DOI: 10.1111/jdv.12107 JEADV

ORIGINAL ARTICLE

The International Criteria for Behçet's Disease (ICBD): a collaborative study of 27 countries on the sensitivity and specificity of the new criteria

International Team for the Revision of the International Criteria for Behçet's Disease (ITR-ICBD)
F. Davatchi,[1,*,†] S. Assaad-Khalil,[2,†] K.T. Calamia,[3,†] J.E. Crook,[4,†] B. Sadeghi-Abdollahi,[1,†] M. Schirmer,[5,†] T. Tzellos,[6,†] C.C. Zouboulis,[6,7,8,*,†] M. Akhlagi,[1] A. Al-Dalaan,[9] Z.S. Alekberova,[10] A.A. Ali,[11] A. Altenburg,[7,8] E. Arromdee,[12] M. Baltaci,[13] M. Bastos,[14] S. Benamour,[15] I. Ben Ghorbel,[16] A. Boyvat,[17] L. Carvalho,[18] W. Chen,[19] E. Ben-Chetrit,[20] C. Chams-Davatchi,[1] J. A. Correia,[18] J. Crespo,[21] C. Dias,[22] Y. Dong,[23] F. Paixão-Duarte,[24] K. Elmuntaser,[25] A.V. Elonakov,[10] J. Graña Gil,[26] A.-A. Haghdoost,[1] R.M. Hayani,[27] H. Houman,[16] A.R. Isayeva,[28] A.R. Jamshidi,[1] P. Kaklamanis,[29] A. Kumar,[30] A. Kyrgidis,[6] W. Madanat,[31] A. Nadji,[1] K. Namba,[32] S. Ohno,[32] I. Olivieri,[33] J. Vaz Patto,[34] N. Pipitone,[35] M.V. de Queiroz,[36] F. Ramos,[36] C. Resende,[37] C.M. Rosa,[36] C. Salvarani,[35] M.J. Serra,[38] F. Shahram,[1] H. Shams,[1] K.E. Sharquie,[27] M. Sliti-Khanfir,[16] T. Tribolet de Abreu,[39] C. Vasconcelos,[18] J. Vedes,[40] B. Wechsler,[41] Y.K. Cheng,[42] Z. Zhang,[43] N. Ziaei[1]

[1]Rheumatology Research Center, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran
[2]Department of Internal Medicine, Alexandria Faculty of Medicine, Alexandria University, Alexandria, Egypt
[3]Department of Internal Medicine,
[4]Division of Health Sciences Research, Mayo Clinic College of Medicine, Jacksonville, FL, USA
[5]Department of Internal Medicine I, Innsbruck Medical University, Innsbruck, Austria
[6]Division of Evidence Based Dermatology, Dessau Medical Center, Dessau, Germany
[7]Departments of Dermatology, Venereology, Allergology and Immunology, Dessau Medical Center, Dessau, Germany
[8]German Registry of Adamantiades-Behçet's Disease, Dessau, Germany
[9]Department of Internal Medicine, King Faisal Specialist Hospital and Research Centre, College of Medicine, King Saud University, Riyadh, Saudi Arabia
[10]Department of Systemic Rheumatic Diseases, State Institute of Rheumatology, Moscow, Russia
[11]Department of Internal Medicine, Aga Khan University Hospital, Karachi, Pakistan
[12]Department of Internal Medicine, Division of Rheumatology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand
[13]Department of Dermatology and Venereology, Innsbruck Medical University, Innsbruck, Austria
[14]Department of Internal Medicine, Hospital São Teotónio, Viseu, Portugal
[15]Department of Internal Medicine, CHU Ibn Rochd, Casablanca, Morocco
[16]Department of Internal Medicine, La Rabba Hospital, Tunis, Tunisia
[17]Department of Dermatology, School of Medicine, Ankara University, Ankara, Turkey
[18]Department of Internal Medicine, Hospital Geral de Santo António, Porto, Portugal
[19]Department of Dermatology, Chang Gung Memorial Hospital Kaohsiung, Chang Gung University, Kaohsiung, Taiwan
[20]Department of Internal Medicine, Hadassah-Hebrew University Medical Centre, Jerusalem, Israel
[21]Department of Internal Medicine, Coimbra University Hospital, Coimbra, Portugal
[22]Department of Internal Medicine, Hospital de S. João, Porto, Portugal
[23]Department of Rheumatology and Immunology, Peking Union Medical College Hospital, Beijing, China
[24]Department of Medicine II, Fernando da Fonseca Hospital, Lisbon, Portugal
[25]Kadisia Clinic, Tripoli, Libya
[26]Rheumatology Unit, Hospital Juan Canalejo, La Coruña, Spain
[27]Department of Dermatology and Venereology, College of Medicine, Medical City Teaching Hospital, University of Baghdad, Baghdad, Iraq
[28]Department of Therapeutic Dentistry, Azerbaijan Medical University, Baku, Azerbaijan
[29]Department of Rheumatology, Medical Center, Athens, Greece
[30]Department of Internal Medicine, Clinical Immunology and Rheumatology Service, All India Institute of Medical Sciences, New Delhi, India
[31]Department of Internal Medicine, Jordan Hospital, Amman, Jordan
[32]Department of Ophthalmology and Visual Sciences, Hokkaido University Graduate School of Medicine, Sapporo, Japan
[33]Rheumatology Department of Lucania, San Carlo Hospital, Potenza and Madonna delle Grazie Hospital, Matera, Italy
[34]Portuguese Institute of Rheumatology, Lisbon, Portugal

