

Impact of the positive pathergy test on the performance of classification/diagnosis criteria for Behcet's disease

Fereydoun Davatchi · Bahar Sadeghi Abdollahi · Cheyda Chams-Davatchi · Farhad Shahram · Zahra Ghodsi · Abdolhadi Nadji · Massoomeh Akhlaghi · Tahereh Faezi · Hormoz Shams · Roghieh Larimi · Farima Ashofteh

Received: 26 December 2011 / Accepted: 20 February 2012 / Published online: 4 April 2012
© Japan College of Rheumatology 2012

Abstract

Background The only diagnostic test that currently exists for Behcet's disease (BD) is the pathergy test. A positive pathergy test (PPT) is an important component of many of the 16 sets of classification/diagnosis criteria used to diagnose BD. The aim of this study was to determine the importance of a PPT in the performance of the diagnosis/classification criteria for BD.

Patients and methods All patients listed in the BD registry of the Rheumatology Research Center, Tehran (6,727) and 4,648 BD controls were enrolled in the study. The diagnosis was clinical when no other diagnosis could explain the patient's manifestations. The criteria were tested with and without PPT results. Sensitivity, specificity, and accuracy were calculated.

Results Without PPT, all sets of criteria lost sensitivity, gained specificity, and lost accuracy, with the exception of the Cheng–Zhang criteria. The largest loss in sensitivity was for the Hubault–Hamza (35 %) and Dilsen (17.3 %) criteria; the least was for the Curth (1.9 %) and ICBD (6.5 %) criteria. The largest gain on specificity was for the Dilsen (4.7 %) and Curth (3.1 %) criteria; the least was for the Japan (0.1 %) and Japan revised (0.1 %) criteria. The greatest loss in accuracy was for the Hubault–Hamza

(20.4 %) and Dilsen revised (9.3 %) criteria; the least was for the ICBD (3.6 %), while Curth gained 0.3 %.

Conclusion Without PPT as a criterion for the diagnosis of BD, the sensitivity and accuracy of the sets of classification/diagnosis criteria decrease, while the specificity improves.

Keywords Behcet's disease · Diagnosis · Criteria for diagnosis · Vasculitides

Introduction

Behcet's disease (BD) has long been recognized as a disease entity (fifth century BC), and its description can be found in the third book of Hippocrates on endemic diseases [1]. In the context of modern medicine, it was first described by a Greek ophthalmologist, Adamantiades, in 1930 and 1931 [2, 3] and then as a separate entity in 1937 by Behçet, a Turkish dermatologist [4]. The disease is called BD, but also Adamantiades–Behcet's disease (ABD), by some authors [5]. There is no specific laboratory test for the diagnosis of the disease, with the exception of the pathergy test that uses the pathergy phenomenon of the disease [6]. This test was first described in 1937 by Blobner [7] and quickly became an important tool for the diagnosis of the disease [8, 9]. Different methods are used to perform the pathergy test. However, the basic technique consists of puncturing the skin with a needle. In the case of a positive pathergy test (PPT), within 24–48 h post-puncture, a papulopustular reaction will appear at the site of the needle puncture, which is surrounded by an erythematous reaction [6, 9–13]. The shape (21–25 gauge needle) and sharpness of the needle, the angle of penetration into the skin, and the time delay for reading the reaction varies depending on the

F. Davatchi (✉) · B. Sadeghi Abdollahi · C. Chams-Davatchi · F. Shahram · Z. Ghodsi · A. Nadji · M. Akhlaghi · T. Faezi · H. Shams · R. Larimi · F. Ashofteh
Behcet's Disease Unit, Rheumatology Research Center, Shariati Hospital, Tehran University of Medical Sciences, Jalal Al Ahmad/Kargar Avenue, 14117 Tehran, Iran
e-mail: fddh@davatchi.net

center performing the test. The pathergy test can measure disease progress, such as attacks and remissions, as well as other manifestations of the disease [14]. The incidence of PPT varies depending on the country. Overall, the incidence is higher in the ‘Silk Road’ countries [15], but it is declining gradually in some countries [16]. Although the sensitivity of the test has decreased, its specificity has increased along with its positive predictive value (PPV), positive likelihood ratio (PLR), and the diagnostic odds ratio (DOR) [17]. The pathergy test is used in many sets of classification/diagnosis criteria [18].

