Cureus
Part of SPRINGER NATURE

Open Access Review Article

Diagnostic Accuracy of the Skin Pathergy Test for Behçet's Disease: A Systematic Review of Sensitivity, Specificity, and Methodological Determinants

Review began 11/07/2026
Review ended 11/07/2026
Published 11/11/2026

© Copyright 2026
Abu Taleb et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.119503

Radwan Abu Taleb [1], Osama A. Mobarki [2], Fatima A. Aqili [3], Nesrin H. Eshaq [2], Noran I. Henishi [2], Ramis M. Manni [2], Derwesh I. Al-Ahmed [4], Ghena M. Alsaadi [5], Alhanouf A. Bedaiwi [6], Aeshah H. Sabi [7], Adil A. Alyahmady [4], Jawad M. Alghzwi [8]

1. Department of Dermatology, Jazan General Hospital, Jazan, SAU 2. Faculty of Medicine, Jazan University, Jazan, SAU 3. College of Medicine, Jazan University, Jazan, SAU 4. College of Medicine, King Khalid University, Abha, SAU 5. Department of General Practice, Meena Medical Group (Tawuniya), Jeddah, SAU 6. Faculty of Medicine, University of Tabuk, Tabuk, SAU 7. College of Pharmacy, Jazan University, Jazan, SAU 8. College of Medicine, Dar Al Uloom University, Riyadh, SAU

Corresponding author: Osama A. Mobarki, osmobarki@gmail.com

Abstract

The skin pathergy test is included in major classification criteria for Behçet's disease (BD), but its diagnostic performance and the factors influencing test results remain incompletely understood. This systematic review evaluated the ability of the skin pathergy test to identify BD and examined how variations in testing methodology affect diagnostic performance. Studies assessing the pathergy test in individuals with BD and non-BD comparison groups were reviewed, and diagnostic accuracy data were analyzed. Across the included studies, the test demonstrated consistently high specificity but variable sensitivity. Evidence from studies comparing different testing approaches suggested that procedural factors, particularly needle characteristics and interpretation thresholds, substantially influence test positivity rates. Use of certain needle types was associated with higher detection rates, while more stringent criteria for defining a positive reaction reduced sensitivity. Overall, the findings indicate that a positive skin pathergy test may provide strong support for the diagnosis of BD, whereas a negative result does not reliably exclude the disease. Variability in testing methods appears to be an important contributor to differences in reported diagnostic performance. Greater standardization of test procedures and interpretation is needed to improve consistency and clarify the role of the skin pathergy test in the classification and evaluation of BD.

Categories: Rheumatology
Keywords: behçet's disease, behçet's syndrome, cutaneous hyperreactivity, diagnostic accuracy, diagnostic performance, diagnostic test evaluation, pathergy reaction, sensitivity, skin pathergy test, specificity

Introduction And Background

Behçet's disease (BD) is a systemic inflammatory disorder without a pathognomonic laboratory test. Diagnosis relies on clinical features and established classification criteria [1,2]. The skin pathergy test assesses cutaneous hyperreactivity following needle pricking, typically producing an erythematous papule or pustule within 24-48 hours [2]. Despite its long-standing role in BD classification, its diagnostic performance remains variable [1,3].

The skin pathergy test is incorporated into established classification criteria for BD, making its diagnostic accuracy clinically relevant. A positive result may support disease classification, whereas a negative result cannot reliably exclude the disease [1-4].

Several factors complicate the interpretation of the test. Reported positivity varies substantially across geographical regions and over time. Differences in needle type, application technique, and assessment of skin reactions may contribute to this variability, suggesting that procedures described as the same test can produce different results [5-9]. Furthermore, because the pathergy test is incorporated into classification criteria, studies using these criteria to identify BD may be affected by incorporation bias, in which the index test contributes to the reference standard [10,11].

