diagnostic test is used as selection parameter in primary care.

including (1) insidious onset of back pain, (2) morning stiffness in the lower back, (3) improvement of back pain with exercise, (4) no improvement with rest, (5) awakening at night or early morning because of back pain and (6) alternating buttock pain. Defined sets of IBP criteria are the Calin criteria,¹ Berlin criteria² and the Assessment of SpondyloArthritis international Society (ASAS) IBP experts criteria³; the latter was applied in the ASAS classification criteria for axSpA study.^{4,5} The published sensitivities (70.1%–95%) and specificities (72.5%–81.3%) for various sets of IBP criteria reveal calculated



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Martin Rudwaleit;
martin.rudwaleit@klinikumbielefeld.de

positive likelihood ratios (LR+), a measure of diagnostic utility, of 2.9–3.9. Accordingly, a LR+ of 3–4 for IBP has been proposed as best estimate for diagnostic purposes in daily practice.⁶ Assuming a background prevalence of 5% of axSpA among patients with chronic back pain in primary care, it has further been estimated that the presence of IBP increases the likelihood of axSpA by 9%–11% (from 5% to 14%–16%).⁶ IBP has also been proposed⁷ and has been proven effective as a parameter for selecting patients with chronic back pain in primary care for referral to rheumatology.^{7–13} No diagnostic accuracy study on IBP in the rheumatology setting is available to date.

METHODS

Study centres

The Diagnostic Accuracy of Inflammatory Back pain study (DIVERS) was designed according to recommendations in the field of diagnostic studies including the Standards for Reporting of Diagnostic Accuracy recommendations.^{14 15} DIVERS was conducted in six rheumatology centres across Germany: four general rheumatology practices (secondary care) distributed across the country, one large hospital for rheumatic diseases and one university hospital (tertiary care).

Patients

Eligible patients were routinely referred to rheumatology, had to have chronic back pain of ≥ 3 months and age at onset of ≤ 45 years and no clear diagnosis. Patients with a known diagnosis of axSpA were excluded. To avoid selection bias, participating centres were strongly encouraged to include consecutively all newly referred patients with chronic back pain with suspicion of axSpA. No specific referral strategies were set up for this study.

Study procedures

In each centre, two rheumatologists were involved: R-care and R-blind. R-care took routine care of the patient, ordered diagnostic tests as needed and made the final diagnosis of axSpA or no-axSpA. In contrast, R-blind took the clinical history regarding IBP features only, but was blinded to all other disease features otherwise. Both R-care and R-blind documented their findings independently of each on a prespecified case report form. The clinical diagnosis (axSpA or non-SpA) made by R-care served as reference standard. Since the study was conducted in routine rheumatology care, the sequence of consultation by either R-care or R-blind was left to the discretion of the participating centres and was driven primarily by feasibility aspects (R-care first 76%, R-blind first 24%).

In addition, patients were asked to complete a short self-reported questionnaire on IBP features in the rheumatologist's waiting room prior to the consultation, with answering modalities of 'yes' or 'no' to each IBP question.

The time period between first presentation to the rheumatology centre including the assessment of IBP by R-blind and final diagnosis (axSpA or no-axSpA) was in the range of 2–8 weeks. The first patient was included in March 2009 and the last patient in June 2010.

Study end points and data analysis

The diagnostic test of interest (presence/absence of IBP) was assessed by R-blind according to four definitions: (1) IBP by global judgement by the rheumatologist (ie, judgement on IBP independent of formal IBP criteria; yes/no), (2) Calin criteria, (3) Berlin criteria and (4) ASAS criteria for IBP. The global judgement (yes/no) on IBP was further categorised into 'uncertain', 'moderately confident', 'confident' and 'very confident'. Sensitivity, specificity, positive and negative LR (LR+ and LR-), positive and negative predictive values (PPV and NPV) with corresponding 95% CI of IBP and the net increase (%) in disease probability of axSpA were calculated. The agreement on the presence of IBP between R-blind and R-care was assessed by percentage agreement and Cohen's kappa; the latter interpreted according to the method of Landis and Koch.¹⁶

RESULTS

Of 476 eligible patients, 13 patients did not fulfil the inclusion criteria and were excluded. R-blind decided on the presence/absence of IBP in 461/463 patients. IBP by global judgement (R-blind) was present in 306/461 patients (66.4%). The level of confidence on IBP by R-blind in these 461 patients was as following: 'very confident' in 17.1% of patients, 'confident' in 62.5%, 'moderately confident' in 18.9% and 'uncertain' in 1.5%. The final clinical diagnosis by R-care (reference standard) was missing in four and considered uncertain in 54 patients. Thus, 403 patients with an assessment of IBP by R-blind (67.3% had IBP) and a definite diagnosis by R-care (180 with definite axSpA—88 with AS, 92 with nr-axSpA—and 223 with no-axSpA) were included in the final analysis (patient flow chart shown in figure 1). The clinical characteristics of the patients are presented in table 1. The prevalence of IBP was generally higher in patients with axSpA (with no difference between AS and nr-axSpA) as compared with patients without axSpA (table 2).