BD, although considered to be a ‘young’ disease, has already 16 sets of classification/diagnosis criteria. Curth presented the first set of criteria in 1946 [19], less than 10 years after the official description of the disease appeared. Other sets of classification/diagnosis criteria have followed: in 1969 by Hewitt et al. [20] and Mason and Barnes [21], respectively, in 1971 by Hewitt and colleagues (revision of their original criteria [22]), in 1972 by the Japan Research Committee for BD (Japan criteria [23]), in 1974 by Hubault and Hamza [24] and O’Duffy [25], respectively, in 1980 by Cheng and Zhang [26], in 1986 by Dilsen et al. [27], and in 1988 by Japan Research Committee for BD (Japan revised criteria [28]). In near 42 years since the first description of the BD, ten sets of classification/diagnosis criteria have been proposed for BD, and there is as yet no consensus on any of them. In 1990, the International Study Group on BD, formed by seven countries (France, Iran, Japan, Tunisia, Turkey, UK, and USA), proposed the ISG criteria [29]. However, due to their low sensitivity and high specificity [30–36], no consensus was reached on these criteria, and other criteria have continued to be presented, including the Iran criteria in 1993 [30], the Classification Tree in 1993 [37], the Dilsen revised criteria in 2000 [38], and the Korean criteria in 2003 [39, 40]. In 2004, an International Team from 27 countries (Austria, Azerbaijan, China, Egypt, France, Germany, Greece, India, Iran, Iraq, Israel, Italy, Japan, Jordan, Libya, Morocco, Pakistan, Portugal, Russia, Saudi Arabia, Singapore, Spain, Taiwan, Thailand, Tunisia, Turkey, and USA) was formed. This Team presented the International Criteria for Behcet’s disease (ICBD) in 2006 to the 12th International Conference on BD [41, 42]. The ICBD was again presented to the 2007 American College of Rheumatology (ACR) congress [43], and has since been validated in Germany [44], China [45], and Iran [46].

The pathergy test is a critical criterion/component of 12 of the 16 sets of criteria used for diagnosing BD (the Hewitt, Mason and Barnes, Hewitt revised, and O’Duffy criteria do not use the pathergy test criterion). The aim of this study was to determine the impact of the pathergy test on the performance of those 12 criteria sets.

Patients and methods

Patients

All patients on the BD registry of the BD Unit, Rheumatology Research Center, Tehran University of Medical Sciences (Tehran, Iran) were enrolled in the study. They were diagnosed by expert clinicians as BD based on their clinical manifestations in the absence of other diseases that could explain them. The diagnosis was made basically on the presence of oral aphthosis, genital aphthosis, skin manifestations, ocular lesions, vascular lesions, and neurological symptoms. The combination of two or more manifestations led to the diagnosis of BD when no other diseases could explain their presence together and their appearance seemed to be related to each other. Each of the patients was classified at least by one of the known diagnosis/classification criteria, and the large majority of patients were classified by several diagnosis/classification criteria. Controls were selected from patients referred to the BD unit due to suspected BD, but who after a thorough evaluation were found not to have the disease.

Pathergy test

Three tests were performed on the forearm of the patient, all of which consisted of an intradermal disposable needle prick after thorough asepsis of the skin with povidone iodine 10 % (Betadine®). One prick was done with a 21-gauge needle, the second with a 25-gauge needle, and the third with a 25 gauge needle, all with the injection of one to two drops of normal saline. The test was read 24 h later by one of the two dermatologists of the BD clinic and cross-checked by one of the rheumatologists (the same rheumatologist throughout the study). A test was considered to be positive (PPT) when a papule or pustule formed on the site of the needle prick, surrounded by an erythema [6, 11, 13, 16, 17].

Methods

All criteria (except the Hewitt original criteria) were checked in all patients. The percentage of patients and controls fulfilling the criteria were calculated. The result of the pathergy test was then set to negative for all patients and controls, and the calculations were redone.

Statistical analysis

Sensitivity, specificity, and accuracy (percentage agreement) were calculated [47] together with the respective 95 % confidence interval (95 % CI). The PPV, the negative predictive value (NPV), the PLR, the negative likelihood

ratio (NLR), the DOR, and the Youden index were calculated [48–51]. For comparison purposes, the chi-square test (Pearson’s test) was used [52].

Results

A total of 6,727 BD patients and 4,648 controls were enrolled in the study. The pathergy test was positive for 51.1 % of the BD patients (95 % CI 49.9–52.3) and for 5.9 % of the control patients (95 % CI 5.2–6.6).

The BD and control patients were examined using each of the 15 sets of diagnosis/classification criteria for BD in terms of the sensitivity, specificity, and accuracy of the sets of criteria. To check the impact of the pathergy test on the performance of the criteria, we then set the results of the pathergy test to negative for all patients and controls and determined once again the sensitivity, specificity, and accuracy of the sets of criteria.

Sensitivity refers to the number of BD patients correctly classified by the criteria and in our study is expressed as the percentage of patients fulfilling the criteria. The sensitivity was 81.7 % for the Dilsen criteria, 86 % for the Japan revised criteria, 77.7 % for the ISG criteria, 97.2 % for the Classification Tree, 86.2 % for the Korean criteria, and 98.3 % for the ICBD. Full details are given in Table 1. These results for the Curth, Mason and Barnes, Hewitt, Japan’s original, Hubault and Hamza, O’Duffy, Cheng and Zhang, Iran, and Dilsen revised sets of criteria are also given in Table 1.

The sensitivity without the pathergy test fell to 64.4 % for the Dilsen criteria, 78.4 % for the Japan revised criteria, 61.7 % for the ISG criteria, 89.7 % for the Classification Tree, 76.4 % for the Korean criteria criteria, and 91.8 % for the ICBD. Full details of all criteria are given in Table 1. The loss of sensitivity was 17.3 % for the Dilsen criteria, 7.6 % for the Japan revised criteria, 16 % for the ISG criteria, 7.5 % for the Classification Tree, 9.8 % for the Korean criteria, and 6.5 % for the ICBD.