Although previous reviews have discussed the pathergy test and the diagnostic performance of BD classification criteria, a systematic assessment of the test itself remains necessary [1-4]. This systematic review examines the reported diagnostic accuracy of the skin pathergy test, evaluates the risk of bias in the available evidence, and explores methodological and clinical factors that may explain differences in test performance. The findings aim to clarify the test's strengths and limitations and inform its interpretation in clinical practice and its role in BD classification.

How to cite this article
Abu Taleb R, Mobarki O A, Aqili F A, et al. (November 11, 2026) Diagnostic Accuracy of the Skin Pathergy Test for Behçet's Disease: A Systematic Review of Sensitivity, Specificity, and Methodological Determinants. Cureus 18(11): e119503. DOI 10.7759/cureus.119503

Review

Methods

This systematic review was conducted and reported in accordance with the PRISMA statement [12]. The review was designed to evaluate the diagnostic accuracy of the skin patch test in BD. The review question is summarized in Table *1*.

Component	Description
Population (P)	Individuals with suspected or established BD and non-BD comparison groups
Index test (I)	Skin pathergy test and its variants (conventional needle test, blunt needle techniques, oral pathergy, laser-induced pathergy, and other procedural modifications)
Comparator (C)	Individuals without BD, including healthy controls and patients with alternative diagnoses
Outcome (O)	Diagnostic accuracy measures of the pathergy test, including sensitivity, specificity, true-positive, false-positive, true-negative, and false-negative results

TABLE 1: PICO framework for the systematic review of the diagnostic accuracy of the skin pathergy test in BD.

This table defines the review question according to the PICO framework. It outlines the target population, the pathergy test methods evaluated, the comparison groups, and the diagnostic accuracy outcomes assessed in this systematic review.

BD, Behçet's disease

Eligibility Criteria

Studies were eligible if they evaluated any form of the skin patch test in people with suspected or established BD and included a non-BD comparison group, with sufficient data to assess diagnostic performance. Studies published before 1990 were excluded because diagnostic criteria were rarely specified, limiting assessment of the reference standard. This restriction was introduced as a protocol amendment. We also excluded studies without a comparison group, studies targeting conditions other than BD, studies investigating procedure-related pathergy rather than a deliberately performed test, mechanistic studies using pathergy status only for subgrouping, publications without primary comparative data, and studies with fewer than 10 participants in either group.

Index Test

The conventional skin pathergy test was defined as a dry needle prick assessed clinically after 24-48 hours, with an erythematous papule or pustule of at least 2 mm considered positive. Each study's diagnostic threshold was recorded as reported. Data were extracted on needle type and gauge, number of pricks, application route, substances used, test site, skin preparation, reading time, assessment method, and reader blinding. Variants involving blunt needles, intravenous administration, saline, oral mucosa, laser induction, or alternative grading methods were treated as distinct index tests. Within-study comparisons were used to examine differences between techniques.

Target Condition and Reference Standard

The target condition was BD. Acceptable reference standards included established international or national classification criteria and documented expert clinical diagnosis. Because some classification criteria incorporate the pathergy test, studies were categorized into three prespecified verification strata: V1, reference standards independent of the pathergy result; V2, reference standards incorporating pathergy with an indeterminate contribution; and V3, reference standards in which pathergy was demonstrably necessary for classifying a substantial proportion of cases. These categories were used to assess the potential influence of incorporation bias, whereby the index test contributes to the reference standard.

Search Strategy and Study Selection

PubMed/MEDLINE was searched from inception to September 3, 2025 using complementary strategies targeting the pathergy test and studies validating diagnostic criteria or reporting pathergy results within

larger cohorts. The Cochrane Central Register of Controlled Trials (CENTRAL), Web of Science Core Collection, and Scopus were searched using equivalent strategies. No language or date restrictions were applied at the search stage, and no diagnostic-accuracy filter was used. The index-test strategy was validated against 12 previously identified eligible studies and retrieved all 12 (Table *2*).

