

# Preliminary Validation of a Self-reported Screening Questionnaire for Inflammatory Back Pain

STEPHANIE O. KEELING, SUMIT R. MAJUMDAR, BARBARA CONNER-SPADY, MICHELE C. BATTIÉ, LINDA J. CARROLL, and WALTER P. MAKSYMOWYCH

**ABSTRACT.** *Objective.* Inflammatory back pain (IBP) is an important feature of axial spondyloarthritis (SpA) that is poorly recognized in primary care, perhaps delaying diagnosis of SpA. We aimed to develop and validate a self-report questionnaire using important domains reported by patients with IBP.

*Methods.* We developed a 6-item questionnaire assessing spinal/hip stiffness, nocturnal pain, diurnal variation, effects of exercise/rest, and peripheral joint pain/swelling. This was compared with the Calin questionnaire and the domains comprising the Assessment of Spondyloarthritis International Society (ASAS) criteria for IBP in 220 patients with established axial SpA and 66 patients with mechanical back pain followed in tertiary care rheumatology clinics. The classification utility of each item was evaluated using sensitivity, specificity, and likelihood ratio (LR). Multivariable logistic regression was used to analyze different combinations of items to develop candidate scoring systems.

*Results.* The single item “diurnal variation” had the highest combination of sensitivity (49%) and specificity (92%) for IBP (positive LR 5.95, 95% CI 2.54–13.94), outperforming the Calin and ASAS IBP criteria, which had sensitivities of 83% and 59%, specificities 42% and 66%, positive LR 1.42 and 1.72, negative LR 0.41 and 0.62, respectively. Classification utility of this item was even higher in SpA patients with disease duration < 6 years (sensitivity 48%, specificity 96%, positive LR 12, negative LR 0.54). The other 5 items did not improve classification utility in any combination.

*Conclusion.* Assessment of a single self-reported item, “diurnal variation,” had substantial classification utility for IBP. This domain is not addressed in existing criteria for IBP, indicating a potentially important omission. (J Rheumatol First Release Feb 15 2012; doi:10.3899/jrheum.110537)

## Key Indexing Terms:

INFLAMMATORY BACK PAIN  
DIURNAL VARIATION

SPONDYLOARTHRITIS  
ASAS INFLAMMATORY BACK PAIN CRITERIA  
CALIN CRITERIA  
CLASSIFICATION UTILITY

Inflammatory back pain (IBP) is one feature of ankylosing spondylitis (AS) that is incorporated into the modified New York classification criteria<sup>1</sup> and more recently the Assessment of Spondyloarthritis International Society (ASAS) classification criteria for axial SpA<sup>2</sup>. There are no widely accepted diagnostic criteria for AS and so the classification criteria are often used for the purpose of diagnosis. However, diagnostic

delay in AS is problematic because symptoms often predate diagnosis by 8 to 11 years<sup>3,4,5</sup>. Given that the symptom severity in these patients is comparable to those with established disease and that the benefit of tumor necrosis factor inhibition in these patients is well recognized<sup>6</sup>, this delay is no longer acceptable<sup>7,8,9,10</sup>.

A lack of recognition of IBP among primary care physicians likely contributes to this delay in IBP diagnosis and subsequent referral to rheumatology<sup>11</sup>. A recent postal survey of general practitioners suggested great variability in understanding the features of IBP and indicated a need for continuing education<sup>12</sup>. However, SpA is found in only 3%–5% of patients presenting with lower back pain in primary care, so few of these physicians will attain sufficient experience to diagnose IBP with confidence<sup>11</sup>. Recognition of IBP may be facilitated by the development of screening strategies that include self-report questionnaires<sup>12,13</sup>. Improved IBP screening would help both primary care physicians and rheumatologists identify patients who may have SpA and expedite their formal assessment.

Until 2009, only 2 screening criteria for IBP existed, the Calin criteria<sup>14</sup> and the more recent Berlin criteria<sup>15</sup>. The Calin criteria were developed in 1977 and have been incorpo-

---

*From the Department of Medicine, School of Public Health, and Department of Physical Therapy, University of Alberta, Edmonton; and Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada.*

*Prof. Majumdar is a Health Scholar supported by Alberta Innovates – Health Solutions.*

*S.O. Keeling, MSc, MD, FRCPC, Assistant Professor, Department of Medicine; S.R. Majumdar, MD, MPH, Professor, Department of Medicine and School of Public Health, University of Alberta; B. Conner-Spady, PhD, Department of Community Health Sciences, University of Calgary; M.C. Battié, PhD, Professor, Department of Physical Therapy; L.J. Carroll, PhD, Professor, School of Public Health; W.P. Maksymowych, MB ChB, FRCPC, Professor, Department of Medicine, University of Alberta.*

*Address correspondence to Prof. S. Keeling, Department of Medicine, University of Alberta, 562 Heritage Medical Research Center, Edmonton, Alberta T6G 2S2, Canada. E-mail: stephanie.keeling@ualberta.ca*  
*Accepted for publication December 7, 2011.*

---

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2012. All rights reserved.