[35]Rheumatology Unit, Hospital Santa Maria Nuova, Reggio Emilia, Italy
[36]Department of Rheumatology and Metabolic Bone Diseases, Hospital Santa Maria, Lisbon, Portugal
[37]Rheumatology Research Unit, Institute of Molecular Medicine, Medical Faculty, University of Lisbon, Lisbon, Portugal
[38]Department of Internal Medicine, Santo António dos Capuchos Hospital, Lisbon, Portugal
[39]Department of Internal Medicine, Hospital Espőrito Santo, Évora, Portugal
[40]Department of Internal Medicine, Hospital Sousa Martins, Guarda, Portugal
[41]Internal Medicine Department, Pitie Salpetriere Hospital, Paris, France
[42]Allergy, Arthritis and Rheumatism Clinic, Gleneagles Medical Centre, Singapore, Singapore
[43]Rheumatology and Clinical Immunology Department, Peking University First Hospital, Peking, China
*Correspondence: F. Davatchi, C.C. Zouboulis. E-mails: fddh@davatchi.net; christos.zouboulis@klinikum-dessau.de.

Abstract

Objective Behçet's disease (BD) is a chronic, relapsing, inflammatory vascular disease with no pathognomonic test. Low sensitivity of the currently applied International Study Group (ISG) clinical diagnostic criteria led to their reassessment.

Methods An International Team for the Revision of the International Criteria for BD (from 27 countries) submitted data from 2556 clinically diagnosed BD patients and 1163 controls with BD-mimicking diseases or presenting at least one major BD sign. These were randomly divided into training and validation sets. Logistic regression, 'leave-one-country-out' cross-validation and clinical judgement were employed to develop new International Criteria for BD (ICBD) with the training data. Existing and new criteria were tested for their performance in the validation set.

Results For the ICBD, ocular lesions, oral aphthosis and genital aphthosis are each assigned 2 points, while skin lesions, central nervous system involvement and vascular manifestations 1 point each. The pathergy test, when used, was assigned 1 point. A patient scoring ≥ 4 points is classified as having BD. In the training set, 93.9% sensitivity and 92.1% specificity were assessed compared with 81.2% sensitivity and 95.9% specificity for the ISG criteria. In the validation set, ICBD demonstrated an unbiased estimate of sensitivity of 94.8% (95% CI: 93.4–95.9%), considerably higher than that of the ISG criteria (85.0%). Specificity (90.5%, 95% CI: 87.9–92.8%) was lower than that of the ISG-criteria (96.0%), yet still reasonably high. For countries with at least 90%-of-cases and controls having a pathergy test, adding 1 point for pathergy test increased the estimate of sensitivity from 95.5% to 98.5%, while barely reducing specificity from 92.1% to 91.6%.

Conclusion The new proposed criteria derived from multinational data exhibits much improved sensitivity over the ISG criteria while maintaining reasonable specificity. It is proposed that the ICBD criteria to be adopted both as a guide for diagnosis and classification of BD.

Received: 14 December 2012; Accepted: 15 January 2013

Conflict of interest

None declared.

Funding sources

No funding source.

Introduction

Behçet's disease (BD) is classified among inflammatory vascular diseases, affecting vessels of all kinds and sizes.[1] The aetiopathogenesis of the disease remains unknown.[2] Every tissue and organ of the body can practically be affected. It is particularly prevalent in the 'Silk Route' populations; however, it has a global distribution.[3] It seems to be strongly dependent on the geographical area of BD patients' residence, thus indicating the implication of environmental factors.

Although the disease rates and the clinical expression vary to some extent by ethnic origin, recurrent mucocutaneous lesions, skin lesions, ocular findings and reactivity of the skin to needle prick or injection (pathergy test) constitute common clinical hallmarks of BD.[4–6] As there is a lack of a universally recognized pathognomonic test, BD diagnosis is primarily based on clinical criteria. Oral aphthosis (OA), genital aphthosis (GA), cutaneous lesions (such as inflammatory papulopust-

[†]These authors contributed equally to this study as first authors.

ular lesions, erythema nodosum and skin ulcers) and positive pathergy (PP) reaction as well as uveitis/retinitis/hypopyon-iritis have been used in several sets of diagnostic criteria. Interestingly, though, rates of PP reaction not only vary widely in different populations but also there are indications that its sensitivity is declining over time.[7]

Taking into account the above mentioned evidence, it is not strange that perhaps no other disease has given rise to such a large number of diagnostic or classification criteria as BD. The International Study Group (ISG) criteria for BD were developed as a collaboration of scientists from seven countries to bring international agreement on one set of diagnostic criteria.[8] However, when subsequently evaluated in individual countries, it was repeatedly found to have low sensitivity relative to other criteria that have been proposed.[9] ISG criteria also did not allow for variations in the symptoms of BD, since OA was considered an obligatory manifestation for BD diagnosis.[9]

As a result of the weaknesses of the ISG criteria, an International Team for the Revision of the International Criteria for BD (ITR-ICBD) was formed under the auspices of the Epidemiology Research Group of the International Society for Behçet's disease. The aim of this team was to re-assess the sensitivity and specificity of the existing criteria, including ISG, on a large cohort of patients from 27 countries. If the properties of the ISG criteria were found to be unacceptable, then the team aimed to create a new set of criteria, with the goal of identifying a scheme that has good discriminatory properties regardless of country and that would be intuitive and easy to use in a wide variety of settings. In particular, the new set of criteria would be developed as a joint effort of an international team of BD experts thus providing a greater level of consistency among BD experts. Furthermore, the gathering together of the expert knowledge in a systematic manner would offer a tool that would be useful for mass screening and identification of possible BD by non-experts, thus maximizing clinical utility in different clinical settings.