Item	Description
Databases searched	PubMed/MEDLINE, Cochrane CENTRAL, Web of Science Core Collection, and Scopus
Coverage period	Database inception to September 3, 2025
Language restrictions	None
Date restrictions	None during searching; studies published before 1990 excluded during eligibility assessment
Core concepts	BD, pathergy test, skin hyperreactivity, diagnostic accuracy, sensitivity, specificity, validation
Example PubMed search	("Behcet Syndrome"[Mesh] OR Behcet* OR Behçet*) AND (pathergy OR "pathergy test" OR "skin pathergy" OR hyperreactivity)
Additional strategy	Complementary searches for studies validating Behçet classification criteria and reports containing pathergy-test diagnostic data
Study selection process	Deduplication, title/abstract screening, full-text review, and inclusion according to predefined eligibility criteria

TABLE 2: Electronic database search strategy used for identification of eligible studies.

This table summarizes the literature search methods used in the systematic review, including databases searched, coverage period, search concepts and terms, date of the last search, eligibility restrictions, and study selection procedures. The strategy was designed to identify studies evaluating the diagnostic performance of the skin pathergy test in BD.

BD, Behçet's disease

After deduplication, records were screened against the eligibility criteria, and exclusion reasons were documented. Potentially eligible full texts were assessed using the same criteria. Data were extracted independently by two reviewers, with disagreements resolved through discussion or arbitration by a third reviewer. Extracted information included study design, setting, participant numbers, comparison-group characteristics, testing technique, positivity threshold, reference standard, blinding, verification stratum, and data required to assess diagnostic performance. No values were inferred unless they could be derived arithmetically from reported data. When multiple publications arose from the same cohort or registry, the report providing the most complete reconstructible diagnostic-accuracy data was designated the primary study for quantitative assessment. Companion reports from the same cohort were retained only for supplementary narrative information and were not treated as independent studies contributing additional diagnostic-accuracy data.

Risk of Bias and Data Synthesis

Risk of bias was assessed independently by two reviewers using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool [13]. Prespecified rules addressed patient selection, blinding of test interpretation, potential incorporation bias, and differential verification. Two-gate studies were considered at high risk of bias in patient selection. Failure to report blinding was considered high risk for test interpretation.

Studies in verification stratum V3 were considered at high risk of bias in the reference standard domain because the pathergy test formed part of the reference standard and was demonstrably necessary for classification in a substantial proportion of cases. Studies in verification stratum V2 were judged as having unclear or potentially elevated risk of bias because the degree of incorporation of the pathergy test into the reference standard could not be determined.

Results

Study Selection

Study selection is summarized in Figure *1*. After screening and full-text review according to the prespecified eligibility criteria, 11 studies were included in the review. Seven provided reconstructible two-by-two contingency tables, while four contributed descriptive findings only. Three of the narrative-only reports

originated from the same Tehran BD registry as the largest included study. These reports were not considered independent study populations and were included solely to provide supplementary descriptive information regarding temporal trends and the contribution of pathergy testing to classification criteria. Their participants were not counted as additional unique subjects (Table 3).