rated to some extent into the European Spondylarthropathy Study Group (ESSG)<sup>16</sup>, the modified New York criteria for AS<sup>1</sup>, and the Amor criteria for SpA<sup>17</sup>. The Calin criteria, based on nonstandardized questions administered by the interviewing clinician that were felt to be important in the clinical history for IBP, have been studied in various populations, but the same high sensitivity (95%) and specificity (76%) of the original study has not been reproduced<sup>1,18,19</sup>. The Berlin criteria were derived from an exploratory study that assessed 101 patients with AS and 112 control patients with mechanical back pain through face-to-face interviews by the same examiner<sup>15</sup>. When 3 out of 4 IBP measures were positive (“morning stiffness,” “improvement of back pain with exercise and not rest,” “nocturnal pain,” and “alternating buttock pain”), sensitivity and specificity were 33.6% and 97.3%, respectively, with a positive likelihood ratio (LR) of 12.4<sup>15</sup>. Limitations of the Berlin criteria include difficulties and variation in interpretation of the wording in specific domains, such as “alternating buttock pain,” lack of standardization of the wording to the questions, and the use of an established cohort of patients with AS and mechanical back pain controls<sup>15</sup>.

In 2009, the ASAS IBP criteria were reported and these were based on the expert judgment of the rheumatologist as the “gold standard” for diagnosing IBP in 20 patients with chronic back pain of unknown origin<sup>20</sup>. These new candidate IBP criteria administered by the interviewing clinician included the domains “improvement with exercise,” “nocturnal pain,” “age at onset < 40 years,” and “no improvement with rest.” They were then validated in a distinct cohort of 648 patients presenting to the rheumatologist with new-onset back pain, and were shown to have a sensitivity of 79.6% and specificity of 72.4%<sup>20</sup>. Limitations of the ASAS IBP criteria include the lack of standardization of the questions, so that elicitation and affirmation of IBP items is dependent on the expertise of the interviewer. As stated, this expertise is typically lacking among primary care physicians. Concerns have also been raised regarding the omission of morning stiffness due to a lack of statistical significance, even though this represents a concept considered fundamental to the recognition of IBP<sup>21</sup>. The aim of our study was to use expert consensus to develop and then pilot a standardized set of patient self-reported questions that address domains considered relevant to a diagnosis of IBP and which could be employed in primary care to identify patients who require further assessment for SpA.

## MATERIALS AND METHODS

*Development of survey questions.* Three rheumatologists at the University of Alberta (A.S. Russell, W.P. Maksymowych, S.O. Keeling) reviewed the literature and the existing criteria for IBP in general to identify content domains considered relevant to the diagnosis of IBP. At the time of survey development and data collection, the Berlin and ASAS IBP criteria were unpublished. The 3 rheumatologists reviewed the Calin criteria independently and then as a group. At both stages, each individual Calin question item was scrutinized (Figure 1) to determine to what degree the question item addressed the domains of interest, which domains were not addressed, and the optimal for-

mulation of the wording of questions addressing each domain in the context of the extensive clinical background of the investigators in managing patients with SpA. This resulted in a 6-item questionnaire.

To confirm that the 6-item questionnaire did not reflect a particular site bias for the assessment of IBP, the questionnaire was sent to 11 members of the Spondyloarthritis Research Consortium of Canada (SPARCC), and this panel of national experts reviewed each item on a 10-point scale and ranked its importance in the diagnosis of IBP. The questionnaire was further scrutinized for feasibility and comprehension by patient members of the Canadian Spondylitis Association. All the questions/domains were considered appropriate.

*Candidate items.* The important domains for IBP identified by expert consensus included (1) morning stiffness, (2) age at onset of back and/or hip pain, (3) nocturnal pain, (4) diurnal variation in symptoms, (5) peripheral joint pain and swelling, (6) response to exercise, (7) response to rest, and (8) response to exercise and rest. The Calin and 6-item self-report questionnaires did not include exactly the same domains (Figure 1). Additional domains included in the 6-item questionnaire but not the Calin criteria included (1) nocturnal pain, (2) diurnal variation, (3) peripheral joint pain and swelling, and (4) response to rest.

*Face validity and pilot testing in an established back pain cohort.* Thirteen rheumatologists (7 at the University of Alberta, including 3 who derived the 6-item questionnaire, 6 community rheumatologists in the city of Edmonton) administered the Calin and 6-item questionnaire to consecutive patients with an established cause of back pain, using the rheumatologist’s assessment and diagnosis as the gold standard. While the 3 rheumatologists who derived the 6-item questionnaire included their own patients in the study, 48% (n = 136) of patients were contributed by other rheumatologists. The SpA diagnoses for the patients with IBP by the rheumatologists were not based formally on either the ASAS axial SpA criteria<sup>2</sup> or ESSG criteria<sup>16</sup> but constituted their expert opinion. The limitations of the ESSG criteria have been cited<sup>2</sup>, while the ASAS criteria were not published at the time the study was designed. Any patient with back pain of known cause was included if they were 18 years of age or older, spoke English, and agreed to participate in the study. Rheumatologists categorized the patient as “mechanical back pain (MBP)” or “inflammatory back pain (IBP)” and included the specific diagnosis (when available) and duration of back pain symptoms. The questionnaires were administered in random order to address possible order effects. All patients provided written informed consent and the study was approved by the Health Ethics Research Board of the University of Alberta (Edmonton, Alberta, Canada).