Materials and methods

This was a prospective, international, multi-centre diagnostic accuracy study. Scientists from 32 countries joined the ITR-ICBD and those of 27 (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) sent data on BD patients and controls. This study conducted in accordance with the Declaration of Helsinki and was performed according to national laws and/or approved by the relevant authorities and the Human Ethics Review Committee of each country prior to patient enrolment. Recruitment of BD patients and controls was based on presenting symptoms and was performed in a consecutive manner. As there is no pathognomonic test for the diagnosis for BD, diagnosis was made on clinical grounds based on presenting symptoms by an expert in each relevant centre. Experts performing the diagnosis were specifically advised not to use any previous diagnostic algorithm for the final diagnosis. Control patients were selected from patients with other final diagnoses mimicking BD, or patients having at least one major sign of BD (OA, GA, papulopustular skin lesions, erythema nodosum, superficial thrombophlebitis, uveitis/retinitis). Table 1 indicates the symptoms and demographic data that were considered and recorded for every patient. Oral aphthous lesions and genital aphthous lesions and no other kind of oral and genital ulcers were considered BD signs. Regarding skin lesions, pseudofolliculitis (BD pustulosis), erythema nodosum and characteristic skin ulcers (skin aphthosis) were included. Special attention was given to the differentiation of acne lesions from pseudofolliculitis, since acne lesions can be seen in BD patients, but are not characteristic of BD. Ocular lesions were evaluated by an ophthalmologist. Anterior and posterior uveitis was evaluated with fundus examination and/or angiography. Neurological manifestations were evaluated by neurologists with the use of CT, MRI, clinical examination and liquor examination when appropriate. Pathergy test was performed with at least three skin punctures. A papular reaction ≥ 2 mm of diameter surrounded by erythema, or a pustule reaction after 24–48 h was considered a positive reaction. Haemorrhage due to needle trauma was not taken as a positive reaction. Complete data were required on all patients except for pathergy testing, HLA typing and family history. Symptoms and data before the initiation of treatment were recorded. Presence or absence of symptoms was evaluated by a team comprising dermatologists, ophthalmologists, rheumatologists and neurologists, as appropriate. The final decision regarding the relevance of symptoms and signs presented in Table 1 and recorded for any given patient was based on clinical evaluation in all centres participating in this study. The assessment of the presence or absence of symptoms in each centre was performed prior to the final diagnosis made by BD experts.

Data entry software was developed specifically for this study to ensure the quality of the data. The full data set was divided randomly into two groups, stratified by country and case/control status. These were designated to the training and validation sets. The training set was to be used for initial comparison of the ISG criteria with other criteria as well as for the development of the revised criterion; once the form of the new criteria had been decided upon, it was to be compared with the other criteria using the validation set.

All patients' characteristics were summarized in the training set cases and controls, both overall and separately by country so as to gain a sense of the country-to-country differences and to learn which variables were the most useful discriminators of cases and controls from an empirical perspective. To ensure that those variables that had clear potential for discriminant

Table 1 Patient characteristics in training data set

Characteristic	Behçet's disease (*N* = 1278)		Controls (*N* = 582)		
	N	% sensitivity	*N*	%	Specificity (%)
Gender (male)	730	57	267	46	54
Oral aphthosis*	1248	98	378	65	35
Genital aphthosis*	950	74	35	6	94
Skin manifestations*, †	899	70	76	13	87
Pseudofolliculitis*	674	53	48	8	92
Erythema nodosum*	406	32	27	5	95
Skin aphthosis*	56	4	10	2	98
Ocular lesions*, †	703	55	139	24	76
Anterior uveitis*	508	40	120	21	79
Posterior uveitis*	482	38	76	13	87
Retinal vasculitis*	288	23	21	4	96
Joint manifestations*, †	652	51	213	37	63
Arthralgia	443	35	161	28	72
Arthritis*	289	23	76	13	87
Ankylosing spondylitis	27	2	11	2	98
Neurological manifestations*, †	212	17	16	3	97
Peripheral*	63	5	8	1	99
Central*	155	12	9	2	98
Vascular manifestations*, †	242	19	36	6	94
Arterial thrombosis	40	3	5	1	99
Large vein thrombosis*	86	7	12	2	98
Phlebitis*	111	9	20	3	97
Superficial phlebitis*	70	5	7	1	99
Gastrointestinal*, †	80	6	17	3	97
Chronic diarrhoea	46	4	14	2	98
Proctorrhagia	25	2	5	1	99
Epididymitis*	91	7	9	2	98
Cardiac manifestations	24	2	3	1	99
Pleuro-pulmonary	31	2	15	3	97
Pathergy test done	891	70	413	71	–
Pathergy test positive‡	419	47	33	8	92
HLA B51 test done	514	40	286	49	–
HLA B51 test positive‡	260	51	84	29	71
Family history – asked	951	74	432	74	–
Family history positive‡	105	11	26	6	94

*These symptoms showed evidence of association with disease status with $P < 0.01$ and all except sex were considered for potential inclusion in the new classification scheme.

†Presence of any of the subsequent indented characteristics defines the presence of this more encompassing characteristic.