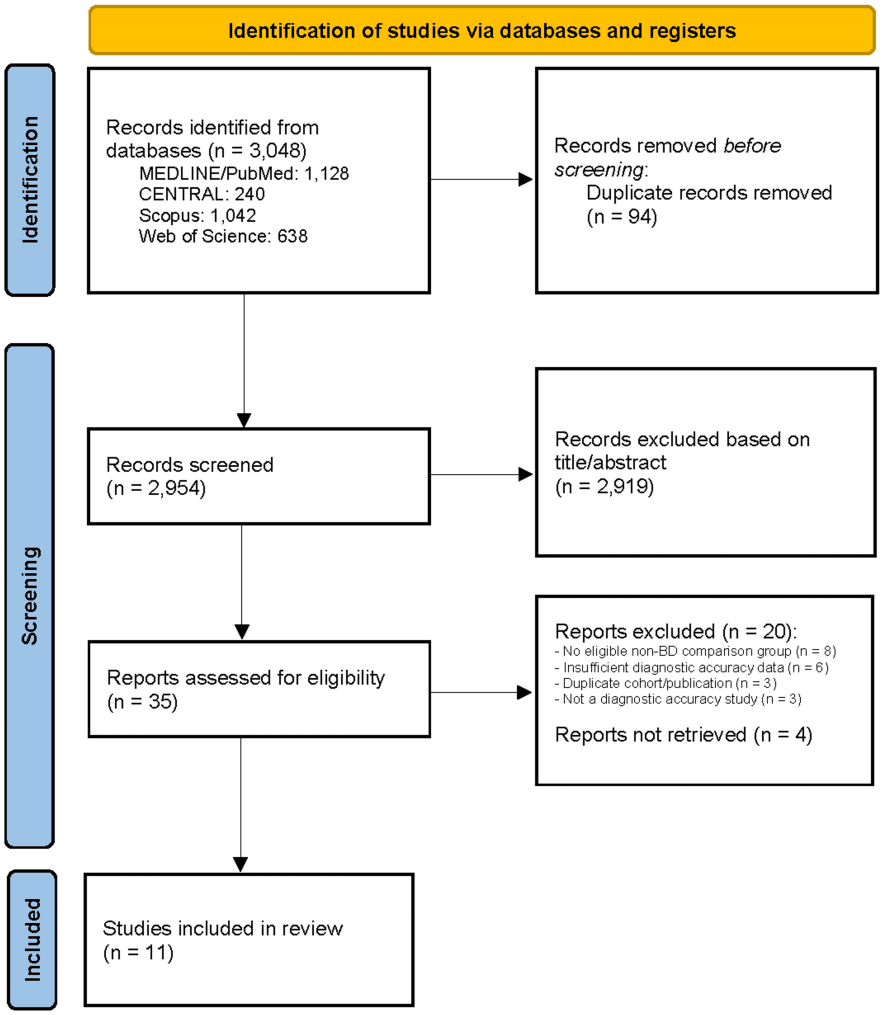


FIGURE 1: PRISMA flow diagram depicting the study selection process for the systematic review.

Following the identification of records through database searching and other sources, duplicates were removed, and the remaining records were screened. Full-text articles were assessed for eligibility, with exclusions documented along with reasons. Studies meeting the inclusion criteria were included in the final synthesis.

Study	Country	Design	BD tested (n)	Controls tested (n)	Control group	Reference standard	Stratum	Contribution
Davatchi et al. [1]	Iran	Registry	6484	4179	Referred patients in whom BD was excluded	Expert clinical diagnosis	V1	2 × 2
Dilşen et al. [2]	Turkey	Two-gate	92	192	64 healthy participants and 128 participants with non-BD diseases	Dilşen criteria	V2	2 × 2
Akmaz et al. [3]	Turkey	Two-gate	43	15	Dermatological disease	Not stated	Unclear	2 × 2
Sharquie et al. [4]	Iraq	Two-gate	83	120	90 healthy participants and 30 participants with other skin diseases	ISG 1990 criteria	V2	2 × 2
Budak-Alpdoğan et al. [5]	Turkey	Two-gate	30	132	Inflammatory arthropathy, recurrent oral ulcers, CML with or without interferon treatment, and healthy participants	ISG 1990 criteria	V2	2 × 2
Aydın et al. [6]	Turkey	Two-gate	31	62	31 participants with familial Mediterranean fever and 31 healthy participants	ISG 1990 criteria	V2	2 × 2
Sayarlıoğlu et al. [7]	Turkey	Two-gate	35	28	Healthy volunteers	ISG 1990 criteria	V2	2 × 2
Davatchi et al. [8]	Iran	Registry	6727	4648	Referred patients in whom BD was excluded	Expert clinical diagnosis	V1	Narrative (no denominators)
Davatchi et al. [9]	Iran	Registry	7504	-	Referred patients in whom BD was excluded	Expert clinical diagnosis	V1	Narrative (no controls)
Davatchi et al. [10]	Iran	Registry	4717	2407	Referred patients in whom BD was excluded	Expert clinical diagnosis	V1	Narrative (no pathergy, 2 × 2 data)
Matucci-Cerinic et al. [11]	Multinational	Two-gate	18	-	Autoinflammatory confounding diseases	Expert consensus	V1	Narrative (control data incomplete)

TABLE 3: Characteristics of the 11 studies included in the systematic review.

