Diagnosis/Classification Criteria for Behcet’s Disease

Fereydoun Davatchi

Behcet’s Unit, Rheumatology Research Center, Shariati Hospital, Tehran University of Medical Sciences, Tehran 14117, Iran

Correspondence should be addressed to Fereydoun Davatchi, fddh@davatchi.net

Received 10 June 2011; Accepted 23 June 2011

Academic Editor: Ümit Tursen

Copyright © 2012 Fereydoun Davatchi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Historical Background. The ISG criteria for Behcet’s, created in 1990, have excellent specificity, but lack sensitivity. The International Criteria for Behcet’s Disease (ICBD) was created in 2006, as replacement to ISG. The aim of this study was to compare their performance.

ISG and ICBD Criteria. For ISG oral aphthosis is mandatory. The presence of any two of the following (genital aphthosis, skin lesions, eye lesions, and positive pathergy test) will diagnose/classify the patient as BD. For ICBD, vascular lesions were added, while oral aphthosis is no more mandatory. Getting 3 or more points diagnose/classify the patient as BD (genital aphthosis 2 points, eye lesions 2 points, and the remaining each one point).

Performance and Comparison of ISG and ICBD. Their sensitivity, specificity, and accuracy (percent agreement), were tested in three independent cohort of patients from Far-East (China), Middle-East (Iran), and Europe (Germany). The sensitivity for ISG was respectively 65.4%, 78.1%, 83.7% and for ICBD 87%, 98.2%, and 96.5%. The specificity for ISG was 99.2%, 98.8%, 89.5% and for ICBD 94.1%, 95.6%, and 73.7%. The accuracy for ISG was 74.2%, 85.5%, 85.5% and for ICBD 88.9%, 97.3%, and 89.5%. Conclusion. ICBD has better sensitivity, and accuracy than ISG.

1. Historical Background

Although Behcet’s Disease (BD) is relatively a young disease (described in 1937), it has already 16 sets of diagnosis/classification criteria. The first of them was proposed by Curth in 1946, less than 10 years after the description of the disease [1]. It was followed by Hewitt et al. in 1969 [2], Mason and Barnes in 1969 [3], Hewitt et al. revised in 1971 [4], Japan in 1972 [5], Hubault and Hamza in 1974 [6], O’Duffy in 1974 [7], Chen in 1980 [8], Dilsen et al. in 1986 [9], Japan revised in 1988 [10], International Study Group (ISG) in 1990 [11], Iran in 1993 [12], Classification Tree in 1993 [13], Dilsen revised in 2000 [14], Korea in 2003 [15, 16], and the International Criteria for Behcet’s Disease (ICBD) in 2006 [17–19].

The ISG criteria were created in 1990 to bring a consensus on one set of criteria by the collaboration of France, Iran, Japan, Tunisia, Turkey, UK, and USA. With the sensitivity of ISG criteria being low [12, 20–25], during the first International Workshop of Behcet’s Disease in Kuhfair (Austria), it was decided to create an international team to evaluate the performance of ISG criteria and to compare it with the existing BD criteria and revise it if necessary.

The ITR-ICBD team was founded in 2004 with the participation of 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). The International Criteria for Behcet’s Disease (ICBD) were presented to the International Conference of Behcet’s Disease in Lisbon (Portugal) in 2006. Originally it had two formats, like the Iranian criteria. Later, it was decided to keep only the traditional format [17–19]. The ICBD were presented to the 2007 World Congress of Dermatology in Argentina and to 2009 ACR congress of Rheumatology in the USA [19].

2. ISG and ICBD Criteria

The ISG criteria [11] use 5 items. Two items are mucous membrane manifestations. They are oral aphthosis (OA) and genital aphthosis (GA). The third item is skin manifestations, comprising pseudofolliculitis (PF) and erythema nodosum (EN). The forth item is ocular manifestations. They are anterior uveitis (AU), posterior uveitis (PU), and retinal
vasculitis (RV). The fifth item is the presence of pathergy phenomenon (PP). It is detected by the pathergy test [26–30]. In ISG criteria, the presence of OA is mandatory. Two other items from the 4 remaining (GA, skin, eye, PP) are necessary to classify a patient as having BD.

For the international criteria, the ICBD [17–19], vascular manifestations (VMs) have been added to the 5 items of ISG criteria, because they are one of the characteristics of BD, and were used in many criteria before the advent of ISG (Mason and Barnes, Hewitt, Hubault and Hamza, Dilsen, Japan revised, and Dilsen revised criteria). VM is defined as superficial phlebitis, deep vein thrombosis, large vein thrombosis, arterial thrombosis, and aneurysm. Therefore, ICBD use six items: OA, GA, skin (PF, EN), eye lesions (AU, PU, RV), VM, and PP. In the ICBD, genital aphthous lesions and eye lesions have more diagnostic value than the others. They get each 2 points. The other 4 items (OA, skin, VM, PP) get one point each. A patient has to get 3 or more points to be diagnosed/classified as having BD.