‡Percent is calculated among those in whom test is done.

utility would only be included in the final model, inclusion of those variables that had a $P < 0.01$ in a test of association with disease status (logistic regression likelihood ratio test adjusting for gender and country) were only considered. Those variables that were not routinely collected on all patients were not considered for inclusion in the development of the new scheme for two main reasons: (i) difficulty in assessing the reasons for missing data to be sure that it would be dealt with adequately, but more importantly and (ii) the desire for a scoring system that is readily available to all and that does not depend on information that may not be available. In particular, these considerations applied to pathergy testing, the HLA-B51 test and family history.

A modified version of stepwise logistic regression with forward selection was used as the primary tool for creating a scoring scheme, with modifications included primarily to limit the influence of data from any one country so as to maximize the general applicability of the final scheme. One modification was to weight cases and controls in the analysis so as to allow each country to contribute no more than 10% of the information for cases and no more than 10% of the information for controls. Rather than using a fixed *P*-value or similar criterion to decide at which point to stop adding variables into the model, a leave-one-country-out-at-a-time cross-validation approach along with inspection of receiver operator characteristic (ROC) curves of estimated sensitivity vs. specificity was used to assess the utility of additional variables. Interaction terms were considered as part of this process. A number of proposals with simplified scoring and similar properties were identified as a result of this process; these were discussed among the first authors to come to a decision on the preferred scheme taking into account the a priori defined goals and comparisons with properties of the existing criteria in the training data set. The agreed upon scheme was then finally assessed by estimating sensitivity and specificity in the validation set in comparison with the previous ISG criteria and other published criteria.

As an additional step, we examined to what extent the properties of the chosen scheme might improve by modification to allow for the addition of pathergy test results when known.

Results

Participants' demographic data

A total of 2556 patients with BD and 1163 controls were recruited between January 1, 2005 and June 1, 2006. The clinical diagnosis of BD or other clinical mimickers and the conduct of the symptom evaluations was performed prior to treatment administration. Patients were divided randomly into two series, the training data and the validation data (Table 2). The training set consisted of 1278 cases and 582 controls and characteristics of patients in this group are presented in Table 1. The spectrum

Table 2 Numbers of cases and controls in training and validation sets by country

Country	Training set (n = 1860)		Validation set (n = 1859)		Total (n = 3719)
	Behçet's disease (n = 1278)	Controls (n = 582)	Behçet's disease (n = 1278)	Controls (n = 581)	
Iran	146	177	147	177	647
Portugal	120	71	121	71	383
Turkey	117	18	117	17	269
Egypt	102	17	101	17	237
Greece	57	44	58	43	202
PR China	56	41	57	41	195
Japan	95	0	96	0	191
USA	86	4	85	4	179
Tunisia	52	28	51	29	160
Italy	52	28	53	27	160
Morocco	44	27	44	27	142
Germany	43	19	43	19	124
Saudi Arabia	35	21	36	20	112
Russia	24	23	24	24	95
Spain	43	4	42	4	93
Singapore	26	20	26	20	92
Jordan	20	20	20	21	81
India	34	4	33	5	76
Iraq	32	0	32	0	64
Libya	26	5	26	4	61
Thailand	15	8	15	7	45
France	16	0	16	0	32
Azaerbaijan	15	0	15	0	30
Austria	7	3	6	3	19
Taiwan	8	0	8	1	17
Israel	6	0	6	0	12
Pakistan	1	0	0	0	1

Table 3 Most common diagnoses of the control patients

Bullous autoimmune diseases
• Bullous pemphigoid
• Cicatricial mucous membrane pemphigoid
Stevens–Johnson syndrome (±recurrent ulcerative stomatitis)
Recurrent genital herpes infection
Recurrent oral apthosis (±arthralgia)*
Panuveitis (±arthralgia)*
Anterior uveitis (±arthralgia)*
Retinal vasculitis*
Herpetic keratouveitis
HLA-B27 positivity
• Uveitis (±ankylosing spondylitis)
• Recurrent iritis

*Other origin than Behçet's Disease.

of presenting symptoms across countries, along with the rates of conduct of pathergy testing, HLA B51 testing and availability of family history information are shown in supplementary files. A list of the most common diagnoses for control patients is presented in Table 3.

Performance of existing criteria in the training set

Performance of the existing criteria[8,10–22] was examined in the training set (Table 4). ISG criteria yielded an estimated sensitivity of 81% and a specificity of 96%, ranking 13th for sensitivity and second for specificity. Highest sensitivity was obtained with Curth criteria (98%), followed by Cheng and Zhang (97%) although both had specificity lower than 90%. The Iran Classification Tree had good sensitivity (94%) along with reasonable specificity (91%) although pathergy testing was included in this criteria. Highest specificity was obtained with Hubault and Hamza criteria (97%) followed jointly by the ISG and Hewitt revised (96%). The least specific were the Curth criteria (84%) and the Cheng and Zhang criteria (87%). Tested against large series of patients and controls gathered from the 27 countries in the training set, the ISG criteria was considered to have failed, yielding unacceptably low sensitivity,

Table 4 Sensitivity and specificity of various criteria in training and validation sets