Specificity is the number of controls correctly classified as not having the disease and is expressed as the percentage of controls not fulfilling the criteria. The specificity was 95 % for the Dilsen criteria, 97.7 % for the Japan revised criteria, 99.1 % for the ISG criteria, 97.4 % for the Classification Tree, 98.2 % for the Korean criteria, and 96 % for the ICBD. Full details of all criteria are given in Table 2.

The specificity without the pathergy test improved to 99.7 % for the Dilsen criteria, 97.8 % for the Japan revised criteria, 99.8 % for the ISG criteria, 98.1 % for the Classification Tree, 98.9 % for the Korean criteria, and 96.8 % for the ICBD. Full details are given in Table 2. The increase in specificity was 4.7 % for the Dilsen criteria,

Table 1 Sensitivity

Criteria	Patients (n = 6,727)						Loss (%)
	With PPT			No PPT			
	No.	%	95 % CI	No.	%	95 % CI	
Curth	6,704	99.7	0.1	6,580	97.8	0.4	1.9
Mason and Barnes ^a	4,357	64.8	1.1	4,357	64.8	1.1	–
Hewitt ^a	2,546	37.8	1.2	2,546	37.8	1.2	–
Japan original	5,711	84.9	0.9	5,189	77.1	1.0	7.8
Hubault and Hamza	4,062	60.4	1.2	1,706	25.4	1.0	35
O’Duffy ^a	4,651	69.1	1.1	4,651	69.1	1.1	–
Cheng and Zhang	6,321	94.0	0.6	6,321	94.0	0.6	0
Dilsen	5,497	81.7	0.9	4,333	64.4	1.1	17.3
Japan revised	5,788	86.0	0.8	5,277	78.4	1.0	7.6
ISG	5,229	77.7	1.0	4,149	61.7	1.2	16
Iran	6,090	90.5	0.7	5,183	77.0	1.0	13.5
Classification Tree	6,541	97.2	0.4	6,036	89.7	0.7	7.5
Dilsen revised	5,325	79.2	1.0	4,244	63.1	1.2	16.1
Korea	5,798	86.2	0.8	5,138	76.4	1.0	9.8
ICBD	6,611	98.3	0.3	6,173	91.8	0.7	6.5

PPT Positive pathergy test, 95 % CI 95 % confidence interval, ISG International Study Group Criteria, ICBD International Criteria for Behcet’s disease

^a These sets of criteria do not include the pathergy test as a criterion

0.1 % for the Japan revised criteria, 0.7 % for the ISG criteria, 0.7 % for the Classification Tree, 0.7 % for the Korean criteria, and 0.8 % for the ICBD.

Accuracy (percentage agreement) is the overall number of patients and controls correctly classified by the criteria and is expressed by the percentage of patients fulfilling the criteria and the percentage of controls not fulfilling the criteria. It shows the overall performance of the criteria. The accuracy was 87.1 % for the Dilsen criteria, 90.8 % for the Japan revised criteria, 86.5 % for the ISG criteria, 97.3 % for the Classification Tree, 91.1 % for the Korean criteria, and 97.4 % for the ICBD. Full details for all criteria are given in Table 3.

The accuracy without the pathergy test fell to 78.8 % for the Dilsen criteria, 86.3 % for the Japan revised criteria, 77.3 % for the ISG criteria, 93.2 % for the Classification Tree, 85.6 % for the Korean criteria, and 93.8 % for the ICBD. Full details are given in Table 3. The loss in accuracy was 8.3 % for the Dilsen criteria, 4.5 % for the Japan revised criteria, 9.2 % for the ISG criteria, 4.1 % for the Classification Tree, 5.5 % for the Korean criteria, and 3.6 % for the ICBD.

The PPV, NPV, PLR, NLR, DOR, and Youden index are indicators of the different performance aspects of the sets of criteria and are all calculated from the sensitivity and the specificity of the criteria. These were calculated for each of the criteria sets, with and without the pathergy test.

Table 2 Specificity

Criteria	Patients (<i>n</i> = 4,648)						Gain (%)
	With PPT			No PPT			
	No.	%	95 % CI	No.	%	95 % CI	
Curth	4,051	87.2	1.0	4,199	90.3	0.9	3.1
Mason and Barnes ^a	4,637	99.8	0.1	4,637	99.8	0.1	–
Hewitt ^a	4,638	99.8	0.1	4,638	99.8	0.1	–
Japan original	4,543	97.7	0.4	4,547	97.8	0.4	0.1
Hubault and Hamza	4,607	99.1	0.3	4,645	99.9	0.1	0.8
O'Duffy ^a	4,593	98.8	0.3	4,593	98.8	0.3	–
Cheng and Zhang	4,340	93.4	0.7	4,340	93.4	0.7	0
Dilsen	4,415	95.0	0.6	4,633	99.7	0.2	4.7
Japan revised	4,540	97.7	0.4	4,544	97.8	0.4	0.1
ISG	4,605	99.1	0.3	4,639	99.8	0.1	0.7
Iran	4,511	97.1	0.5	4,548	97.8	0.4	0.7
Classification Tree	4,525	97.4	0.5	4,561	98.1	0.4	0.7
Dilsen revised	4,601	99.0	0.3	4,634	99.7	0.2	0.7
Korea	4,565	98.2	0.4	4,598	98.9	0.3	0.7
ICBD	4,464	96.0	0.6	4,499	96.8	0.5	0.8