This table summarizes the country, study design, numbers of participants with BD and controls, control-group characteristics, reference standards, verification strata, and contribution of each study to the diagnostic evidence. Studies are classified according to the reference standard used: V1 indicates a reference standard independent of the pathergy test, V2 indicates a reference standard incorporating the pathergy test with an indeterminate contribution, and V3 indicates a reference standard in which the pathergy test was demonstrably necessary for classifying a substantial proportion of cases. The contribution column indicates whether a study provided reconstructible two-by-two contingency data or descriptive findings only.

Studies [8-10] represent additional analyses of the same Tehran registry underlying study [1]. These reports were not considered independent cohorts and did not contribute separate diagnostic-accuracy datasets.

BD, Behçet's disease; CML, chronic myeloid leukemia; ISG, International Study Group; 2 × 2, two-by-two contingency table

Characteristics of the Included Studies

The 11 included studies evaluated the skin pathergy test in different populations and settings. The seven studies with reconstructible diagnostic data were conducted in Turkey, Iraq, and Iran and included 6,798 patients with BD and 4,728 non-BD participants [1-7]. Comparison groups comprised patients with other rheumatic or inflammatory diseases, mixed conditions that could mimic BD, dermatological diseases, and healthy volunteers. Reference standards included established classification criteria and expert clinical diagnosis. Five studies used criteria that incorporated the pathergy test, one used expert clinical diagnosis, and one did not clearly specify the diagnostic criteria.

Diagnostic Performance

Reported sensitivity varied substantially across studies, ranging from 0.11 to 0.69, whereas specificity was generally high. Five of the seven studies with reconstructible diagnostic data reported a specificity of 1.00. The remaining studies reported specificities of 0.94 and 0.93, respectively. The lowest specificity occurred in a study whose comparison group included patients with chronic myeloid leukemia receiving interferon treatment [5]. These findings indicate that the test's ability to identify BD varied considerably despite generally high specificity (Table *4*).

Study	Country	BD tested (n)	Controls tested (n)	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Davatchi et al. [1]	Iran	6484	4179	3393	266	3091	3913	0.52 (0.51-0.54)	0.94 (0.93-0.94)
Dilşen et al. [2]	Turkey	92	192	10	0	82	192	0.11 (0.05-0.17)	1.00 (1.00-1.00)
Akmaz et al. [3]	Turkey	43	15	5	0	38	15	0.12 (0.02-0.21)	1.00 (1.00-1.00)
Sharquie et al. [4]	Iraq	83	120	57	0	26	120	0.69 (0.59-0.79)	1.00 (1.00-1.00)
Budak-Alpdoğan et al. [5]	Turkey	30	132	15	9	15	123	0.50 (0.32-0.68)	0.93 (0.89-0.97)
Aydın et al. [6]	Turkey	31	62	13	0	18	62	0.42 (0.25-0.59)	1.00 (1.00-1.00)
Sayarlıoğlu et al. [7]	Turkey	35	28	9	0	26	28	0.26 (0.11-0.40)	1.00 (1.00-1.00)

TABLE 4: Two-by-two contingency data and diagnostic accuracy of the conventional skin pathergy test.

This table presents the numbers of TP, FP, FN, and TN results for the conventional skin pathergy test in the seven studies providing reconstructible two-by-two contingency data. Sensitivity and specificity are reported with their corresponding 95% CIs. The final row presents the bivariate summary estimates across the seven studies. TP indicates a positive pathergy test in a participant with BD; FP, a positive test in a participant without BD; FN, a negative test in a participant with BD; and TN, a negative test in a participant without BD.