Criteria*	Training				Validation			
	Sensitivity (*N* = 1278)		Specificity (*N* = 582)		Sensitivity (*N* = 1278)		Specificity (*N* = 581)	
	n	95% CI †	*n*	95% CI	*n*	95% CI	*n*	95% CI
Curth[10]	1255	98% (97–99)	486	84% (80–86)	1265	99% (98–99)	475	82% (78–85)
Mason/Barnes[11]	1046	82% (80–84)	554	95% (93–97)	1046	82% (80–84)	554	95% (93–97)
Hewitt revised[12]	755	59% (56–62)	558	96% (94–97)	731	57% (54–60)	555	96% (94–97)
Japan (original)[13]	1089	85% (83–87)	539	93% (90–95)	1125	88% (86–90)	536	92% (90–94)
Hubault and Hamza[14]	701	55% (52–58)	566	97% (96–98)	741	58% (55–61)	562	97% (95–98)
O'Duffy[15]	1115	87% (85–89)	534	92% (89–94)	1123	88% (86–90)	523	90% (87–92)
Cheng and Zhang[16]	1232	97% (95–97)	505	87% (84–89)	1249	98% (97–98)	484	83% (80–86)
Dilsen (original)[17]	1094	86% (84–87)	527	91% (88–93)	1130	88% (87–90)	527	91% (88–93)
Japan (revised)[18]	1125	88% (86–90)	533	92% (89–93)	1160	91% (89–92)	527	91% (88–93)
ISG[8]	1038	81% (79–83)	558	96% (94–97)	1086	85% (83–87)	558	96% (94–97)
Iran traditional[19]	1119	88% (86–89)	537	92% (90–94)	1149	90% (88–92)	536	92% (90–94)
Iran Classification Tree[20]	1199	94% (92–95)	528	91% (88–93)	1223	96% (94–97)	522	90% (87–92)
Dilsen (revised)[21]	1057	83% (81–85)	556	96% (94–97)	1106	87% (85–88)	557	96% (94–97)
Korea[22]	1139	89% (87–91)	542	93% (91–95)	1179	92% (91–94)	536	92% (90–94)
ICBD‡	1200	94%	536	92%	1211	95% (93–96)	526	91 (88–93)

*For those criteria that use the pathergy test result this was assumed negative for the patients who did not have the test for the purposes of creating this table.

†CI: confidence interval.

‡ICBD: International Criteria for Behçet's Disease.

thus providing the evidence of need for an improved international criteria.

Development of New International Criteria for Behcet's Disease (ICBD)

The proportions of patients with each symptom were calculated in the training set cases and controls (Table 1), thus providing estimates of sensitivity and specificity for each symptom separately. Of note, no symptom alone is adequate to classify BD patients. Table 1 indicates the symptoms that showed sufficient evidence of association with BD ($P < 0.01$) to be considered for potential inclusion in the new classification scheme. The symptoms showing greatest discriminatory utility individually, in the sense of having the highest average of sensitivity and specificity, were genital aphthosis (84%), skin manifestations (79%), pseudofolliculitis (72%), oral aphthosis (66%) and ocular lesions (66%).

The final proposed criteria based on the training set data and agreed upon by the research team included oral aphthosis, genital aphthosis, ocular lesions (anterior uveitis, posterior uveitis, or retinal vasculitis), neurological manifestations, skin lesions (pseudofolliculitis, skin aphthosis, erythema nodosum) and vascular manifestations (arterial thrombosis, large vein thrombosis, phlebitis or superficial phlebitis). Oral aphthosis, genital aphthosis and ocular lesions were each given 2 points, whereas 1 point was assigned to each of skin lesions, vascular manifestations and neurological manifestations. A patient scoring 4 points or above was classified as having BD (Table 5). With the training data set, this scheme exhibited an estimated 93.9% sensitivity and 92.1% specificity. When estimates were created that limited the contribution of any one country to no more than 10%, adjusted estimates were 93.9% for sensitivity and 89.6% for specificity.

Performance of existing criteria in validation data set

As the new scoring scheme, the ICBD, was developed using the training data, it could only be compared objectively with the other criteria using the validation data set (Table 4, Fig. 1). Japan revised, Iran traditional and Iran Classification Tree, Korea and the newly developed ICBD were the five existing criteria that achieved estimates of both sensitivity and specificity that were higher than 90%. ICBD demonstrated an estimated sensitivity of 94.8% (95% CI: 93.4–95.9%), considerably higher than that of the original ISG criteria (85.0%); the specificity was lower than that of the original criteria at 90.5% (95% CI: 87.9–92.8%) compared to 96.0%, yet still acceptably high. Distribution of scores in cases and controls in both validation and training data sets along with a proposed plausibility scale of BD diagnosis is presented in Table 6.

Inclusion of pathergy testing

Pathergy testing was not performed in all patients and the proportions of patients in whom tests were done varied from

country to country with results not available for approximately 30% of patients overall. Also, within countries the proportion tested was sometimes quite different for cases and controls suggesting selectivity over use of the test and/or use of pathergy test result in diagnosis (see supplementary files). For these reasons, pathergy testing was not included in the primary development of the scoring system. However, empirical evidence suggests that building in pathergy testing into the scheme may improve classification performance, been, indeed, included in many previous classification schemes for BD. We thus examined in a more empirical way to what extent the properties of the proposed simplified scoring system may improve with the addition of pathergy test results. We restricted our assessment to countries where at least 90% of cases and at least 90% of controls had a pathergy test and imputed pathergy test results for those individuals without the test in those countries according to the distribution of pathergy test results among those of the same ICBD score. Among these countries, adding 1 point for a positive pathergy test result increased the estimate of sensitivity from 95.5% to 98.5% while barely reducing specificity, from 92.1% to 91.6%. Thus, although the main scoring system remained unaltered, it was decided by the team that it was appropriate to add an additional point for a positive pathergy result when available.