*Data analysis.* Baseline characteristics of the mechanical and inflammatory back pain cohort were compared with chi-square, Mann Whitney U, or Student t tests as appropriate. The responses to the Calin items are dichotomous (yes/no) whereas our 6-item questionnaire included a combination of dichotomous responses (yes or no) and multiple choice answers. We pre-specified which responses for each individual item would be considered indicative of IBP in the 6-item questionnaire prior to the data analysis, due to the potential combinations of dichotomous and multiple choice answers (Figure 1). The affirmative response reflected the consensus of the expert panel of what represented IBP versus MBP for each individual item. Positive and negative LR, sensitivities, and specificities were calculated for all items, and analyses were repeated stratifying patients by age, sex, and symptom duration. For the latter, symptom duration was divided into tertiles of the cumulative distribution (< 6 years, 6–18 years, > 18 years).

Multivariate logistic regression analysis of all statistically significant items ( $p < 0.05$ ) from univariate analyses of the Calin and 6-item questionnaire, respectively, was performed with the diagnosis of IBP as the dependent variable. A second analysis using stepwise forward logistic regression was performed to explore various combinations of individual items. We then undertook 2 possible analytic approaches that might form the basis for an IBP screening tool. In one approach, each question item that was significant by regression analysis received a score of 1, and scores were summed (additive tool). In a second approach, beta-coefficients for significant items by logistic

### The Calin Questionnaire

1. Have you ever had trouble with your back (excluding the neck)? [presence of back pain ever]  
Yes\_\_ No\_\_
2. Have you had any discomfort (pain and/or stiffness) in your back today? [presence of back pain today]  
Yes\_\_ No\_\_
3. Has the discomfort ever gone on for three months [chronicity of pain]\*  
Yes\_\_ No\_\_
4. Has the back been stiff, especially in the morning? [morning stiffness]\*  
Yes\_\_ No\_\_
5. At what age did you discover your back discomfort (yr)? [age of onset]\*  
Age = \_\_ years old
6. Did the problem begin slowly? [insidious onset]\*  
Yes\_\_ No\_\_
7. Has the discomfort been improved by exercise? [improves with exercise]\*  
Yes\_\_ No\_\_

### The 6-item Questionnaire

1. Do you experience stiffness in your back and/or hips? [morning stiffness]  
Yes\_\_ No\_\_  
If yes, when is this most noticeable? Please tick only one option  
A. When I get out of bed\_\_\*\*  
B. Afternoon\_\_  
C. Evening\_\_  
D. No consistent time\_\_
2. Have you been woken up by back and/or hip pain during sleep? [nocturnal pain]  
Please tick one option:  
Often\_\_\*\* Sometimes\_\_\*\* Almost never\_\_  
If yes, then do you wake up:  
Early during the night?\_\_  
In the early morning?\_\_\*\*
3. At what time of day are your back and/or hip symptoms the worst? [diurnal variation]  
Please only tick one option:  
Morning\_\_\*\* Afternoon\_\_ Evening\_\_ Night\_\_ Not applicable\_\_
4. Have you ever had painful or swollen joints? [peripheral joint involvement]  
Yes\_\_\*\* No\_\_  
If yes, what joints?\_\_\_\_\_
5. What effect does exercise have on your back and/or pain? [response to exercise]  
Please only tick one option:  
A. Usually makes it worse\_\_  
B. No consistent effect\_\_  
C. Usually makes it better\_\_\*\*
6. What effect does lying down and taking a rest have on your back and/or hips stiffness? [response to rest]  
Please only tick one option:  
A. Usually makes it worse\_\_\*\*  
B. No consistent effect\_\_\*\*  
C. Usually makes it better\_\_  
D. Not applicable\_\_

\*Questions 1 and 2 are screening for the presence of back pain. The actual Calin criteria include questions 3–7 inclusive. These represent the 5 Calin question items that screen for inflammatory back pain. If 4 out of 5 of questions 3–7 are positive, then the sensitivity for IBP is 95% and specificity for IBP is 76% compared with the mechanical back pain group<sup>14</sup>. \*\*Indicates the prespecified positive responses for IBP in the 6-item questionnaire and which responses were therefore dichotomized as ‘yes’ versus ‘no’ for the statistical analysis.

Figure 1. The Calin questionnaire and the 6-item questionnaire with [associated domains].

regression were used to derive weights for the contribution of that item to the total score (weighted tool). With both the additive and weighted tools, all domains to which the patient answered affirmatively for IBP received a value

of 1, and were entered into the formula. Therefore, when patients responded to “diurnal variation” (At what time of day are your back and/or hip symptoms the worst?) as “morning,” or to “response to exercise” (What effect does



exercise have on your back and/or pain?) as “usually makes it better,” or to “response to rest” (What effect does lying down and taking a rest have on your back and/or hip stiffness?) as “usually makes it worse,” they received a value of 1 per domain. Any of the other possible responses per domain (question item) would therefore be negative for IBP and receive a score of 0. The final score reflected the sum total (additive score range = 0–3; weighted score =  $0.6 \times$  diurnal variation +  $0.24 \times$  response to exercise +  $0.16 \times$  response to rest, range 0–1). Receiver operating characteristic (ROC) analysis to calculate the area under the curve (AUC) as well as the balance between sensitivity, specificity, and likelihood ratios were used to provide optimal cutoffs for each tool in screening for IBP. These cutoffs were then used to compare sensitivity, specificity, and likelihood ratios for IBP with the Calin criteria and individual items from the 6-item questionnaire. In general,  $AUC \leq 0.5$  is likened to chance, 0.6–0.7 of slight clinical value, 0.7–0.8 of modest clinical value, and  $> 0.8$  considered clinically very useful<sup>22</sup>. Subgroup analyses were performed to determine classification utility according to sex, age, and symptom duration, evaluating the performance of the questionnaires for early disease populations (age  $< 40$  years, symptom duration  $< 6$  years, male). Chi-square analysis was used to assess for a statistical difference between the AUC for the Calin and respective candidate tools.