BD, Behçet's disease; FN, false negative; FP, false positive; TN, true negative; TP, true positive

Variation According to Testing Technique

Within-study comparisons suggested that needle type influenced test positivity. In one study involving 92 patients tested using four methods, a blunt needle produced positive results in 48% of patients, compared with 11% using a sharp needle; no positive results occurred among 192 controls [2]. Another study reported positive results in 72% of patients tested with a blunt intradermal needle, compared with 12% of those tested with a disposable sharp needle; no positive results occurred among controls [3].

The interpretation threshold also affected sensitivity. Applying a stricter grading method to the same test reactions in 83 patients reduced sensitivity from 0.69 to 0.47 without changing specificity. An oral mucosal version of the test, performed on the lower lip, had a sensitivity of 0.47 and a specificity of 1.00 in the same participants [4]. By comparison, argon laser-induced inflammation produced a sensitivity of 0.23, compared with 0.26 for conventional needle testing in the same participants [7].

Changes in Diagnostic Performance Over Time

The largest included study examined 6,484 patients with BD and 4,179 controls across four consecutive periods. Sensitivity declined from 0.64 in the earliest period to 0.36 in the latest, while specificity increased from 0.87 to approximately 0.98-0.99 [1]. These changes occurred within a single center and cohort, suggesting that variation in diagnostic performance over time cannot be explained solely by differences between study populations.

The three companion reports from this registry provided additional descriptive findings. One reported a positive pathergy test in 51.1% of 6,727 patients and 5.9% of 4,648 controls [8]. Another reported positive results in 3,782 of 7,504 patients (50.4%) [9]. A separate analysis found that removing the pathergy item reduced the sensitivity of the International Study Group (ISG) criteria from 82% to 66% and that of the Dilşen criteria from 86% to 68%, with slight improvements in specificity [10]. In a pediatric validation cohort, the test was positive in seven of 18 patients with confirmed BD, although the control-group denominators were not reported [11].

Risk of Bias and Limitations of the Evidence

All seven studies with reconstructible diagnostic data were judged to have a high risk of bias in patient selection. Six recruited cases and controls separately, while one did not adequately describe its sampling frame. None explicitly reported interpreting test readings without knowledge of the diagnosis, resulting in a high risk of bias in the index-test domain under the prespecified assessment rules (Table *5*).

Study	Patient selection	Index test	Reference standard	Flow and timing	Applicability: patients	Applicability: index test	Applicability: reference standard
Davatchi et al. [1]	Unclear	High	Unclear	Low	Low	Low	Low
Dilşen et al. [2]	High	High	High	Low	Low	High	Low
Akmaz et al. [3]	High	High	Unclear	Low	Low	High	Unclear
Sharquie et al. [4]	High	High	High	Low	Unclear	Unclear	Low
Budak-Alpdoğan et al. [5]	High	High	High	Unclear	Low	Low	Low
Aydın et al. [6]	High	High	High	Low	Low	Low	Low
Sayarlıoğlu et al. [7]	High	High	High	Low	High	Low	Low

TABLE 5: Risk-of-bias and applicability assessments of the seven studies contributing two-by-two contingency data.

Risk of bias and applicability were assessed using the QUADAS-2 tool. Risk of bias was evaluated across four domains: patient selection, index test, reference standard, and flow and timing. Applicability was assessed across three domains: patient selection, index test, and reference standard. Each domain was judged as having a low, high, or unclear risk of bias or applicability concern. Index-test risk-of-bias judgements were high for all seven studies because none explicitly reported blinded interpretation of pathergy test results. High applicability concern for the index test reflects the use of blunt reusable needles or multiple simultaneous applications rather than methods representative of current practice.

QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies 2

Five studies used reference standards incorporating the pathergy test and were therefore considered at high risk of incorporation bias. One used expert clinical diagnosis, while another did not specify its diagnostic criteria. Flow and timing were judged to be at low risk of bias in five studies.