3. Performance and Comparison of ISG and ICBD

Many ways and methods can be used to evaluate the performance of a criteria set. The most common used are sensitivity, specificity, and accuracy. Other methods are the positive predictive value, the negative predictive value, the positive likelihood ratio, the negative likelihood ratio, the diagnostic odds ratio, and Youden’s index [35–39].

Sensitivity is the number of BD patients correctly classified (diagnosed) by the criteria. It is expressed as percentage (number of diagnosed BD patients, divided by the total number of BD patients, and then multiplied by 100) [35]. The sensitivity of ISG in their cohort of 886 patients was 92% [11]. The 95% confidence interval (95% CI) was 90% to 93.6%. The sensitivity of ICBD in their cohort of 2556 BD patients was 96.1% (95% CI 95.3–96.8). By chi-square test the difference between the two sets of criteria is statistically significant ($\chi^2 = 23.439, P < 0.001$). The sensitivity of ISG in the ICBD cohort of patients was 82.4% (95% CI 80.9–83.9).

It is important to look at the sensitivity of the two criteria in independent cohort of patients. Three studies validated the ICBD in their cohort of patients: Germany in 2008 [31], China in 2008 [32], and Iran in 2010 [33]. The sensitivity of ISG was, respectively, 83.7% (95% CI 74.3–90.1), 65.4% (95% CI 60.2–70.5), and 78.1% (95% CI 77–79.1). The sensitivity of ICBD was, respectively, 96.5% (95% CI 89.7–99.2), 87% (95% CI 82.8–90.2), and 98.2% (95% CI 97.8–98.5).

Table 1 shows the sensitivity of ISG in different cohort of patients from different parts of world [11, 12, 17–24, 31–33].

Specificity is the number of non-BD patients, correctly recognized as not having BD. It is expressed as percentage (number of non-BD patients correctly recognized as not having BD, divided by the total number of non-BD patients, then multiplied by 100) [35]. The specificity of ISG criteria in their own cohort of patients was 97% (95% CI 90.8–99.3). However, the number of control patients was only 97, and all other control patients having oral aphthosis were discarded from the original cohort of control patients [34]. The specificity of ICBD in their cohort of patients was 88.7% (95% CI 86.8–90.4). The specificity of ISG and ICBD in Germany, China, and Iran was, respectively, 89.5%, 99.2% and 98.8% (ISG), and 73.7%, 94.1%, and 95.6% (ICBD). Table 2 shows the specificity of different criteria in different studies.

Accuracy or percent agreement is the ability of the criteria to correctly recognize BD patients from the non-BD patients. It is also expressed by percentage (number of diagnosed BD patients + number of non-BD patients correctly recognized as not having BD, divided by the total number of BD patients + total number of non-BD patients, and then multiplied by 100) [35]. The accuracy of ISG in their own cohort of patients was 92% (95% CI 90.1–93.5). The accuracy of ICBD in their own cohort of patients was 93.8% (95% CI 93–94.5). The accuracy of ISG and ICBD in Germany, China, and Iran was, respectively, 85.5%, 7402% and 85.5% (ISG),
and 89.5%, 88.9%, and 97.3% (ICBD). Table 3 shows the accuracy of different criteria in different studies.

Positive predictive value (PPV) demonstrates the probability that the positive test be true positive. PPV is more influenced by specificity than sensitivity. A criteria set with 90% sensitivity and 90% specificity will have a PPV of 90. If sensitivity increases to 95, PPV will improve to 90.5%, while if specificity increases to 95%, PPV will improve to 94.8%. PPV is also greatly influenced by the prevalence of the disease. Taking the above example, the PPV remains the same (90) in a dedicated BD clinic, where 50% of patients have BD and 50% are controls (patients mimicking BD but are not true BD). In the general population, with a prevalence of 80 for 100,000 inhabitants, the PPV becomes only 0.72. Therefore the results calculated in a specific setting cannot be used in another setting [33]. The PPV was higher for ISG than ICBD criteria in the 3 independent set of patients; however, the difference was very small in the Iranian patients, only 2.8% (Table 4).

Negative predictive value (NPV) indicates the probability of a negative test to be a true negative. The NPV also is influenced by the prevalence of the disease. On the contrary of PPV, the NPV is more influenced by specificity than sensitivity. It is also highly influenced by the prevalence of the disease [33].