We also conducted a posthoc comparison of the performance of the recently reported ASAS IBP criteria and the 6-item questionnaire by comparing sensitivity, specificity, LR, and ROC analysis when at least 4 out of the 5 domains from the ASAS IBP criteria were present as required for the diagnosis of IBP (improvement with exercise, pain at night, insidious onset, age at onset  $< 40$  yrs, no improvement with rest). All these domains are included between the Calin criteria and the 6-item questionnaire. A posthoc analysis comparing to the Berlin criteria could not be performed because the study was designed before these were published and therefore particular domains including “alternating buttock pain” were not included<sup>15</sup>. Analyses were conducted using Stata 9.

## RESULTS

**Patient characteristics.** The established back pain cohort consisted of 286 patients, 220 with IBP and 66 with MBP (Table 1). Overall, mean age was 44 years, 195 were men, and average symptom duration was 14 years. The mean age for patients with MBP was 49 years (range 18–82) compared to 40 years (range 14–73) for the IBP group ( $p \geq 0.5$ ). Mean disease duration for the MBP group was 13 years (range 0.3–61) and for the IBP group 15 years (range 0.2–52;  $p \geq 0.5$ ). A diagnosis of AS was made in 209 of the patients with IBP, with the definite diagnosis unavailable for 11 patients with IBP. The etiology of MBP for the 66 patients included degenerative disc disease, facet arthritis, and herniated disc. In the majority of cases, however, the specific MBP diagnosis was not

recorded in the questionnaire. Eighty-one patients (28%) had symptom duration  $< 6$  years, including 53 patients (24%) with IBP and 28 patients (42%) with MBP. There was a male predominance in the IBP group (170 men, 50 women) in contrast to a female predominance in the MBP group (41 women, 25 men). The frequency of HLA-B27 positivity was 82% (106 patients), with HLA-B27 status known for 130 patients with IBP.

**Univariate analysis and screening utility of individual items.** The frequency of positive responses to the 6-item and Calin question items (domains) that achieved statistical significance in differentiating IBP and MBP patients in the univariate analysis was similar between the 2 questionnaires, except for “morning stiffness” (Table 2). The Calin question identified morning stiffness in 199 of those with IBP (90%) versus 45 MBP patients (68%). The 6-item question for morning stiffness, which was broken down into 2 parts (“Do you experience stiffness in your back and/or hips”; “If yes, when is this most noticeable”), did not identify as many IBP patients with morning stiffness [ $n = 95$  (48%)] compared with the Calin question. While 207 IBP patients (94%) answered “yes” to the presence of stiffness in the back and/or hips, only 95 of these IBP patients (48%) answered in the affirmative when the “time of day of stiffness” was added to the question. However, specificity was considerably higher at 75%, compared to the Calin morning stiffness question (32%).

The domains in the 6-item questionnaire with the greatest difference in frequency between IBP and MBP included “diurnal variation” ( $p = 0.0001$ ), “response to rest” ( $p = 0.01$ ), and “response to exercise” ( $p = 0.001$ ; Table 2). “Diurnal variation” was reported in only 5 MBP patients (8%) compared with 99 IBP patients (49%). “Response to exercise” was significantly different between IBP and MBP for both questionnaires. “Insidious onset,” a domain recorded only in the Calin criteria, was not helpful in differentiating back pain, being present in 80% of IBP and 76% of MBP patients. “Nocturnal pain” had low sensitivity and specificity for IBP and was therefore not significant in univariate analysis.

Overall, more individual domains were strongly associated with IBP in the 6-item questionnaire (Table 2) compared to the Calin criteria in the univariate analysis. While the individual Calin criteria domains strongly associated with IBP included “age of onset  $< 40$  yrs” (not tested in the 6-item questionnaire), “response to exercise,” and “morning stiffness,” the 6-item questionnaire included 2 additional domains significantly associated with IBP, namely “diurnal variation” (not tested in the Calin criteria) and “response to rest.” The domain “diurnal variation” in the 6-item questionnaire had the highest classification utility of all the individual question items from the Calin and 6-item questionnaires, with a sensitivity of 49% and specificity of 92% and positive LR 6.0, negative LR 0.6.

**Multivariate analysis and screening utility of combined items.** Multivariate logistic regression analysis showed that the “diurnal variation” domain from the 6-item questionnaire was

Table 1. Background characteristics of the patients with mechanical and inflammatory back pain. Data are mean  $\pm$  SD (95% CI).