The formal agreement on the global judgement on the presence of IBP between R-blind and R-care was moderate (kappa 0.45; 95% CI 0.36 to 0.54) with percentage agreement 74.9%. Similar rates of agreement between R-blind and R-care were obtained for the various defined IBP criteria: kappa 0.43 (95% CI 0.32 to 0.53; percentage agreement 80.2%) for the Calin criteria; kappa 0.52 (95% CI 0.43 to 0.61; percentage agreement 80.0%) for Berlin criteria and kappa 0.46

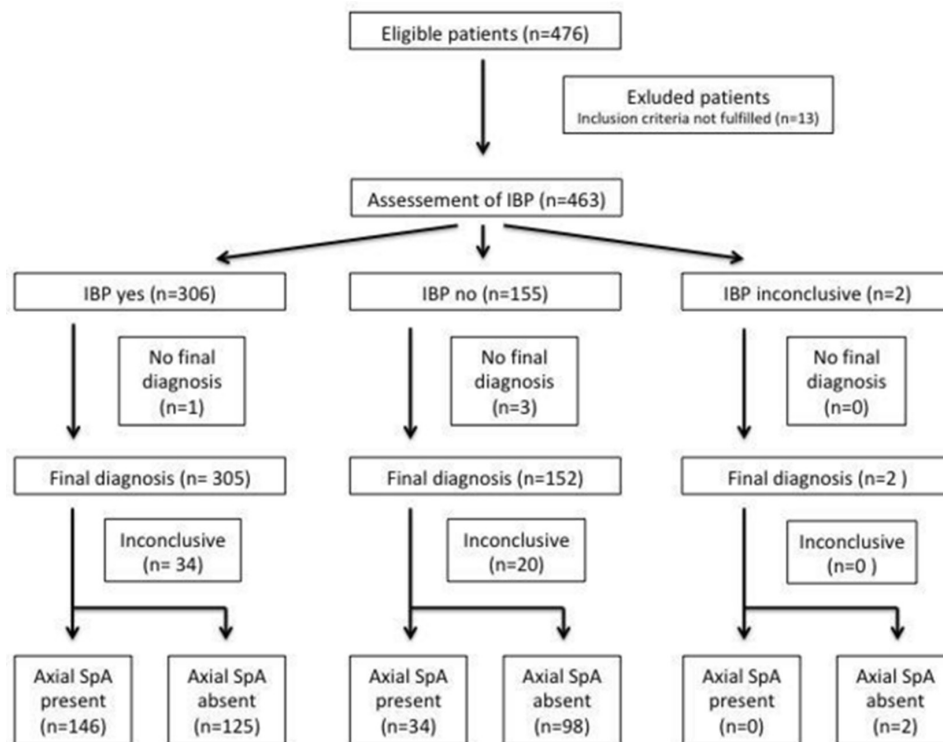


Figure 1 Flow chart of patients with chronic back pain included in diagnostic accuracy of inflammatory back pain study. IBP, inflammatory back pain; SpA, spondyloarthritis.

Table 1 Clinical, laboratory and imaging characteristics of patients with chronic back pain who had judgement on IBP and who received a final diagnosis