Positive likelihood ratio (PLR) demonstrates the odds of having the disease. If PLR is superior to 5, it means that the test is related to the disease. It is highly influenced by specificity, as is the PPV. It is why the PLR is much higher for ISG criteria than ICBD (Table 4). Higher PLR for ISG means that, if ISG is positive, the chance of having BD is very high, but unfortunately ISG was negative in around 18% of subjects, in the 3 independent sets (Table 1).

Negative likelihood ratio (NLR) shows the odds of not having the disease. It is highly influenced by the sensitivity, as for the NPV. It has therefore better values for ICBD than for ISG criteria (Table 4). The high NLR for ICBD means that, if ICBD are negative, there are little chances for the patient to have BD (only 2% error rate for the Iranian patients: Table 4).

Diagnostic odds ratio (DOR) is a new way to show how much a test is reliable, like combining the PLR and NLR results. If DOR is 1, it means the test (criteria) does not discriminate between the patient and the control. The power of discrimination increases with higher values of DOR. The DOR of ISG is 294 and of ICBD is 1185 in the Iranian patients, demonstrating the high discriminative power of ICBD over ISG (Table 4).

Youden’s index (YI) is a rather old (1950) and simple calculation, combining the results of sensitivity and specificity, to show the performance of the diagnosis criteria. The result goes from zero to one. The more the result approaches 1, the higher the performance of the test is. The ideal is one, meaning a sensitivity and a specificity of 100%. A sensitivity and a specificity of 90% will give a YI of 0.8. The YI of ISG is inferior to ICBD in China and Iran (Table 4).

4. Conclusion

ICBD are the latest diagnosis/classification criteria, created by the participation of 27 countries from different parts of the world. The large number of Behcet’s disease patients and control patients, from inside and outside of the Silk Road, assures the variability needed to create an international criteria that can work in any country with different ethnicities.
Table 4: Predictive value, likelihood ratio, diagnostic odds ratio, and Youden's index.

<table>
<thead>
<tr>
<th>Study</th>
<th>Reference</th>
<th>Criteria</th>
<th>PPV</th>
<th>NPV</th>
<th>PLR</th>
<th>NLR</th>
<th>DOR</th>
<th>YI</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISG 1990</td>
<td>[11]</td>
<td>ISG</td>
<td>96.8</td>
<td>92.4</td>
<td>30.7</td>
<td>0.08</td>
<td>371.8</td>
<td>0.89</td>
</tr>
<tr>
<td>Iran 1993</td>
<td>[12, 13]</td>
<td>ISG</td>
<td>97.2</td>
<td>87.6</td>
<td>34.5</td>
<td>0.14</td>
<td>243.6</td>
<td>0.84</td>
</tr>
<tr>
<td>APLAR 1998</td>
<td>[21]</td>
<td>ISG</td>
<td>99.0</td>
<td>78.1</td>
<td>103.1</td>
<td>0.28</td>
<td>368.4</td>
<td>0.71</td>
</tr>
<tr>
<td>Russia 2000</td>
<td>[22]</td>
<td>ISG</td>
<td>99.7</td>
<td>83.2</td>
<td>399</td>
<td>0.20</td>
<td>1971</td>
<td>0.80</td>
</tr>
<tr>
<td>ICBD 2006</td>
<td>[17–19]</td>
<td>ISG</td>
<td>95.4</td>
<td>84.5</td>
<td>20.6</td>
<td>0.18</td>
<td>112.4</td>
<td>0.78</td>
</tr>
<tr>
<td>Germany 2008</td>
<td>[31]</td>
<td>ISG</td>
<td>88.8</td>
<td>84.6</td>
<td>7.97</td>
<td>0.18</td>
<td>43.8</td>
<td>0.73</td>
</tr>
<tr>
<td>China 2008</td>
<td>[32]</td>
<td>ISG</td>
<td>98.8</td>
<td>74.1</td>
<td>81.7</td>
<td>0.35</td>
<td>234.4</td>
<td>0.65</td>
</tr>
<tr>
<td>IRAN 2010</td>
<td>[33]</td>
<td>ISG</td>
<td>98.5</td>
<td>81.9</td>
<td>65.1</td>
<td>0.22</td>
<td>294</td>
<td>0.77</td>
</tr>
</tbody>
</table>

PPV: positive predictive value, NPV: negative predictive value, PLR: positive likelihood ratio, NLR: negative likelihood ratio, DOR: diagnostic odds ratio, and YI: Youden's index.

The validation of the criteria in the Far East, Middle-East, and Europe demonstrates its validity.

References


Submit your manuscripts at http://www.hindawi.com