Characteristic	IBP, n = 220	MBP, n = 66
Age, yrs	39.7 $\pm$ 12.6 (38.0–41.4)	48.7 $\pm$ 15.6 (44.9–52.6)
Symptom duration, yrs	15.3 $\pm$ 11.2 (13.8–16.8)	12.7 $\pm$ 14.4 (9.2–16.3)
Pain numerical rating scale (0–10)	5.8 $\pm$ 2.5 (5.3–6.2)	6.2 $\pm$ 2.3 (5.4–7.0)
Sex, male:female	170:50	25:41

IBP: inflammatory back pain; MBP: mechanical back pain.

Table 2. Classification utility of response to IBP domain questions in patients with inflammatory back pain (IBP) (n = 220) or mechanical back pain (MBP) (n = 66).

Domain	Criteria	No. IBP (%)	No. MBP (%)	OR (95% CI)	p	Sensitivity, %	Specificity, %	LR + (95% CI)	LR - (95% CI)
Morning stiffness	Calin	199 (90)	45 (68)	4.42 (2.23–8.78)	0.0001	90	32	1.33 (1.12–1.57)	0.30 (0.18–0.51)
	6-item	95 (48)*	16 (25)*	2.80 (1.49–5.25)	0.001	48	75	1.94 (1.24–3.04)	0.69 (0.57–0.84)
Response to exercise	Calin	136 (67)	21 (38)	3.43 (1.86–6.36)	0.0001	67	63	1.80 (1.26–2.55)	0.52 (0.39–0.69)
	6-item	116 (53)*	20 (30)*	2.62 (1.45–4.71)	0.001	53	70	1.76 (1.19–2.58)	0.67 (0.54–0.83)
Diurnal variation	6-item	99 (49)*	5 (8)	10.66 (4.10–27.71)	0.0001	49	92	5.95 (2.54–13.94)	0.56 (0.48–0.65)
Response to rest	6-item	131 (60)	28 (42)	2.07 (1.18–3.61)	0.010	60	58	1.42 (0.69–1.05)	0.69 (0.53–0.9)
Nocturnal pain	6-item	97 (51)*	22 (38)*	1.71 (0.93–3.12)	0.080	51	62	1.35 (0.94–1.92)	0.79 (0.62–1.01)
Age of onset (< 40 yrs)	Calin	203 (93)	45 (70)*	5.71 (2.70–27.71)	0.0001	93	30	1.32 (1.12–1.56)	0.23 (0.13–0.43)
Insidious onset	Calin	170 (80)	48 (76)	1.24 (0.63–2.41)	0.534	80	24	1.05 (0.90–1.22)	0.85 (0.51–1.42)

\* Denominator is not 220 for IBP or 66 for MBP due to missing responses. LR: likelihood ratio positive/negative.

independently associated with IBP regardless of whether the domains “response to rest” and “response to exercise” were combined or separated in the analyses (adjusted OR 11.18 and 9.37, respectively,  $p = 0.0001$  for both; Table 3). Two additional domains from the 6-item questionnaire were independently associated with IBP in multivariate analyses, including “response to rest” and “response to exercise” (Table 3). In the 6-item questionnaire, the “morning stiffness” domain was not independently associated with IBP ( $p > 0.5$ ) in contrast to the Calin item for “morning stiffness” (adjusted OR 3.59, 95% CI 1.57–8.22,  $p = 0.002$ ).

We combined items from the 6-item questionnaire that were independently associated with IBP in multivariate analysis by either simple summation (additive score, range 0–3) or by application of item weighting based on the beta-coefficient from logistic regression analysis (weighted score =  $0.6 \times$  diurnal variation +  $0.24 \times$  response to exercise +  $0.16 \times$  response to rest). Optimal cutoff scores for the additive and weighted scores in ROC analysis were 2 and 0.4, respectively. The weighted score had a higher sensitivity (60%) compared to the additive score (54%) with comparable specificity (89%–90%), but neither was superior to the classification utility of the

single item “diurnal variation” (sensitivity 49%, specificity 92%; Table 4). Both combined item scores and “diurnal variation” outperformed the Calin criteria (sensitivity 83%, specificity 42%). The AUC from ROC analysis for the weighted score (0.77) and single domain “diurnal variation” (0.70) were greater than that seen with the Calin criteria (0.62;  $p = 0.026$  and  $p = 0.027$ , respectively). The AUC for the Calin criteria (0.62) was not statistically different from the additive (0.75) score ( $p = 0.248$ ). The ASAS IBP criteria did not perform as well as either the single item “diurnal variation” or the additive or weighted scores in identifying IBP (Table 4).

*Subgroup analysis according to sex and symptom duration.* The classification utility of the additive and weighted scores and “diurnal variation” was further assessed in those patients in the lowest tertile of symptom duration, which was < 6 years (Table 5). Both the additive and weighted scores and “diurnal variation” performed even better, with high specificities (> 95%) and increased positive LR (12 to 15) in these patients. Classification utility was also somewhat higher in women, although sample size was small and 95% CI were wide.

Table 3. Multivariate logistic regression analysis of inflammatory back pain (IBP) items significant by univariate analysis to determine independent association with IBP.