Table 5 International Criteria for Behçet's Disease – point score system: scoring ≥ 4 indicates Behçet's diagnosis

Sign/symptom	Points
Ocular lesions	2
Genital aphthosis	2
Oral aphthosis	2
Skin lesions	1
Neurological manifestations	1
Vascular manifestations	1
Positive pathergy test*	1*

*Pathergy test is optional and the primary scoring system does not include pathergy testing. However, where pathergy testing is conducted one extra point may be assigned for a positive result.

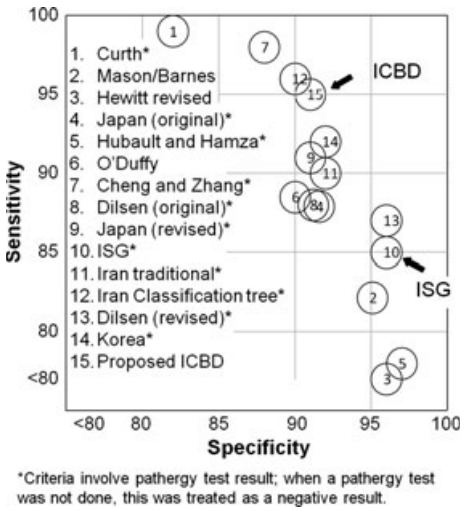

Figure 1 Sensitivity and specificity of existing and ICBD criteria in the validation data set.

Discussion

ICBD criteria, derived from an evidence-based protocol, showed good discriminatory properties with data contributed from 27 countries and brings an international agreement on a new set of criteria. The gathering together of expert knowledge in this systematic manner is anticipated to provide a greater level of consistency among BD experts as well as non-experts, in the future. Overall, this multinational diagnostic accuracy study demonstrated that the new proposed ICBD criteria along with the point score system showed a good balance of sensitivity and specificity.

Furthermore, the ICBD criteria include a wide variety of symptoms such as vascular manifestations, skin lesions and neurological manifestations and permit the early diagnosis of special cases as well as facilitate their potential use as a mass screening tool for early referral to expert centres. On the other hand, it also allows a uniform approach for identifying BD patients for inclusion in multi-centred international randomized controlled clinical trials.

A main advantage of the ICBD criteria is the fact that they can be used with or without the conduction of pathergy testing. Pathergy testing is not routinely performed in all countries and the

Table 6 Distribution of scores in cases and controls (testing and validation set combined)

Score*	Percent of cases	Percent of controls	Plausibility of BD	Simple classification
≤ 1	<1	11	Almost certainly not BD	Not BD
2	1	72	BD very unlikely	
3	4	9	Possible, but unlikely BD	
4	14	5	Probable BD	BD
5	32	3	BD highly likely	
≥ 6	48	<1	Almost certainly BD	

*This table does not incorporate the pathergy testing results.
BD, Behçet's disease.

rate of PP varies widely in different populations.[7] In the proposed point system pathergy test is optional thus allowing early and accurate diagnosis in countries where pathergy testing is not routinely conducted and in countries where BD patients show a low rate of pathergy test positivity. On the other hand, it does allow the influence of pathergy test in the final scoring where it is done, by addition of an additional point for a positive result. Interestingly, highest pathergy positivity prevalences have been detected in the countries of Silk Route, where BD also shows highest prevalence.[7] Moreover, a recent study in Iran indicates that over the years the sensitivity of pathergy testing has declined and its specificity has increased.[7] Taking this into account, pathergy testing may be considered especially important for patients who have baseline ICBD scores near to the cut-off point (3, 4) after exclusion of other possible diagnoses at least in countries with high prevalence of BD. Promoting this optional, selective use of pathergy testing among a limited subset of patients also reduces costs and thus may be more amenable to sites that would prefer to do pathergy testing on a selective basis only.

The primary limitation of this study is that BD has no specific test and the gold standard for final diagnosis remains expert opinion. BD experts in different countries may have different expert opinion on a given patient due to different intrinsic thresholds for certainty of a diagnosis or placing more weight on some symptoms than others according to differing prevalence by site. Identification of controls shares the same limitations. These considerations could only be addressed effectively through conduct of an agreement study. Despite these caveats, our approach is an important attempt to come as close as possible to a consensus among experts regarding the diagnosis of BD.

One downside of the introduction of the new ICBD criteria is the argument that many previous studies were performed using ISG criteria and future studies using ICBD criteria will not be fully comparable to them. However, such an argument is not strong enough to justify the further use of ISG criteria with such low performance. In addition, since ICBD criteria exhibit higher sensitivity than the ISG criteria, classification with the ICBD criteria will likely include the vast majority of patients who would be classified according the ISG criteria.

Delays of diagnoses of BD have been documented.[23] The higher sensitivity of the ICBD criteria will allow for earlier recognition and referral to expert centres, thus leading to earlier diagnosis and treatment with salutary results. For all these reasons, we suggest that the new criteria with improved sensitivity be adopted so as to better serve both the clinician and the patient in correct and early diagnosis of BD.

Panel: Research in context

Systematic review

We performed a Pubmed search for all diagnostic accuracy studies regarding Behcet disease using the search terms 'Behcet', 'classification', 'diagnostic accuracy', 'diagnosis', 'sensitivity', 'specificity' and also contacted Behcet's disease experts worldwide. We applied no language or data restrictions to our search.

Interpretation

Our findings suggest that the new proposed criteria/point score which derived from multinational data exhibits much improved sensitivity over the ISG criteria, while still maintaining reasonable specificity. It provides a greater level of consistency among Behcet's disease experts, is useful for mass screening and its introduction maximizes clinical utility in different clinical settings. For these reasons, it is suggested that the new criteria should be adopted.