Parameter	AxSpA			All (n=403)
	AS (n=88)	Nr-axSpA (n=92)	No-axSpA (n=223)	
Age, years (mean±SD)	36.1±10.2	32.8±9.3	42.7±11.3	39.0±11.4
Male sex, n (%)	60 (68.2)	54 (58.7)	79 (35.4)	193 (47.9)
HLA-B27 positive, n (%)	79 (89.8)	70 (76.1)	83 (37.2)	232 (57.6)
Back pain duration, years (mean±SD)	9.0±7.8	5.3±6.9	12.5±10.9	10.1±9.9
Peripheral oligoarthritis, n (%)	12 (13.6)	20 (21.7)	32 (14.3)	64 (15.9)
Enthesitis, n (%)	21 (23.9)	16 (17.4)	33 (14.8)	71 (17.6)
Dactylitis, n (%)	2 (2.3)	5 (5.4)	10 (4.5)	17 (4.2)
Uveitis, n (%)	13 (14.8)	18 (19.6)	11 (4.9)	42 (10.4)
History of IBD, n (%)	2 (2.3)	4 (4.4)	0 (0)	6 (1.5)
Psoriasis, n (%)	2 (2.3)	13 (14.1)	16 (7.2)	31 (7.7)
SpA family history, n (%)	17 (19.3)	24 (26.1)	45 (20.1)	86 (21.3)
CRP, mg/L (mean±SD)	11.0±16.4	7.1±9.2	2.9±4.3	5.7±9.9
ESR, mm/hour (mean±SD)	22.7±20.8	15.2±13.3	12.0±10.3	15.2±14.6
Radiographic sacroiliitis according to the mNY criteria, n/N (%)	78/85 (91.8)	0/85	0/208	85/378 (22.5)
Active inflammatory changes in the SIJ on MRI, n/N (%)	37/48 (77.1)	51/70 (72.9)	2/77 (2.6)	90/195 (46.2)
Chronic inflammatory changes in the SIJ on MRI, n/N (%)	32/48 (66.7)	38/70 (54.3)	0/77 (0)	70/195 (35.9)
Active inflammatory changes in the spine on MRI, n/N (%)	15/36 (41.7)	13/35 (37.1)	5/84 (6.0)	33/155 (21.3)
Chronic inflammatory changes in the spine on MRI, n/N (%)	11/36 (30.6)	3/35 (8.6)	3/84 (3.6)	17/155 (11.0)

AS, ankylosing spondylitis; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; HLA, human leucocyte antigen; IBD, inflammatory bowel disease; MRI, MRI resonance imaging; SIJ, sacroiliac joint; SpA, spondyloarthritis; mNY criteria, modified New York criteria; nr-axSpA, non-radiographic axial spondyloarthritis.

Table 2 The prevalence of IBP (%) according to different criteria in patients referred because of chronic back pain

Assessor	IBP according to...	AxSpA			All (n=403)
		AS (n=88)	nr-axSpA (n=92)	No-axSpA (n=223)	
Blinded rheumatologist	Global evaluation	80.7	81.5	56.1	67.3
	Calin's criteria	79.6	79.4	74.9	76.9
	Berlin Criteria	81.8	80.4	67.7	73.7
	ASAS criteria	75.0	73.9	60.5	66.8
Diagnosing rheumatologist	Global evaluation	92.1	88.0	41.3	63.0
	Calin's criteria	87.5	82.6	73.1	78.4
	Berlin criteria	85.2	82.6	55.2	68.0
	ASAS criteria	89.8	78.3	64.6	73.2
Patient	Calin's criteria	73.6	76.7	79.5	77.5
	Berlin criteria	87.4	85.6	81.7	83.8
	ASAS criteria	79.3	68.9	69.0	71.2

AS, ankylosing spondylitis; ASAS, Assessment of SpondyloArthritis international Society; IBP, inflammatory back pain; SpA, spondyloarthritis; nr-axSpA, non-radiographic axial spondyloarthritis.

(95% CI 0.37, 0.56; percentage agreement 77.3%) for the ASAS criteria.

Diagnostic accuracy of IBP

For diagnostic accuracy analyses of IBP, data from AS and nr-axSpA were pooled (table 3). IBP by R-blind had a sensitivity of 74%–81% and a specificity of 25%–44% for the diagnosis of axSpA, depending on the IBP definition applied. Interestingly, global judgement on

IBP by R-blind numerically exceeded the three sets of defined IBP criteria in terms of sensitivity and specificity; yet, the resulting positive LR_s overall were low, ranging from 1.1 (Calin criteria) to 1.4 (global judgement of IBP) (table 3). The results were similar when we stratified for single study sites or for type of centre: for example, the LR₊ for IBP according to ASAS criteria by R-blind was 1.3 (hospital based) and 1.1 (private