Criteria	Domain	OR (95% CI)	p
Calin	Morning stiffness	3.59 (1.57–8.22)	0.002
	Age of onset < 40 yrs	3.07 (1.17–8.10)	0.023
	Response to exercise	3.26 (1.70–6.26)	0.0001
6-item questionnaire	Morning stiffness	0.61 (0.26–1.40)	0.243
	Diurnal variation	11.18 (3.66–34.16)	0.0001
	Response to exercise	2.38 (1.21–4.70)	0.012
	Response to rest	1.99 (1.02–3.87)	0.043
6-item questionnaire	Morning stiffness	0.80 (0.35–1.82)	0.589
	Diurnal variation	9.37 (3.11–28.29)	0.0001
	Combination “response to rest” + “response to exercise”	4.71 (1.57–14.17)	0.006

Table 4. Classification utility of candidate tools for inflammatory back pain (IBP): (1) additive score, (2) weighted score, (3) “diurnal variation” single item, (4) Calin criteria, and (5) ASAS IBP according to expert criteria.

Score	Sensitivity, %	Specificity, %	LR+ (95% CI)	LR- (95% CI)	AUC
Additive score* $\geq 2$	54	90	5.52 (2.55–11.92)	0.51 (0.43–0.60)	0.75
Weighted** $\geq 0.4$	60	89	5.51 (2.57–10.56)	0.45 (0.37–0.55)	0.77
Diurnal variation single item	49	92	5.95 (2.54–13.94)	0.56 (0.48–0.65)	0.70
Calin criteria	83	42	1.42 (1.14–1.78)	0.41 (0.27–0.62)	0.62
ASAS IBP criteria	59	66	1.72 (1.19–2.47)	0.62 (0.48–0.79)	0.62

\* Additive tool domains are “diurnal variation,” “response to rest,” “response to exercise” from the 6-item questionnaire. \*\* Weighted score =  $(0.6 \times \text{diurnal variation} + 0.24 \times \text{response to exercise} + 0.16 \times \text{response to rest})$  from the 6-item questionnaire. ASAS: Assessment of Spondyloarthritis International Society; LR: likelihood ratio positive/negative; AUC: area under the curve analysis.

Table 5. Classification utility of candidate tools for inflammatory back pain stratified by sex and symptom duration.

Factor	Scoring Tool	Sensitivity, %	Specificity, %	LR+ (95% CI)	LR- (95% CI)
Sex					
Male	Additive $\geq 2$	55	79	2.64 (1.19–5.82)	0.57 (0.49–0.75)
Female	Additive $\geq 2$	52	97	19.30 (2.74–136.00)	0.49 (0.36–0.67)
Male	Weighted $\geq 0.4$	59	75	2.38 (1.18–4.82)	0.54 (0.40–0.73)
Female	Weighted $\geq 0.4$	61	97	22.52 (3.21–157.85)	0.40 (0.28–0.58)
Male	Diurnal variation	47	79	2.26 (1.02–5.02)	0.67 (0.52–0.86)
Female	Diurnal variation	54	100	—*	0.46 (0.34–0.62)
Symptom duration, yrs					
$\geq 6$	Additive $\geq 2$	54	86	3.89 (1.70–8.89)	0.53 (0.43–0.66)
< 6	Additive $\geq 2$	55	96	13.78 (1.99–95.57)	0.47 (0.34–0.64)
$\geq 6$	Weighted $\geq 0.4$	59	83	3.52 (1.67–7.40)	0.50 (0.39–0.63)
< 6	Weighted $\geq 0.4$	63	96	15.82 (2.29–109.21)	0.38 (0.26–0.56)
$\geq 6$	Diurnal variation	58	89	5.48 (1.46–20.61)	0.47 (0.35–0.64)
< 6	Diurnal variation	48	96	12 (1.72–83.67)	0.54 (0.41–0.72)

\* Incalculable (100% specificity). LR: likelihood ratio positive/negative.

## DISCUSSION

We developed a 6-item questionnaire to classify IBP in the clinical setting with future intention to validate the final product in target populations and primary care to ultimately screen for early SpA. We found that our 6-item self-report questionnaire more effectively identified IBP than either the Calin or the ASAS IBP criteria in a population with established back pain. In particular, the single domain “diurnal variation” had the strongest association with IBP in our study. This has not been evaluated previously in any screening tools for IBP. In contrast, the similar domain “morning stiffness” had limited classification utility in both the 6-item self-report and Calin questionnaires and combinations of items did not reveal performance that was superior to the single item diurnal variation domain. Diurnal variation also outperformed the Calin and ASAS IBP criteria.

The Calin criteria were developed from the clinical history of IBP in established AS, whereas the 6-item questionnaire was derived by expert consensus followed by development of standardized questions. The Calin questionnaire is not completed by the patient but rather the rheumatologist, and lacks precise wording for each domain. The questionnaire is only

completed if the patient has ever had back pain, whereas the 6-item questionnaire is completed if one has had back and/or hip pain, thereby acknowledging that many patients refer to the buttock region as the “hip”<sup>23</sup>. “Insidious onset” was omitted from the 6-item questionnaire because investigators felt the term “insidious” is vague and not well defined. Nocturnal pain has been suggested as a characteristic feature of IBP, but it did not attain statistical significance as set out in the 6-item questionnaire. This may relate to patient difficulty in interpreting the wording of this question. The expected answer for IBP would be for pain in the early morning or “second half of the night,” but interpretation of this question and its answer for the IBP compared to the MBP patient was evidently varied and did not achieve statistical significance<sup>15</sup>. A patient’s understanding of “second half of the night” may be mixed, leading to heterogeneity in responses between MBP and IBP patients (e.g., second half of the night might be midnight onward versus after 3 AM). It is difficult to develop a standardized question that adequately conveys the meaning of this complex domain and our data should not necessarily be interpreted as diminishing the importance of this feature in the evaluation of IBP.