Authors' contribution

F. Davatchi and C.C. Zouboulis, as the corresponding authors, confirm that they had access to all the data in the study and have final responsibility for the decision to submit for publication. All authors declare that they participated in the study as declared below and that they have seen and approved the final version of the manuscript.

S. Assaad-Khalil, K.T. Calamia, F. Davatchi, M. Schirmer and C.C. Zouboulis were the leaders of the International Team for the Revision of the International Criteria for Behçet's disease, planned the study and co-analysed the final data.

F. Davatchi coordinated the data collection and primary global analysis, which was performed at the Rheumatology Research Center, Tehran University for Medical Sciences.

J. E. Crook, B. Sadeghi-Abdollahi, T. Tzellos and A. Kyrgidis analysed the global data, performed the statistical evaluation of the global data, and re-evaluated the global data set.

S. Assaad-Khalil, K.T. Calamia, J.E. Crook, F. Davatchi, M. Schirmer, T. Tzellos and C.C. Zouboulis co-authored the manuscript.

M. Schirmer, M. Baltaci (Austria), A.R. Isayeva (Azerbaijan), Y. Dong, Z. Zhang (China), S. Assaad-Khalil (Egypt), B. Wechsler (France), C.C. Zouboulis, A. Altenburg (Germany), P. Kaklamanis (Greece), A. Kumar (India), F. Davatchi, F. Shahram, A. Nadji, C. Chams-Davatchi, H. Shams, N. Ziaei, M. Akhlagi, A.R. Jamshidi, A.A. Haghdoost (Iran), K.E. Sharquie, R.M. Hayani (Iraq), E. Ben-Chetrit (Israel), I. Olivieri, C. Salvarani, N. Pipitone (Italy), S. Ohno, K. Namba (Japan), W. Madanat (Jordan), K. Elmuntaser (Libya), S. Benamour (Morocco), A.A. Ali (Pakistan), J. Crespo, T. Tribolet de Abreu, C. Vasconcelos, J. Correia, L. Carvalho, M. Bastos, M.J. Serra, C. Resende, F. Ramos, M. Rosa, V. de Queiroz, J. Vedes, C. Dias, J. Vaz Patto, F.P. Duarte (Portugal), Z.S. Alekberova, A.V. Elonakov (Russia), A. Al Dalaan (Saudi Arabia), Y.K. Cheng (Singapore), J. Grana Gil (Spain), W. Chen (Taiwan), A. Emvalee (Thailand), H. Houman, I. Ben Ghorbel, M. Sliti-Khanfir (Tunisia), A. Boyvat (Turkey), K.T. Calamia (USA) collected and evaluated the data in the respected countries.

M. Baltaci, A.R. Isayeva, Y. Dong, Z. Zhang, B. Wechsler, A. Altenburg, P. Kaklamanis, A. Kumar, F. Shahram, A. Nadji, C. Chams-Davatchi, H. Shams, N. Ziaei, M. Akhlagi, A.R. Jamshidi, A.A. Haghdoost, K.E. Sharquie, R.M. Hayani, E. Ben-Chetrit, I. Olivieri, C. Salvarani, N. Pipitone, S. Ohno, K. Namba, W. Madanat, K. Elmuntaser, S. Benamour, A.A. Ali, J. Crespo, T. Tribolet de Abreu, C. Vasconcelos, J. Correia, L. Carvalho, M. Bastos, M.J. Serra, C. Resende, F. Ramos, M. Rosa, V. de Queiroz, J. Vedes, C. Dias, J. Vaz Patto, F.P. Duarte, Z.S. Alekberova, A.V. Elonakov, A. Al Dalaan, Y.K. Cheng, J. Grana Gil, W. Chen, A. Emvalee, H. Houman, I. Ben Ghorbel, M. Sliti-Khanfir, A. Boyvat, K.T. Calamia, J.E. Crook, T. Tzellos and A. Kyrgidis made suggestions to the final manuscript.