The limited number of studies, variation in testing methods and reference standards, and incomplete reporting restricted comparisons across subgroups. Only one study used a reference standard independent of the pathology test, preventing a meaningful comparison between studies with independent and potentially incorporated reference standards. The evidence therefore supports a descriptive assessment of diagnostic performance but does not establish the independent contribution of each methodological factor.

Discussion

The present review found that the skin pathergy test has consistently high specificity but highly variable sensitivity in BD. These findings indicate that a positive test result can provide strong support for the diagnosis in an appropriate clinical setting, whereas a negative result does not reliably exclude the disease. This observation is consistent with the role of the pathergy test in established and contemporary classification systems, including the ISG criteria and the International Criteria for Behçet's Disease, in which pathergy serves as a supportive rather than definitive diagnostic finding [14,15].

A major finding of this review was the influence of testing methodology on diagnostic performance. Variations in needle type, application technique, and positivity thresholds were associated with substantial differences in test positivity rates across studies. Previous reports have similarly documented geographic and temporal variation in pathergy positivity, with higher rates historically reported in populations from regions along the traditional Silk Road and declining positivity observed in more recent cohorts [2,15]. Together, these observations suggest that both procedural and population-related factors contribute to the heterogeneity reported in the literature and support the need for greater standardization of testing procedures and interpretation criteria.

The available evidence should be interpreted in light of important limitations. Most included studies originated from Turkey, Iran, and Iraq, potentially limiting generalizability to other populations. In addition, many studies used case-control designs, did not report blinded assessment of test results, and employed reference standards that incorporated the pathergy test itself, creating a risk of incorporation bias. Despite

these limitations, the evidence supports the continued use of the skin pathergy test as a valuable adjunctive tool in the assessment of BD, particularly when interpreted alongside clinical findings. Future prospective studies using standardized protocols and independent reference standards may help clarify its contemporary diagnostic role and improve consistency across different populations [14,15].

Limitations

The small number of studies, geographical concentration, and substantial risk of bias limited the evidence. Most studies were conducted in Turkey, Iraq, and Iran, limiting generalizability to other populations. All seven studies with reconstructible diagnostic data had a high risk of bias in patient selection, and none reported blinded test interpretation. One study contributed most of the participants, further limiting the independence and representativeness of the findings. Four additional eligible studies lacked sufficient data for complete diagnostic assessment. The small evidence base also prevented meaningful subgroup analyses and robust assessment of incorporation bias.

Conclusions

The skin pathergy test remains a useful adjunct in the evaluation and classification of BD because of its generally high specificity, indicating that a positive result can provide strong supportive evidence for the diagnosis in an appropriate clinical context. However, the wide variation in sensitivity observed across studies demonstrates that a negative test result cannot reliably exclude the disease and should not deter further clinical assessment when suspicion remains high. The findings suggest that differences in testing methodology, including needle characteristics, application techniques, and interpretation thresholds, are important contributors to variability in reported diagnostic performance. Interpretation of the available evidence is limited by methodological weaknesses in the primary studies, including risks of selection bias, lack of blinded assessment, and potential incorporation bias associated with some reference standards. Overall, the evidence supports the continued use of the skin pathergy test as a complementary diagnostic tool rather than a stand-alone test. Greater standardization of test procedures, consistent reporting practices, and the use of rigorous study designs in representative patient populations are needed to improve comparability across studies and to better define the test's role in contemporary clinical practice and future classification frameworks for BD.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Fatima A. Aqili, Osama A. Mobarki

Critical review of the manuscript for important intellectual content: Fatima A. Aqili, Radwan Abu Taleb, Osama A. Mobarki, Ramis M. Manni, Derwesh I. Al-Ahmed, Ghena M. Alsaadi, Alhanouf A. Bedaiwi, Aeshah H. Sabi, Jawad M. Alghzwi

Acquisition, analysis, or interpretation of data: Radwan Abu Taleb, Nesrin H. Eshaq, Noran I. Henishi, Ramis M. Manni, Derwesh I. Al-Ahmed, Ghena M. Alsaadi, Alhanouf A. Bedaiwi, Aeshah H. Sabi, Adil A. Alyahmady, Jawad M. Alghzwi