^a These sets of criteria do not include the pathergy test as a criterion

PPV was calculated as (number of positive BD patients) divided by (number of positive BD patients + number of positive controls). The PPV, taking in account the result of the pathergy test or not, was 94.2 versus 99.5 for the Dilsen criteria, 97.4 versus 97.3 for the Japan revised criteria, 98.8 versus 99.7 for the ISG criteria, 97.4 versus 97.9 for the Classification Tree, 97.9 versus 98.6 for the Korean criteria, and 96.1 versus 96.6 for the ICBD. The NPV, taking in account the result of the pathergy test or not, was calculated as (number of negative control patients) divided by (number of negative controls + number of negative BD patients). The NPV was 83.8 versus 73.7 for the Dilsen criteria, 87.5 versus 81.9 for the Japan revised criteria, 81.6 versus 72.3 for the ISG criteria, 97.2 versus 90 for the Classification Tree, 87.7 versus 80.7 for the Korean criteria, and 98.3 versus 92.2 for the ICBD. The full results are given in Table 4.

PLR was calculated as (sensitivity) divided by (1 – sensitivity). The PLR, taking into account the result of the pathergy test or not, was 16.3 versus 215 for the Dilsen criteria, 37 versus 36 for the Japan revised criteria, 86 versus 308 for the ISG criteria, 37 versus 47 for the Classification Tree, 48 versus 69 for the Korean criteria, and 25 versus 29 for the ICBD. NLR was calculated as (1 – sensitivity) divided by (specificity). The NLR, taking into account the result of the pathergy test or not, was 0.19 versus 0.36 for the Dilsen criteria, 0.14 versus 0.22 for the Japan revised criteria, 0.23 versus 0.38 for the ISG criteria, 0.03 versus 0.10 for the Classification Tree, 0.14 versus 0.24 for the Korean criteria, and 0.02 versus 0.08 for the ICBD. The full results are given in Table 4.

Table 3 Accuracy

Criteria	Patients (<i>n</i> = 11,375)						Loss%
	With PPT			No PPT			
	No.	%	95 % CI	No.	%	95 % CI	
Curth	10,755	94.5	0.4	10,779	94.8	0.4	–0.3
Mason and Barnes ^a	8,994	79.1	0.7	8,994	79.1	0.7	–
Hewitt ^a	7,184	63.2	0.9	7,184	63.2	0.9	–
Japan	10,254	90.1	0.5	9,736	85.6	0.6	4.5
Hubault and Hamza	8,669	76.2	0.8	6,351	55.8	0.9	20.4
O'Duffy ^a	9,244	81.3	0.7	9,244	81.3	0.7	–
Cheng and Zhang	10,661	93.7	0.4	10,661	93.7	0.4	0
Dilsen	9,912	87.1	0.6	8,966	78.8	0.8	8.3
Japan revised	10,328	90.8	0.5	9,821	86.3	0.6	4.5
ISG	9,834	86.5	0.6	8,788	77.3	0.8	9.2
Iran	10,601	93.2	0.5	9,731	85.5	0.6	7.7
Classification Tree	11,066	97.3	0.3	10,597	93.2	0.5	4.1
Dilsen revised	9,926	87.3	0.6	8,878	78.0	0.8	9.3
Korea	10,363	91.1	0.5	9,736	85.6	0.6	5.5
ICBD	11,075	97.4	0.3	10,672	93.8	0.4	3.6

^a These sets of criteria do not include the pathergy test as a criterion

Table 4 Indicators of the different performance aspects of the sets of criteria

Criteria	Pathergy	Sensitivity	Specificity	PPV	NPV	PLR	NLR	DOR	Youden
Curth	With	99.7	87.2	88.6	99.7	7.8	0	2264	0.87
	Without	97.8	90.3	91	97.6	10.1	0.02	414	0.88
Mason and Barnes ^a	With	64.8	99.8	99.7	73.9	324	0.35	919	0.65
	Without	64.8	99.8	99.7	73.9	324	0.35	919	0.65
Hewitt ^a	With	37.8	99.8	99.5	61.6	189	0.62	303	0.38
	Without	37.8	99.8	99.5	61.6	189	0.62	303	0.38
Japan	With	84.9	97.7	97.4	86.6	37	0.15	239	0.83
	Without	77.1	97.8	97.2	81	35	0.23	150	0.75
Hubault and Hamza	With	60.4	99.1	98.5	71.4	67	0.40	168	0.60
	Without	25.4	99.9	99.6	57.2	254	0.75	340	0.25
O’Duffy ^a	With	69.1	98.8	98.3	76.2	58	0.31	184	0.68
	Without	69.1	98.8	98.3	76.2	58	0.31	184	0.68
Cheng and Zhang	With	94	93.4	93.4	94	14.2	0.06	222	0.87
	Without	94	93.4	93.4	94	14.2	0.06	222	0.87
Dilsen	With	81.7	95	94.2	83.8	16.3	0.19	85	0.77
	Without	64.4	99.7	99.5	73.7	215	0.36	601	0.64
Japan revised	With	86	97.7	97.4	87.5	37	0.14	261	0.84
	Without	78.4	97.8	97.3	81.9	36	0.22	161	0.76
ISG	With	77.7	99.1	98.8	81.6	86	0.23	384	0.77
	Without	61.7	99.8	99.7	72.3	308	0.38	804	0.62
Iran	With	90.5	97.1	96.9	91.1	31	0.10	319	0.88
	Without	77	97.8	97.2	81	35	0.24	149	0.75
Classification Tree	With	97.2	97.4	97.4	97.2	37	0.03	1300	0.95
	Without	89.7	98.1	97.9	90	47	0.10	450	0.88
Dilsen revised	With	79.2	99	98.7	82.6	79	0.21	377	0.78
	Without	63.1	99.7	99.5	73	210	0.37	568	0.63
Korea	With	86.2	98.2	97.9	87.7	48	0.14	341	0.84
	Without	76.4	98.9	98.6	80.7	69	0.24	291	0.75
ICBD	With	98.3	96	96.1	98.3	25	0.02	1388	0.94
	Without	91.8	96.8	96.6	92.2	29	0.08	339	0.89