Diurnal variation may have performed so well in our study because the question wording does not address specific symptoms, but rather the time of day when they are most noticeable. While specificity was high (92%), sensitivity was low, at 49%, in the overall group of patients with AS, although higher (58%) in those with symptom duration < 6 years. It is possible that this might reflect disease activity, with more active patients having more obvious diurnal variation in symptomatology. Our data support the conclusions of the ASAS IBP study, which indicated that morning stiffness does not independently discriminate between IBP and nonspecific causes of back pain<sup>20</sup>.

Our 6-item questionnaire did not consider the Berlin criteria or ASAS expert IBP criteria because they were not published when we developed our questionnaire. However, the Berlin criteria do not detail how the questions can be standardized, leaving specific domains such as “alternating buttock pain” open to interpretation<sup>15</sup>. Moreover, a posthoc analysis in our study incorporating the 5 domains comprising the ASAS IBP criteria did not perform as well as 1 domain, “diurnal variation,” or the additive and weighted scores. The Berlin criteria comprise 4 domains, “morning stiffness,” “improvement of back pain with exercise and not rest,” “nocturnal pain,” and “alternating buttock pain,” while the ASAS IBP criteria do not include “morning stiffness” or “alternating buttock pain” but add “age less than 40” and “insidious onset” to the other 3 domains, including “improvement with exercise,” “nocturnal pain,” and “no improvement with rest.” The Berlin and ASAS criteria may perform better when administered and critically assessed by clinicians with special expertise in musculoskeletal disorders, but our data suggest that they are unlikely to be useful in primary care practice, where physicians may lack sufficient experience with different presentations of IBP<sup>12</sup>. The utility to the primary care physician of a single question, “diurnal variation,” or even the self-administered 6-item questionnaire improves the screening of patients with chronic back pain in the general population for SpA.

The utility of self-report questionnaires with standardized questions was recently demonstrated in a case-ascertainment tool for AS, where patient self-report for IBP screening was used<sup>13</sup>. Weisman, *et al* identified candidate question items with input from an advisory board, and then drafted and revised a case-ascertainment tool that was validated in a case-control study of 145 cases of AS and 308 cases of chronic back pain with mechanical back pain<sup>13</sup>. The resulting patient-reported 12-question items had a sensitivity of 67.4% and specificity of 94.6% and included several features of IBP including “neck and/or hip pain/stiffness” and “improvement with daily physical activity.” Ultimately, both the case-ascertainment tool and single item “diurnal variation” strive to identify patients with higher probabilities of having SpA who warrant further evaluation. The case-ascertainment tool did have a greater sensitivity and specificity than “diurnal varia-

tion.” However, if the ultimate goal is to identify more patients with potential SpA by creating a tool useful in the primary care office or even a Web-based format, the feasibility of a 12-item instrument compared to a single question raises concern. If there is a slight tradeoff in sensitivity and/or specificity but a greater likelihood for regular use of this question in primary care and specialist offices alike, there is a greater likelihood of identifying patients with IBP requiring further investigation for SpA. A recent study evaluating referral recommendations for patients with chronic back pain by primary care physicians and orthopedists demonstrated the utility of educating primary doctors on SpA features such as IBP<sup>24</sup>. However, even that study included more than one IBP domain among other recommendations for screening (e.g., HLA-B27).

The primary limitation of our study is that patients had established disease, and the utility of the “diurnal variation” domain requires further testing in a primary care population. The gold standard for diagnosing SpA was the rheumatologist’s diagnosis rather than formal criteria such as the modified New York criteria, although it is likely that most practicing rheumatologists would have used these criteria, as the study was conducted before the appearance of the ASAS criteria for SpA diagnosis. There is a possibility that screening utility of our questionnaire will be lower in unselected patient populations. Nevertheless, our data suggest that primary care physicians should be made aware of diurnal variation as a potentially important characteristic of IBP, and consider referral to a rheumatologist if the patient is young and especially if HLA-B27-positive<sup>25</sup>. In addition, further studies aimed at developing screening strategies for IBP should incorporate diurnal variation into the study design. Comparing sensitivity/specificity of the 6-item questionnaire to either the Calin or ASAS IBP criteria, respectively, may also be problematic because the Calin and ASAS questionnaires were designed for administration by the interviewing clinician, compared to the 6-item questionnaire that was designed for patient self-report. Therefore, the inferior performance of the Calin or ASAS IBP criteria may reflect this methodological difference in how these questionnaires were designed.

To our knowledge, this is the first report that formally identifies the importance of “diurnal variation” in the diagnosis of IBP. While there are many criteria for IBP based on expert consensus and overlapping domains, none has demonstrated the same initial classification utility as shown with the single domain of diurnal variation. We also suggest that self-report standardized questions can be helpful in screening for a complex disorder such as IBP, and that this may be a more feasible and pragmatic approach than the administration of multiitem tools assessing overlapping domains that are unfamiliar to most primary care physicians. Our proposed approach to screening for IBP, using the single item of diurnal variation completed by the patient, deserves further testing in larger, unselected cohorts of patients and those at higher risk for SpA.