Table 3 Sensitivity, specificity, PPV and NPV of IBP for the diagnosis of axSpA

Assessor	IBP according to...	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV (95% CI)	NPV (95% CI)	LR+ (95% CI)	LR- (95% CI)
Blinded rheumatologist	Global evaluation	81.1 (75.4 to 86.8)	43.9 (37.3 to 50.7)	53.9 (48.0 to 59.8)	74.2 (66.8 to 81.7)	1.4 (1.3 to 1.7)	0.4 (0.3 to 0.6)
	Calin's criteria	79.4 (73.5 to 85.4)	25.1 (19.6 to 31.3)	46.1 (40.6 to 51.9)	60.2 (50.0 to 70.2)	1.1 (1.0 to 1.2)	0.8 (0.6 to 1.2)
	Berlin criteria	81.1 (75.4 to 86.8)	32.3 (26.2 to 38.6)	49.2 (43.3 to 55.0)	67.9 (58.7 to 76.7)	1.2 (1.1 to 1.3)	0.6 (0.4 to 0.8)
	ASAS criteria	74.4 (68.1 to 80.8)	39.5 (33.0 to 46.1)	49.8 (43.8 to 55.8)	65.7 (57.3 to 73.4)	1.2 (1.1 to 1.4)	0.6 (0.5 to 0.9)
Diagnosing rheumatologist	Global evaluation	90.0 (85.6 to 94.4)	58.7 (52.1 to 65.2)	63.8 (57.8 to 69.7)	87.9 (81.9 to 92.7)	2.2 (1.9 to 2.6)	0.2 (0.1 to 0.3)
	Calin's criteria	85.0 (79.8 to 90.2)	26.9 (21.4 to 33.0)	48.4 (42.8 to 54.1)	69.0 (58.4 to 78.2)	1.2 (1.1 to 1.3)	0.6 (0.4 to 0.8)
	Berlin criteria	83.9 (78.5 to 89.3)	44.8 (38.2 to 51.6)	55.1 (49.0 to 61.1)	77.5 (69.7 to 84.4)	1.5 (1.3 to 1.7)	0.4 (0.2 to 0.5)
	ASAS criteria	83.9 (78.5 to 89.3)	35.4 (29.2 to 41.8)	51.2 (45.3 to 57.0)	73.2 (64.2 to 80.8)	1.3 (1.2 to 1.5)	0.5 (0.3 to 0.7)
Patient	Calin's criteria	75.1 (68.8 to 81.5)	20.6 (15.6 to 26.3)	43.3 (37.7 to 49.0)	50.6 (39.8 to 61.0)	0.9 (0.8 to 1.1)	1.2 (0.8 to 1.7)
	Berlin criteria	86.4 (81.4 to 91.5)	18.3 (13.5 to 23.7)	46.1 (40.6 to 51.6)	62.5 (50.0 to 74.0)	1.1 (1.0 to 1.2)	0.7 (0.5 to 1.2)
	ASAS criteria	74.0 (67.6 to 80.5)	31.1 (25.0 to 37.6)	46.5 (40.5 to 52.4)	59.7 (50.5 to 68.7)	1.1 (0.9 to 1.2)	0.8 (0.6 to 1.1)

ASAS, Assessment of SpondyloArthritis international Society; IBP, inflammatory back pain; LR, likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; SpA, spondyloarthritis; axSpA, axial spondyloarthritis.

practices), respectively. Overall, the LR+ as a measure of diagnostic utility of the symptom of IBP for the diagnosis of axSpA was small and substantially smaller than expected from published data (LR+ 3–4) due to low specificity, independently of the definition of IBP being applied.

The ASAS defined IBP criteria had the lowest sensitivity (74.4%), while the Berlin criteria and global judgement of IBP had the highest sensitivity (both 81.1%). The specificity of IBP according to R-blind varied from 24.9% (Calin criteria) to 43.9% (global judgement) (table 3). There were no striking differences between the three sets of IBP criteria with regard to PPV and NPV: the Calin criteria performed slightly less well than the Berlin and ASAS criteria.

In the original publication of the Berlin criteria, it was speculated that the presence of ≥ 3 out of 4 items (instead of ≥ 2 out of 4) would yield a high diagnostic gain (estimated LR+ 12.4; specificity 97.3%, sensitivity 33.6%).² In the settings of DIVERS, however, the sensitivity and specificity of ≥ 3 out of 4 items of the Berlin criteria were 59.8% and 59.9%, respectively, and resulted in a small increase of the LR+ from 1.2 to 1.5 only.

Single IBP features

Analysing single IBP parameters, the highest sensitivity for the diagnosis of axSpA was observed for ‘improvement of back pain with exercise’ (88.9%), which, however, had a low specificity (22.7%) (table 4). ‘Alternating buttock pain’ had the lowest sensitivity (60.0%) but the highest specificity (58.2%), and the highest LR+ of 1.4 for a single IBP feature according to R-blind.