PPV Positive predictive value, NPV negative predictive value (prevalence not taken in account), PLR positive likelihood ratio, NLR negative likelihood ratio, DOR diagnostic odds ratio, Youden Youden index
 Patients, $n = 6,727$; controls, $n = 4,648$; total study cohort, $n = 11,375$
^a These sets of criteria do not include the pathergy test as a criterion

DOR was calculated as (sensitivity X specificity) divided by (1 – sensitivity) X (1 – specificity). The DOR, taking in account the result of the pathergy test or not, was 85 versus 601 for the Dilsen criteria, 261 versus 161 for the Japan revised criteria, 384 versus 804 for the ISG criteria, 1,300 versus 450 for the Classification Tree, 341 versus 291 for the Korean criteria, and 1,388 versus 339 for the ICBD. The full results are given in Table 4.

Youden’s index was calculated as (sensitivity) + (specificity – 1). The Youden’s index, taking in account the result of the pathergy test or not, was 0.77 versus 0.64 for the Dilsen criteria, 0.84 versus 0.76 for the Japan revised criteria, 0.77 versus 0.62 for the ISG criteria, 0.95 versus 0.88 for the Classification Tree, 0.84 versus 0.75 for the Korean criteria, and 0.94 versus 0.89 for the ICBD. The full results are given in Table 4.

Discussion

The sensitivity of all sets of BD criteria in which the pathergy test is a criterion fell significantly when the results of the pathergy test were not accounted for—with the exception of the Cheng and Zhang criteria (Table 1). The greatest loss in sensitivity was seen with the Hubault and Hamza criteria (35 %), followed by the Dilsen criteria (17.3 %), Dilsen revised criteria (16.1 %), ISG criteria (16 %), and Iran criteria (13.5 %), thereby showing the dependence of these sets of criteria on the pathergy test. These tests will not work effectively in countries where the pathergy test is rarely positive, such as Western countries [12, 15, 35, 53, 54]. In these countries, criteria having the least decrease in sensitivity will be the most suitable, such as the Cheng and Zhang criteria (0 %), Curth criteria (1.9 %), ICBD (6.5 %), Classification Tree (7.5 %), Japan

revised criteria (7.6 %), and Japan original criteria (7.8 %) (Table 1). The best overall sensitivity, without taking in account the result of the pathergy test, was obtained for the Curth criteria (97.8 %), Cheng and Zhang criteria (94 %), and ICBD (91.8 %).

The specificity, which is the converse of sensitivity, improved when the results of the pathergy test were disregarded. However, the improvement was not as important as was the decrease in sensitivity. The largest gain was obtained by the Dilsen criteria (4.7 %) and Curth criteria (3.1 %), while the least improvement was obtained for the Cheng and Zhang criteria (0 %), Japan original criteria and Japan revised criteria (each 0.1 %). The other criteria were improved by 0.7–0.8 % (Table 2). The best overall specificity, without the pathergy test, was 99.9 % and was obtained by the Hubault and Hamza criteria, followed by 99.8 % for the Mason and Barnes criteria, Hewitt criteria, and ISG.

The accuracy (percentage agreement), which is the overall result of sensitivity and specificity, lost some of its performance when the results of the pathergy test were disregarded. The loss was 24 % for the Hubault and Hamza criteria, followed by the Dilsen revised criteria (9.3 %), the ISG criteria (9.2 %), the Dilsen criteria (8.3 %), and the Iran criteria (7.7 %). The least affected sets of criteria were those of Cheng and Zhang (0 % decrease), Curth (0.3 %) and the ICBD (3.6 %), the Classification Tree (4.1 %), and the Japan revised criteria (4.5 %) (Table 3). The best overall accuracy, without the pathergy test, was obtained with the Curth (94.8 %) criteria, ICBD (93.8 %), and the Cheng and Zhang criteria (93.7 %).