## ACKNOWLEDGMENT

The authors thank Dr. Anthony S. Russell for his contribution to the initial development of the 6-item questionnaire.

## REFERENCES

1. 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.
2. Rudwaleit M, van der Heijde D, Landewe R, Listing J, Akkoc N, Brandt J, 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.
3. Braun J, Sieper J. Ankylosing spondylitis. *Lancet* 2007;369:1379-90.
4. Braun J, Sieper J. Biological therapies in the spondyloarthritis — the current state. *Rheumatology* 2004;43:1072-84.
5. Mau W, Zeidler H, Mau R, Majweski A, Freyschmidt J, Stangel W, et al. Clinical features and prognosis of patients with possible ankylosing spondylitis. Results of a 10-year follow-up. *J Rheumatol* 1988;15:1109-14.
6. Rudwaleit M, Haibel H, Baraliakos X, Listing J, Marker-Hermann E, Zeidler H, et al. The early disease stage in axial spondyloarthritis. Results from the German Spondyloarthritis Inception Cohort. *Arthritis Rheum* 2009;60:717-27.
7. Braun J, Sieper J. Building consensus on nomenclature and disease classification for ankylosing spondylitis: Results and discussion of a questionnaire prepared for the International Workshop on New Treatment Strategies in Ankylosing Spondylitis, Berlin, Germany, 18-19 January 2002. *Ann Rheum Dis* 2002;61 Suppl 3:61-7.
8. Braun J, Brandt J, Listing J, Zink A, Alten R, Golder W, et al. Treatment of active ankylosing spondylitis with infliximab: A randomised controlled multicentre trial. *Lancet* 2002;359:1187-93.
9. Haibel H, Rudwaleit M, Listing J, Heldmann F, Wong RL, Kupper H, et al. Efficacy of adalimumab in the treatment of axial spondylarthritis without radiographically defined sacroiliitis: Results of a twelve-week randomized, double-blind, placebo-controlled trial followed by an open-label extension up to week fifty-two. *Arthritis Rheum* 2008;58:1981-91.
10. Barkham N, Coates LC, Keen H, Hensor E, Fraser A, Redmond A, et al. Double-blind placebo-controlled trial of etanercept in the prevention of work disability in ankylosing spondylitis. *Ann Rheum Dis* 2010;69:1926-8.
11. Underwood MR, Dawes P. Inflammatory back pain in primary care. *Br J Rheumatol* 1995;34:1074-7.
12. Jois RN, Macgregor AJ, Gaffney K. Recognition of inflammatory back pain and ankylosing spondylitis in primary care. *Rheumatology* 2008;47:1364-6.
13. Weisman MH, Chen L, Clegg DO, Davis JC, Dubois RW, Prete PE, et al. Development and validation of a case ascertainment tool for ankylosing spondylitis. *Arthritis Care Res* 2010;62:19-27.
14. Calin A, Porta J, Fries JF, Schurman DJ. Clinical history as a screening test for ankylosing spondylitis. *JAMA* 1977;237:2613-4.
15. Rudwaleit M, Metter A, Listing J, Sieper J, Braun J. 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.
16. Dougados M, van der Linden S, Juhlin R, Huitfeldt B, Amor B, Calin A, et al. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. *Arthritis Rheum* 1991;34:1218-27.
17. Amor B, Dougados M, Mijiyawa M. Criteria of the classification of spondyloarthropathies [French]. *Rev Rheum Mal Osteoartic* 1990;57:85-9.
18. Gran JT. An epidemiological survey of the signs and symptoms of ankylosing spondylitis. *Clin Rheumatol* 1985;4:161-9.
19. Van der Linden SM, Fahrer H. Occurrence of spinal pain syndromes in a group of apparently healthy and physically fit sportsmen (orienteers). *Scand J Rheumatol* 1988;17:475-81.
20. Sieper J, van der Heijde D, Landewe R, Brandt J, Burgos-Vagas R, Collantes-Estevez E, 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.
21. Bennett AN, Marzo-Ortega H, Emery P, McGonagle D, Leeds Spondyloarthropathy Group. Diagnosing axial spondyloarthropathy. The new Assessment in Spondyloarthritis international Society criteria: MRI entering centre stage. *Ann Rheum Dis* 2009;68:765-7.
22. Ohman EM, Granger CB, Harington RA, Lee KL. Risk stratification and therapeutic decision making in acute coronary syndromes. *JAMA* 2000;284:876-8.
23. Elyan M, Khan MA. Diagnosing ankylosing spondylitis. *J Rheumatol Suppl.* 2006 Sep;78:12-23.
24. Brandt HC, Spiller I, Song I, Vahldiek JL, Rudwaleit M, Sieper J. Performance of referral recommendations in patients with chronic back pain and suspected axial spondyloarthritis. *Ann Rheum Dis* 2007;66:1479-84.
25. Rudwaleit M, van der Heijde D, Khan MA, Braun J, Sieper J. How to diagnose axial spondyloarthritis early. *Ann Rheum Dis* 2004;63:535-43.