Drafting of the manuscript: Nesrin H. Eshaq, Noran I. Henishi, Adil A. Alyahmady

Disclosures

AI use statement: The authors used Microsoft Copilot (Microsoft Corporation, Redmond, WA, USA) as follows: language polishing in Microsoft Copilot was used to assist with grammar, spelling, and language editing. The authors reviewed and approved all revisions and take full responsibility for the content of the manuscript. The authors confirm that they have reviewed and verified all AI-assisted content and take full responsibility for the work. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. 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, 30:1151-5. 10.1007/s10067-011-

1694-5
2. Dilşen N, Koniçe 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. 10.1136/ard.52.11.823
3. Akmaz O, Erel A, Gürer MA: Comparison of histopathologic and clinical evaluations of pathergy test in Behçet's disease. Int J Dermatol. 2000, 39:121-5. 10.1046/j.1365-4362.2000.00884.x
4. Sharquie KE, Al-Araji A, Hatem A: Oral pathergy test in Behçet's disease. Br J Dermatol. 2002, 146:168-9. 10.1046/j.1365-2133.2002.4513_5.x
5. Budak-Alpdogan T, Demirçay, Alpdogan O, et al.: Skin hyperreactivity of Behçet's patients (pathergy reaction) is also positive in interferon alpha-treated chronic myeloid leukaemia patients, indicating similarly altered neutrophil functions in both disorders. Br J Rheumatol. 1998, 37:1148-51. 10.1093/rheumatology/37.11.1148
6. Aydin F, Akpolat T, Senturk N, Bagci H, Yasar Turanli A: Evaluation of pathergy test positivity in familial Mediterranean fever patients and comparison of clinical manifestations of FMF with Behçet's disease. Clin Rheumatol. 2009, 28:1331-5. 10.1007/s10067-009-1237-5
7. Sayarlioglu M, Calka O, Cinal A, et al.: Comparison of argon laser photocoagulation-induced cutaneous inflammation and skin pathergy test in Behçet's disease. Clin Rheumatol. 2010, 29:59-63. 10.1007/s10067-009-1294-9
8. Davatchi F, Sadeghi Abdollahi B, Chams-Davatchi C, et al.: Impact of the positive pathergy test on the performance of classification/diagnosis criteria for Behcet's disease. Mod Rheumatol. 2013, 23:125-32. 10.3109/s10165-012-0626-9
9. Davatchi F, Shahram F, Chams-Davatchi C, et al.: Behcet's disease in Iran: analysis of 7641 cases. Mod Rheumatol. 2019, 29:1023-30. 10.1080/14397595.2018.1558752
10. Davatchi F, Shahram F, Nadji A, et al.: Influence of pathergy test on the accuracy of different diagnosis criteria for Behçet's disease. Adv Exp Med Biol. 2003, 528:109-12. 10.1007/0-306-48382-3_22
11. Matucci-Cerinic C, Palluy H, Al-Mayouf SM, et al.: Validation of the PEDiatric Behçet's Disease classification criteria: an evidence-based approach. Rheumatology (Oxford). 2024, 63:3422-31. 10.1093/rheumatology/kead609
12. Page MJ, McKenzie JE, Bossuyt PM, et al.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews [Article in Portuguese]. Rev Esp Cardiol (Engl Ed). 2021, 74:790-9. 10.1016/j.rec.2021.07.010
13. Whiting PF, Rutjes AW, Westwood ME, et al.: QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011, 155:529-36. 10.7326/0003-4819-155-8-201110180-00009
14. Criteria for diagnosis of Behçet's disease. International Study Group for Behçet's Disease. Lancet. 1990, 335:1078-80.
15. The International Criteria for Behçet's Disease (ICBD): a collaborative study of 27 countries on the sensitivity and specificity of the new criteria. J Eur Acad Dermatol Venereol. 2014, 28:338-47. 10.1111/jdv.12107