The PV is the probability that a test result is a true result, and the PPV is the probability that a positive test result is a true positive. It is highly influenced by the prevalence of the disease in the population in which it is tested. The prevalence of a disease will change depending on where and in which setting the patients are seen. Consequently, the results obtained in different settings will differ. In 2010, the PPV of ICBD in Iran was 95.7 % when the prevalence of BD was not taken into account [46]; this dropped to 91.7 % in the BD clinic (Rheumatology Research Center) where one-third of new patients were true BD patients (prevalence 33 %). The PPV was calculated to fall to 71.3 % in a setting with 10 % BD prevalence, to 18.4 % in a setting with 1 % BD prevalence, and to only 1.8 % if the test was done randomly in a population where the prevalence is 80 per 100,000 inhabitants [46]. The PPV improves substantially with improvement—even a minor one—of the specificity. The specificity was improved in the absence of the pathergy test and in the same manner as the PPV. The best improvement, 5.3 %, was obtained using the Dilsen criteria, followed by the Curth criteria (2.4 %) and the Hubault and Hamza criteria (1.1 %). A PPV not

improved or only slightly improved by disregarding the pathergy result indicates a better reliability of the criteria. The NPV is the probability that a negative test is truly a negative result. The NPV is also influenced by the prevalence of the disease. The higher the NPV, the more reliable is the test; a lower NPV means that fewer true patients were diagnosed. Sets of criteria that were relatively more dependent on the pathergy test, such as the Dilsen criteria, lost relatively more in NPV value than the other sets of criteria, with the highest loss found for the Dilsen criteria (10.1 %), followed by the Iran criteria (10.1 %), Dilsen revised criteria (9.6 %), and the ISG criteria (9.35). The least waste was for the Cheng and Zhang criteria (0 %) and Curth criteria (2.1 %).

The likelihood ratio (LR) shows how much the odds of having the disease may change with a positive or negative result. Prevalence does not influence the LR and, therefore, figures can be used in any disease setting. The PLR shows the odds of having the disease. When the PLR exceeds 5, the test is related to the disease. Both sensitivity and specificity improve the PLR, but specificity more than sensitivity. When the results of the pathergy test were disregarded, the sensitivity increased and the specificity improved; the increase in the former was much greater than the improvement in the latter, but not as much as to decrease the PLR. The PLR of all criteria using the pathergy test showed an improvement in their PLR, with the exception of the Japan original and Japan revised criteria. The ISG criteria showed the highest improvement, with an odds improvement of 222, followed by the Dilsen (199) and Hubault and Hamza (187) criteria. A higher improvement means that those criteria are more dependent on the result of the pathergy test. The Curth criteria, ICBD, and Classification Tree (odds improvement 23, 4, and 4, respectively) showed the lowest dependency on the pathergy test. The Cheng and Zhang criteria did not change when the pathergy results were disregarded, while the Japan criteria (original and the revised) lost 2 and 1 relative units, thereby showing their independence of the pathergy test. The NLR shows the odds of not having the disease. The odds decreased for all criteria when the pathergy test results were disregarded because NLR is more sensitive to sensitivity than specificity. The least impairment was for the Curth criteria (0.02), ICBD (0.06), and Classification Tree (0.07), while the worse impairment was for the Hubault and Hamza (0.35), Dilsen (0.17), and ISG criteria (0.15).

The DOR shows the power of discrimination of the criteria. A value of 1 means the criteria do not discriminate between patients and controls, while higher values indicate better discrimination. DOR is evenly influenced by sensitivity and specificity. The problem with the DOR is that a high sensitivity can mask a dangerous lack of specificity,

and vice et versa, especially when the value is between 99 and 100 %. The DOR decreased in the majority of criteria in the absence of the pathergy test, with the largest loss shown by the Curth criteria (1,850), ICB (1,049), and Classification Tree (850). On the contrary, the DOR improved by 516 with the Dilsen criteria, by 420 with the ISG criteria, and by 191 with the Dilsen revised criteria.

Youden's index is a simple calculation combining the results of sensitivity and specificity to show the precision or accuracy of the test. Sensitivity and specificity equally influence the Youden index. The results can vary from zero to one, with the latter being the most precise. In our study, the Youden index of all sets of criteria decreased when the pathergy test results were disregarded, but especially the Hubault and Hamza criteria (0.45), the ISG criteria (0.15), and the Dilsen revised criteria (0.15). The least impaired criteria were the ICB (0.05) and the Classification (0.07). The Cheng and Zhang criteria did not change as expected, and the Curth criteria improved by 0.01.

Conclusion

The pathergy test is the only paraclinical test that is available for diagnosing BD. Without it, the sensitivity of the majority of sets of classification/diagnosis criteria decrease by 1.9 to 35 %. On the contrary, the specificity improves by 0.1–4.7 %. Overall, the sets of criteria show a loss of performance (accuracy), demonstrating that this parameter is necessary to improve the power of existing classification/diagnosis criteria.

Conflict of interest None.