Blinded as compared to unblinded assessment of IBP

It can be assumed that the judgement on a diagnostic test like IBP, which is subject to interpretation, will be influenced by knowledge of other diagnostic test results

(‘diagnostic bias’). Indeed, both sensitivity 90.0% versus 81.1% and specificity 58.2% versus 44.0% of IBP according to global judgement were higher for R-care (unblinded) than for R-blind (table 3). This indeed demonstrates a moderate diagnostic bias of R-care in the assessment of IBP by knowledge of other SpA features. In contrast, the comparison between R-blind and R-care according to formal sets of IBP criteria (Calin, Berlin, ASAS) revealed no consistent differences in specificity but lower sensitivities for all three sets of criteria when assessed by R-blind (table 3). This suggests that ‘global judgement’ on IBP is more susceptible to diagnostic bias than defined sets of IBP criteria.

IBP self-assessment by the patient

With regard to fulfilment of defined sets of IBP criteria, little differences were found in the self-assessed prevalence of IBP between patients with axSpA and patients without axSpA (table 2): the specificities of defined IBP criteria, and of single IBP features, were even lower if self-assessed by the patient, while the sensitivities were comparable to those as assessed by R-blind (tables 3 and 4), resulting in even slightly lower LR+ of between 0.9 and 1.1 overall.

Diagnostic gain of IBP for the diagnosis of axSpA

In DIVERS, axSpA was diagnosed in 44.6% (pretest probability of axSpA). According to global judgement on IBP by R-blind, the presence of IBP (LR+ 1.4) resulted in a post-test probability of axSpA of 53%. Thus, a moderate diagnostic gain of IBP of 8.4% remained despite the low LR+. For comparison, the presence of IBP according to R-care—unblinded to other patient findings including human leucocyte antigen (HLA)-B27 and imaging, and therefore somewhat biased—had resulted in an increase of axSpA by as much as 19.5% (from 44.6% to 63.9%).

Table 4 Sensitivity and specificity of single IBP parameters as assessed by the diagnosing rheumatologist, the blinded rheumatologist and patient for the diagnosis of axSpA

IBP parameter	Blinded rheumatologist			Diagnosing rheumatologist			Patient		
	Sensitivity, % (95% CI)	Specificity, % (95% CI)	LR+	Sensitivity, % (95% CI)	Specificity, % (95% CI)	LR+	Sensitivity, % (95% CI)	Specificity, % (95% CI)	LR+
Insidious onset	75.6 (69.3 to 81.8)	24.4 (18.8 to 30.1)	1.0	83.9 (78.5 to 89.3)	19.6 (14.4 to 24.7)	1.0	76.2 (69.8 to 82.5)	15.3 (10.5 to 20.1)	0.9
Morning stiffness ≥ 30 min	71.1 (63.8 to 78.3)	33.9 (27.0 to 40.8)	1.1	67.4 (59.8 to 74.9)	37.7 (30.4 to 45.1)	1.1	69.4 (62.5 to 76.3)	30.1 (24.0 to 36.2)	1.0
Improvement with exercise	88.9 (84.3 to 93.5)	22.7 (17.2 to 28.1)	1.2	85.6 (80.4 to 90.7)	30.7 (24.6 to 36.7)	1.2	83.0 (77.4 to 88.7)	18.6 (13.4 to 23.8)	1.0
No improvement with rest	73.9 (67.5 to 80.3)	44.9 (38.4 to 51.4)	1.3	81.6 (75.9 to 87.2)	41.8 (35.3 to 48.2)	1.4	81.1 (75.1 to 87.1)	30.8 (24.5 to 37.0)	1.2
Pain at night	74.4 (68.1 to 80.8)	33.8 (27.6 to 40.0)	1.1	82.8 (77.3 to 88.3)	30.1 (24.8 to 36.9)	1.2	77.6 (71.4 to 83.8)	26.9 (21.1 to 32.8)	1.1
Alternating buttock pain	60.0 (52.8 to 67.2)	58.2 (51.8 to 64.7)	1.4	57.8 (50.6 to 65.0)	63.4 (57.0 to 69.7)	1.6	68.2 (61.3 to 75.1)	43.4 (36.8 to 50.0)	1.2

IBP, inflammatory back pain; LR, likelihood ratio; SpA, spondyloarthritis; axSpA, axial spondyloarthritis.