References

1 Kalayciyan A, Zouboulis CC. An update on Behçet's disease. *J Eur Acad Dermatol Venereol* 2007; **21**: 1–10.
2 Sakane T, Takeno M, Suzuki N, Inaba G. Behçet's disease. *N Engl J Med* 1999; **341**: 1284–1291.
3 Zouboulis CC. Epidemiology of Adamantiades-Behçet's disease. *Ann Med Interne (Paris)* 1999; **150**: 488–498.
4 Krause I, Yankevich A, Fraser A *et al.* Prevalence and clinical aspects of Behcet's disease in the north of Israel. *Clin Rheumatol* 2007; **26**: 555–560.
5 Davatchi F, Shahram F, Chams-Davatchi C *et al.* Behcet's disease in Iran: analysis of 6500 cases. *Int J Rheum Dis* 2010; **13**: 367–373.
6 Zouboulis ChC, Kötter I, Djawari D *et al.* Current epidemiological data from the German Registry of Adamantiades-Behçet's disease. *Adv Exp Med Biol* 2003; **528**: 43–48.
7 Davatchi F, Chams-Davatchi C, Ghodsi Z *et al.* Diagnostic value of pathergy test in Behcet's disease according to the change of incidence over the time. *Clin Rheumatol* 2011; **30**: 1151–1155.
8 International Study Group for Behçet's Disease. Criteria for diagnosis of Behçet's disease. *Lancet* 1990; **335**: 1078–1080.
9 Lee S. Diagnostic criteria of Behçet's disease: problems and suggestions. *Yonsei Med J* 1997; **38**: 365–369.
10 Curth HO. Recurrent genito-oral aphthosis and uveitis with hypopyon (Behçet's syndrome). A report of 2 cases. *Arch Dermatol Syph (Chicago)* 1946; **54**: 179–196.
11 Mason RM, Barnes CG. Behcet's syndrome with arthritis. *Ann Rheum Dis* 1969; **28**: 95–103.
12 Hewitt J, Escande JP, Maness S. Revision des critères diagnostiques du syndrome de Behcet. *Presse Med* 1971; **79**: 901.
13 Behçet's Disease Research Committee of Japan. Behçet's disease guide to diagnosis of Behcet's disease (1972). *Jpn J Ophthalmol* 1974; **18**: 291–294.
14 Hubault A, Hamza M. *La maladie de Behçet en 1974*. In de Seze S *et al.* eds. L'actualité Rhumatologique Paris. Expension Scientifique, Paris 1974; 15: 43–55.
15 O'Duffy JD. Critères proposés pour le diagnostique de la maladie de Behçet et notes therapeutiques. *Rev Med* 1974; **36**: 2371–2379.
16 Cheng SP, Zhang XQ. Some special clinical manifestations of Behcet's disease and review of literature (in Chinese). *Chin J Int Med* 1980; **19**: 15–22.
17 Dilsen N, Konice M, Aral O. Our Diagnostic Criteria of Behçet's Disease – An Overview. In Lehner T, Barnes CG eds. Recent Advances in Behçet's Disease. London Royal Society of Medicine Services. International Congress and Symposium Series 103, London Royal Society of Medicine Services, London, 1986: 177–180.
18 Mizushima Y. Recent research into Behçet's disease in Japan. *Int J Tiss Reac* 1988; **10**: 59–65.
19 Davatchi F, Shahram F, Akbarian M *et al.* Accuracy of Existing Diagnosis Criteria for Behçet's Disease. In Wechsler B, Godeau P eds. Behçet's Disease. Excerpta Medica International Congress Series 1037, Excerpta Medica, Amsterdam, 1993: 225–228.
20 Davatchi F, Shahram F, Akbarian M *et al.* Classification Tree for the Diagnosis of Behcet's Disease. In Wechsler B, Godeau P eds. Behcet's Disease. Excerpta Medica International Congress Series 1037, Excerpta Medica, Amsterdam, 1993: 245–248.
21 Dilsen N. About Diagnostic Criteria for Behçet's Disease: Our New Proposal. In Bang D, Lee ES, Lee S eds. Behçet's Disease. Design Mecca Publishing Co., Seoul, Korea, 2003: 101–104.
22 Chang HK, Lee SS, Bai HJ *et al.* Validation of the classification criteria commonly used in Korea and a modified set of preliminary criteria for Behçet's disease: a multi-center study. *Clin Exp Rheumatol* 2004; **22** (Suppl. 34): S21–S26.
23 Alpsoy E, Donmez L, Bacanli A, Apaydin C, Butun B. Review of the chronology of clinical manifestations in 60 patients with Behçet's disease. *Dermatology* 2003; **207**: 354–356.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table A1. Frequency of male gender by country (in training set).
Table A2. Frequency of oral aphthosis (OA) by country (in training set).
Table A3. Frequency of genital aphthosis (GA) by country (in training set).
Table A4. Frequency of skin manifestations (SKI2) by country (in training set).
Table A5. Frequency of pseudo folliculitis (PF) by country (in training set).
Table A6. Frequency of erythema nodosum (EN) by country (in training set).
Table A7. Frequency of skin aphthosis (SA) by country (in training set).
Table A8. Frequency of ocular lesions (OL) by country (in training set).
Table A9. Frequency of anterior uveitis (AU) by country (in training set).
Table A10. Frequency of posterior uveitis (PU) by country (in training set).
Table A11. Frequency of retinal vasculitis (RV) by country (in training set).
Table A12. Frequency of joint manifestations (JNT) by country (in training set).
Table A13. Frequency of arthralgia (ARG) by country (in training set).
Table A14. Frequency of arthritis (ARTH) by country (in training set).
Table A15. Frequency of ankylosing spondylitis (AS) by country (in training set).
Table A16. Frequency of neurological manifestations (NEUR) by country (in training set).
Table A17. Frequency of peripheral (PER) by country (in training set).

Table A18. Frequency of central (CNS) by country (in training set).
Table A19. Frequency of vascular manifestations (VAS) by country (in training set).
Table A20. Frequency of arterial thrombosis (ATHR) by country (in training set).
Table A21. Frequency of large vein thrombosis (LVT) by country (in training set).
Table A22. Frequency of phlebitis (PHL) by country (in training set).
Table A23. Frequency of superficial phlebitis (SP) by country (in training set).
Table A24. Frequency of gastro-intestinal (GI) by country (in training set).
Table A25. Frequency of chronic diarrhea (DIAR) by country (in training set).
Table A26. Frequency of proctorrhagia (PRO) by country (in training set).
Table A27. Frequency of epididymitis (EPID) by country (in training set).
Table A28. Frequency of cardiac manifestations (CARD) by country (in training set).
Table A29. Frequency of pleuro-pulmonary (PLPU) by country (in training set).
Table A30. Frequency of pathergy testing (PATH?) by country (in training set).
Table A31. Frequency of pathergy positive test (PATH) by country (in training set).
Table A32. Frequency of B51 testing (B51?) by country (in training set).
Table A33. Frequency of B51 positive test (B51) by country (in training set).
Table A34. Frequency of asking about family history (FH?) by country (in training set).
Table A35. Frequency of positive family history (FH) by country (in training set).