References

- Feigenbaum A. Description of Behçet's syndrome in the Hippocratic third book of endemic diseases. *Br J Ophthalmol*. 1956;40:355–7.
- Adamantiades B. A case of relapsing iritis with hypopyon (in Greek). *Archia Iatrikis Etairias (Proc Med Soc Athens)*. 1930:586–93.
- Adamantiades B. Sur un cas d'iritis à hypopyon récidivant. *Ann Ocul (Paris)*. 1931;168:271–8.
- Behçet H. Über rezidivierende Aphthosen, durch ein Virus verursachte Geschwüre am Mund, am Auge und an den Genitalien. *Derm Wschr*. 1937;105:1152–7.
- Zouboulis CC, Kaklamani P. Early descriptions of Adamantiades–Behçet's disease. *Ann Rheum Dis*. 2003;62:691–2.
- Davatchi F (2009) Behçet's disease. In: Syngle A, Deodhar SD, editors. *Rheumatology principles and practice*. New Delhi: Jaypee Brothers Medical Publishers; 2009. p. 249–68.
- Blobner F. Zur rezidivierenden Hypopyonirrit. *Z Augenheilkunde*. 1937;91:129–39.
- Katzenellenbogen I, Feuerman EJ. Beitrag zum Morbus Behçet Die Bedeutung der spezifischen Hauthyperraktivität und der Behçetin Reaktion. *Hautarzt*. 1965;16:13–8.
- Aral O, Dilsen N, Konice M. Positive skin pathergy reactivity as a genetic marker of Behçet's disease. In: Lehner T, Barnes CG, editors. *Recent advances in Behçet's disease*. London: Royal Society of Medical Services; 1986. p. 173–5.
- Mansoori P, Chams C, Davatchi F, Akbarian M, Shahram F. Pathergy phenomenon. In: Behçet's disease, new aspects. In Nasution AR, Darmawan J, Isbagio H, editors. *Rheumatology APLAR 1992*. Tokyo: Churchill Livingstone, Japan K.K.; 1992. p. 111–3.
- Dilsen N, Konice M, Aral O, Ocal L, Inanç M, Gül A. Comparative study of the skin pathergy test with blunt and sharp needles in Behçet's disease: confirmed specificity but decreased sensitivity with sharp needles. *Ann Rheum Dis*. 1993;52:823–5.
- Davatchi F. Behçet's disease. In: Howe HS, Feng PH, editors. *Textbook of clinical rheumatology*. Singapore: National Arthritis Foundation; 1998. p. 298–315.
- Davatchi F, Shahram F, Chams-Davatchi C, Shams H, Nadji A, Akhlaghi M, et al. How to deal with Behçet's disease in the daily practice. *Int J Rheum Dis*. 2010;13:105–16.
- Chams-Davatchi C, Davatchi F, Shahram F, Akbarian M, Gharibdoost F, Nadji A. Longitudinal study of the pathergy phenomenon in Behçet's disease. In: Hamza M, editor. *Behçet's Disease*. Tunisia: Pub Adhoua; 1997. p. 356–8.
- Davatchi F, Shahram F, Chams-Davatchi C, Shams H, Nadji A, Akhlaghi M, et al. Behçet's disease; from east to west. *Clin Rheumatol*. 2010;29:823–33.
- Davatchi F, Chams-Davatchi C, Shahram F, Nadji A, Shams H, Ghodsi Z, et al. Pathergy test in Behçet's disease: change in incidence over the time. *APLAR J Rheumatol*. 2007;10:333–5.
- Davatchi F, Chams-Davatchi C, Ghodsi Z, Shahram F, Nadji A, Shams H, et al. Diagnostic value of pathergy test in Behçet's disease according to the change in incidence over the time. *Clin Rheumatol*. 2011;30:1151–5.
- Barnes CG. Behçet's syndrome—classification criteria. *Ann Med Interne*. 1999;150:477–82.
- Curth HO. Recurrent genito-oral aphthosis with hypopyon (Behçet's syndrome). *Arch Dermatol*. 1946;54:179–96.
- Hewitt J, Escande JP, Laurent PH, Perlemuter L. Criteres de prevision du syndrome de Behçet. *Bulletin de la Societe Francaise de Dermatologie et de Syphiligraphie*. 1969;76:565–8.
- Mason RM, Barnes CG. Behçet's syndrome with arthritis. *Ann Rheum Dis*. 1969;28:95–103.
- Hewitt J, Escande JP, Maness S. Revision des criteres diagnostiques du syndrome de Behçet. *Presse Med*. 1971;79:901.
- Behçet's Disease Research Committee of Japan. Behçet's disease guide to the diagnosis of Behçet's disease (1972). *Jpn J Ophthalmol*. 1974;18:291–4.
- Hubault A, Hamza M. La maladie de Behçet en 1974. In: de Seze S, editor. *L'actualité Rhumatologique 15*. Paris: Expansion Scientifique; 1974. p. 43–55.
- O'Duffy JD. Critères proposés pour le diagnostic de la maladie de Behçet et notes thérapeutiques. *Rev Med*. 1974;36:2371–9.
- Cheng SP, Zhang XQ. Some special clinical manifestations of Behçet's disease—report of illustrative cases and review of literature (in Chinese) (author's translation). *Chin J Intern Med*. 1980;19:15–22.
- Dilsen N, Konice M, Aral O. Our diagnostic criteria of Behçet's disease—an overview. In: Lehner T, Barnes CG, editors. *Recent advances in Behçet's disease*. International congress and symposium series, vol. 103. London: London Royal Society of Medicine Services; 1986. p. 177–80.
- Mizushima Y. Recent research into Behçet's disease in Japan. *Int J Tissue React*. 1988;10:59–65.
- International Study Group for Behçet's Disease. Criteria for diagnosis of Behçet's disease. *Lancet*. 1990;335:1078–80.