DISCUSSION

DIVERS is the first real-life diagnostic accuracy study on IBP as a diagnostic test for axSpA in the rheumatology setting, that is, in secondary and tertiary care. We found a net diagnostic gain of only 2.5%–8.4%, if IBP is present, for the likelihood of a diagnosis of axSpA. Thus, one important finding at first glance is that IBP in the rheumatology setting contributes little to establishing the diagnosis of axSpA. On the other hand, the majority of referred patients had IBP, suggesting an effective selection in primary care of patients presenting with chronic back pain for referral to rheumatology. Moreover, our study shows that in the rheumatology setting, none of the defined sets of IBP criteria (Calin, Berlin, ASAS) clearly outperformed another one; yet, a tendency for the Calin criteria being the least specific set was found.

The sensitivity of IBP according to various defined IBP criteria (74.4%–81.1%) or to global judgement on IBP (81.1%) among patients with axSpA in DIVERS was very similar to published figures of 70.1%–95% in patients with AS/axSpA.^{1–3} However, we found substantially lower specificities for all defined sets of IBP criteria, ranging from 25.2% to 39.5%, as compared with the original publications on IBP (72.5%–81.3%).^{1–3} As a result, the calculated LR+ were low (1.1–1.2) as compared with published LR+ for these IBP definitions of 2.9–3.9.^{1–3} A plausible explanation for the differences in the specificities of IBP criteria could be the fact that in two of the three earlier studies,^{1,2} well-selected patients with either a clear diagnosis of axSpA or of mechanical back pain (no-axSpA) were included (convenience sample), whereas in our study undiagnosed and newly referred patients were included, thereby reflecting better daily rheumatology practice. Since the prevalence of IBP was high among referred patients, one must assume that IBP has operated as a selection parameter in primary care that had triggered referral to the rheumatologist. This unmeasured channelling process led to a population of referred patients with chronic back pain who were enriched for the presence of IBP. In fact, it has been proposed in 2005, and has subsequently proven effective, to select in primary care for referral to rheumatology chronic back patients with age at onset ≤ 45 years and at least one additional SpA feature such as IBP or a positive HLA-B27 test, both of which increase the likelihood of having axSpA.¹³ In epidemiological studies on unselected back pain, prevalence figures for IBP of 5%–15% for acute and of 28%–35% for chronic back pain have been reported.^{17,18} The high prevalence of IBP of 67.3% in our study indeed supports the notion that IBP had been used by primary care physicians to select patients for referral, although this was not specifically intended.

The selection of patients with certain SpA features for referral is also illustrated in our study by the high prevalence of other SpA features including HLA-B27 (57.6%), a positive family history for SpA (21.3%) or

uveitis (10.4%); all of them leading to enrichment of patients with a higher likelihood of having axSpA. In fact, the selected referral of patients with back pain more likely to have axSpA is eventually reflected by the high rate of a final diagnosis of axSpA (44.6%) in DIVERS which is much higher than reported prevalences of axSpA of around 5%–12% in unselected patients with chronic back pain.^{19,20} Interestingly, studies with a study design similar to DIVERS also revealed high prevalence rates of axSpA among referred patients: in the international ASAS classification criteria for axSpA study, the prevalence of axSpA was 66%,⁵ and in the SpondyloArthritis Caught Early cohort, axSpA was diagnosed in 41%.²¹ Although no structured referral protocol was recommended in either of these studies, an unmeasured preselection process had undoubtedly taken place in both,²² suggesting ‘unmeasured’ selection of patients for referral likely occurring in other countries and settings as well.

The low LR+ of 1.1–1.2 for defined sets of IBP criteria suggests at first glance a minor, if any, diagnostic utility of IBP in the rheumatology setting. It seems that the diagnostic utility of IBP has been already ‘used up’ at the time when the patient is referred to the rheumatologist. Yet, a small diagnostic gain is indeed retained, partially because the pretest probability (prevalence of axSpA) was substantially higher in the rheumatology setting than in primary care: the presence of IBP according to global judgement by the blinded assessor (LR+ 1.4) resulted in an increase of the probability of axSpA from 44.6% to 53%, implicating a net diagnostic gain of 8.4%, whereas the increase with fulfilment of Calin criteria (2.4%), Berlin (4.5%) and ASAS criteria (4.5%) was lower than the estimated net diagnostic gain of 9%–11% for IBP among unselected patients with chronic back pain in primary care (probability of axSpA increases in primary care from 5% to 14%–16%).^{6,19,23}