30. Davatchi F, Shahram F, Akbarian M, Gharibdoost F, Nadji A, Chams C, et al. Accuracy of existing diagnosis criteria for Behcet's disease. In: Wechsler B, Godeau P, editors. Behcet's disease. International congress and symposium series, vol. 1037. Amsterdam: Excerpta Medica; 1993. p. 225–8.
31. Dong Y, Yao Q, Wang M. Behcet's disease in China. In: Proc 8th APLAR Congress of Rheumatology. Melbourne; 1996. p. S14.
32. Davatchi F, Shahram F, Akbarian M, Gharibdoost F, Nadji A, Chams C, et al. Behcet's disease—analysis of 3443 cases. *APLAR J Rheum.* 1997;1:2–5.
33. APLAR Subcommittee for Behcet's Disease. APLAR evaluation of Behcet's disease diagnosis criteria. *APLAR J Rheumatol.* 1998;1:237–40.
34. Prokaeva T, Reshetnyak T, Kuzin A, Rabinovich I, Ermakova N. Evaluation of existing diagnosis criteria for Behcet's disease. In: Bang D, Lee E, Lee S, editors. Behcet's disease. Seoul: Design Mecca Publishing; 2000. p. 598–603.
35. Calamia KT, Davatchi F. Analysis of the sensitivity of diagnosis criteria in United States patients with Behcet's disease. In: Bang D, Lee E, Lee S, editors. Behcet's disease. Seoul: Design Mecca Publishing; 2000. p. 121–4.
36. Davatchi F, Shahram F, Kumar A, Cheng YK, Cheong CT, Bendrups A. Comparative analysis of Behcet's disease in the APLAR region. *ALPAR J Rheumatol.* 2004;7:38–43.
37. Davatchi F, Shahram F, Akbarian M, Gharibdoost F, Chams C, Chams H, et al. Classification Tree for the diagnosis of Behcet's disease. In: Wechsler B, Godeau P, editors. Behcet's disease. International congress and symposium series, vol. 1037. Amsterdam: Excerpta Medica; 1993. p. 245–8.
38. Dilsen N. About diagnostic criteria for Behcet's disease: our new proposal. In: Bang D, Lee E, Lee S, editors. Behcet's disease. Seoul: Design Mecca Publishing; 2000. p. 101–4.
39. Chang HK, Kim SY. Survey and validation of the criteria for Behcet's disease recently used in Korea: a suggestion for modification of the International Criteria. *J Korean Med Sci.* 2003;18:88–92.
40. Chang HK, Lee SS, Bai HJ, Lee YW, Yoon BY, Lee CH, et al. Validation of the classification criteria commonly used in Korea and a modified set of preliminary criteria for Behcet's disease: a multicenter study. *Clin Exp Rheumatol.* 2004;22[Suppl 34]:S21–6.
41. International Team for the Revision of the International Criteria for Behcet's Disease. Evaluation of the international criteria for Behcet's Disease (ICBD). *Clin Exp Rheumatol.* 2006;24[Suppl 42]:S13.
42. International Team for the Revision of the International Criteria for Behcet's Disease. Revision of the international criteria for Behcet's Disease (ICBD). *Clin Exp Rheumatol.* 2006;24[Suppl 42]:S14–5.
43. Davatchi F, Schirmer M, Zouboulis C, Assad-Khalil S, Calamia KT, on behalf of the International Team for the Revision of the International Criteria for Behcet's Disease. Evaluation and revision of the International Study Group Criteria for Behcet's Disease. In: 2007 American College of Rheumatology Meeting. Boston; 2007; abstract 1233.
44. Altenburg A, Bonitsis NG, Papoutsis N, Pasak M, Krause L, Zouboulis CC. Evaluation of diagnostic criteria including ICBD (2006) in Adamantiades–Behcet's disease patients in Germany. *Clin Exp Rheumatol.* 2008;26[Suppl 50]:S3.
45. Zhang Z, Zhou W, Hao Y, Wang Y, Dong Y. Validation of the International Criteria for Behcet's disease (ICBD) in China. *Clin Exp Rheumatol.* 2008;26[Suppl 50]:S6–7.
46. Davatchi F, Sadegh Abdollahi B, Shahram F, Nadji A, Chams-Davatchi C, Shams H, et al. Validation of the International Criteria for Behcet's disease in Iran. *Int J Rheum Dis.* 2010;13:55–60.
47. Szklo M, Nieto FJ. Epidemiology, beyond the basics. 2nd ed. Sudbury: Jones and Bartlett Publishers; 2007.
48. Kirshwood BR. Sterne JAC essential medical statistics. 2nd ed. Malden: Blackwell Science; 2003.
49. Youden WJ. Index for rating diagnostic tests. *Cancer.* 1950;3:32–5.
50. Kraemer HC. Reconsidering the odds ratio as a measure of 2×2 association in a population. *Stat Med.* 2004;23:257–70.
51. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol.* 2003;56:1129–35.
52. Plackett RL. Karl Pearson and the Chi-squared test. *Int Stat Rev [Int Stat Inst (ISI)].* 1983;51:59–72.
53. Davatchi F. Behcet's disease: global perspective. *Indian J Rheumatol.* 2007;2:65–71.
54. Altenburg A, Papoutsis N, Orawa H, Martus P, Zouboulis CC. Epidemiologie und Klinik des Morbus Adamantiades–Behcet in Deutschland—Aktuelle pathogenetische Konzepte und therapeutische Möglichkeiten. *J Dtsch Dermatol Ges.* 2006;4:49–66.