In DIVERS, we also analysed single IBP features, among which ‘alternating buttock pain’ had the highest specificity and the highest LR+ of 1.4, followed by ‘no improvement with rest’ (LR+ 1.3), suggesting that these items may provide some diagnostic information in the rheumatology setting. We also addressed the self-assessment of IBP by the patient. The resulting sensitivities were comparable to those by the blinded rheumatologist. However, the specificities were similar or even lower. The specificity of self-assessment of IBP symptoms in unselected patients in primary care is expected to be higher but cannot be properly addressed in our study.^{7,13,24}

The knowledge of SpA features other than IBP that drive towards or away from a diagnosis of axSpA is likely to influence the global judgement on the presence or absence of IBP. Indeed, this potential bias is illustrated by both a higher sensitivity and higher specificity (with a resulting higher LR+ of 2.2 vs 1.4) for the diagnosing (unblinded) as compared with the blinded

rheumatologist in DIVERS and underscores the necessity in diagnostic accuracy studies for a diagnostic test like IBP, which is subject to interpretation, to be assessed in a blinded fashion.^{14 15}

The strengths of our study are the high standards for the conduct and reporting of diagnostic studies including the prospective study design, the independent assessment of the diagnostic test of interest (IBP) without knowledge of other test results,^{14 15} the multicentre study design with secondary care (four private practices) as well as tertiary care centres (one university hospital and one large community hospital specialised in rheumatology) and the large number of consecutive patients. A potential weakness of our study conducted in routine rheumatology care is the fact that not all patients diagnosed as no-axSpA underwent MRI. However, only 22 of patients without MRI (9.8% of all patients without SpA patients) had a clinical context (IBP plus positive HLA-B27) strongly suggesting axSpA: if all of these patients had shown sacroiliitis on MRI and had been diagnosed axSpA, a scenario that is unlikely, the specificity of IBP (global judgement) had increased from 43.9% to 48.8% and the LR+ from 1.4 to 1.56.

Of interest, our study strikingly shows how test characteristics and the resulting LR+ vary, depending on the setting (primary vs secondary/tertiary care) where these tests are applied and depending on whether parameters have operated already as selection parameters for referral. The understanding of these potential shifts in specificity and LR+ of diagnostic tests is of general importance when interpreting data from any study on diagnostic test characteristics in medicine. The results of this study also confirm that a diagnosis of axSpA cannot be made by the presence or absence of single parameters (in this case of IBP) but only by assessment of all available clinical, laboratory and imaging parameters, interpreted by an experienced physician and after careful exclusion of other diagnoses.²³

In summary, rheumatologists must be aware that their global judgement on IBP might be influenced by knowledge of other SpA parameters. Rheumatologists must also be aware that many patients referred to them for a diagnostic workup of axSpA will have IBP because IBP effectively operates in primary care as a selection criterion for referral. Although the specificity of IBP (and the resulting LR+) is low in the rheumatology setting, a small diagnostic gain remains.

Author affiliations

¹Department of Gastroenterology, Infectious Diseases and Rheumatology, Campus Benjamin Franklin, Charité Universitätsmedizin Berlin, Berlin, Germany

²Epidemiology Unit, German Rheumatism Research Centre, Berlin, Germany

³Department of Rheumatology, Rheumazentrum Ruhrgebiet, Herne, Germany

⁴Department of Medicine and Rheumatology, Klinikum Bielefeld Rosenhöhe, Bielefeld, Germany, Charité Universitätsmedizin Berlin, Berlin, Germany, and Gent University, Gent, Belgium

Acknowledgements We are grateful to Beate Buß for monitoring the data and Janis Vahldiek for data management support. We like to thank the following rheumatologists for participation in this project: Jan Brandt, Kirsten Karberg, Klaus

Krüger, Dorothea Pick, Florian Schuch and Jörg Wendler. We further like to thank all colleagues who referred their patients with chronic back pain within this study.

Collaborators Jan Brandt, Klaus Krüger, Kirsten Karberg, Dorothea Pick, Florian Schuch, Jörg Wendler.

Contributors All authors contributed to acquisition, analysis and interpretation of the data and drafting the manuscript.

Funding This work was supported by a research grant from the German Research Foundation (Deutsche Forschungsgemeinschaft (DFG)), GZ RU 681/5-1.

Competing interests None declared.

Patient consent Obtained.

Ethical approval The study was approved by the ethics committee of the Charité Universitätsmedizin Berlin, and thereafter in the respective institutions of all participating centres.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0>

REFERENCES

- Calin A, Porta J, Fries JF, *et al.* Clinical history as a screening test for ankylosing spondylitis. *JAMA* 1977;237:2613–4.
- Rudwaleit M, Metter A, Listing J, *et al.* Inflammatory back pain in ankylosing spondylitis: a reassessment of the clinical history for application as classification and diagnostic criteria. *Arthritis Rheum* 2006;54:569–78.
- Sieper J, van der Heijde D, Landewé R, *et al.* New criteria for inflammatory back pain in patients with chronic back pain: a real patient exercise by experts from the Assessment of SpondyloArthritis international Society (ASAS). *Ann Rheum Dis* 2009;68:784–8.
- Rudwaleit M, Landewé R, van der Heijde D, *et al.* The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. *Ann Rheum Dis* 2009;68:770–6.
- Rudwaleit M, van der Heijde D, Landewé R, *et al.* The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis* 2009;68:777–83.
- Rudwaleit M, van der Heijde D, Khan MA, *et al.* How to diagnose axial spondyloarthritis early. *Ann Rheum Dis* 2004;63:535–43.
- Sieper J, Rudwaleit M. Early referral recommendations for ankylosing spondylitis (including pre-radiographic and radiographic forms) in primary care. *Ann Rheum Dis* 2005;64:659–63.
- Brandt HC, Spiller I, Song IH, *et al.* Performance of referral recommendations in patients with chronic back pain and suspected axial spondyloarthritis. *Ann Rheum Dis* 2007;66:1479–84.
- Hermann J, Giessauf H, Schaffler G, *et al.* Early spondyloarthritis: usefulness of clinical screening. *Rheumatology* 2009;48:812–6.
- Poddubny D, Vahldiek J, Spiller I, *et al.* Evaluation of 2 screening strategies for early identification of patients with axial spondyloarthritis in primary care. *J Rheumatol* 2011;38:2452–60.
- Sieper J, Srinivasan S, Zamani O, *et al.* Comparison of two referral strategies for diagnosis of axial spondyloarthritis: the Recognising and Diagnosing Ankylosing Spondylitis Reliably (RADAR) study. *Ann Rheum Dis* 2013;72:1621–7.
- van Hooft L, Vergouwe Y, de Buck PD, *et al.* External validation of a referral rule for axial spondyloarthritis in primary care patients with chronic low back pain. *PLoS One* 2015;10:e0131963.
- Rudwaleit M, Sieper J. Referral strategies for early diagnosis of axial spondyloarthritis. *Nat Rev Rheumatol* 2012;8:262–8.
- Bossuyt PM, Reitsma JB, Bruns DE, *et al.* Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;326:41–4.
- Sackett DL, Haynes RB. The architecture of diagnostic research. *BMJ* 2002;324:539–41.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- Weisman MH, Witter JP, Reveille JD. The prevalence of inflammatory back pain: population-based estimates from the US National

- Health and Nutrition Examination Survey, 2009-10. *Ann Rheum Dis* 2013;72:369-73.
18. Hamilton L, Macgregor A, Warmington V, *et al.* The prevalence of inflammatory back pain in a UK primary care population. *Rheumatology* 2014;53:161-4.
 19. Underwood MR, Dawes P. Inflammatory back pain in primary care. *Br J Rheumatol* 1995;34:1074-7.
 20. Liao ZT, Pan YF, Huang JL, *et al.* An epidemiological survey of low back pain and axial spondyloarthritis in a Chinese Han population. *Scand J Rheumatol* 2009;38:455-9.
 21. van den Berg R, de Hooze M, van Gaalen F, *et al.* Percentage of patients with spondyloarthritis in patients referred because of chronic back pain and performance of classification criteria: experience from the Spondyloarthritis Caught Early (SPACE) cohort. *Rheumatology* 2013;52:1492-9.
 22. Rudwaleit M. Spondyloarthropathies: identifying axial SpA in young adults with chronic back pain. *Nat Rev Rheumatol* 2016;12:378-80.
 23. Rudwaleit M, Khan MA, Sieper J. The challenge of diagnosis and classification in early ankylosing spondylitis: do we need new criteria? *Arthritis Rheum* 2005;52:1000-8.
 24. Poddubnyy D, van Tubergen A, Landewé R, *et al.* Development of an ASAS-endorsed recommendation for the early referral of patients with a suspicion of axial spondyloarthritis. *Ann Rheum Dis* 2015;74:1483-7.