 DOI 10.1111/j.1365-2133.2007.08412.x

Diagnostic criteria for atopic dermatitis: a systematic review

E.E.A. Brenninkmeijer,* M.E. Schram,* M.M.G. Leeflang,†‡ J.D. Bos* and Ph.I. Spuls*‡

*Department of Dermatology, †Department of Clinical Epidemiology, Biostatistics and Bioinformatics and ‡Dutch Cochrane Centre, Academic Medical Center A0-251, University of Amsterdam, PO Box 22660, 1100 DD Amsterdam, the Netherlands

Correspondence
E.E.A. Brenninkmeijer.
E-mail: e.e.brenninkmeijer@amc.uva.nl

Accepted for publication
1 October 2007

Key words
atopic dermatitis, definition, diagnostic criteria, QUADAS, validity

Conflicts of interest
None declared.

Summary

Background Atopic dermatitis (AD) has a wide spectrum of dermatological manifestations and despite various validated sets of diagnostic criteria that have been developed over the past decades, there is disagreement about its definition. Nevertheless, clinical studies require valid diagnostic criteria for reliable and reproducible results.

Objective To summarize the evidence concerning the validity of diagnostic criteria for AD.

Methods All data sources were identified through searches on Medline, Embase and Cochrane databases. The Quality Assessment of Diagnostic Accuracy tool (QUADAS) was used. Results are presented in a receiver operating characteristic (ROC) plot.

Results Out of the 20 articles that met the criteria, 27 validation studies were identified. In two studies concerning Hanifin and Rajka diagnostic criteria sensitivity and specificity ranged from 87·9% to 96·0% and from 77·6% to 93·8%, respectively. Nineteen validation studies of the U.K. diagnostic criteria showed sensitivity and specificity ranging from 10% to 100% and 89·3% to 99·1%, respectively. Three validation studies concerning the Schultz-Larsen criteria showed sensitivity from 88% to 94·4% and specificity from 77·6% to 95·9%. In one article concerning the criteria of Diepgen, the sensitivity ranged from 83·0% to 87·7% and the specificity from 83·9% to 87·0%. One article studied the Kang and Tian criteria and reported 95·5% sensitivity and 100% specificity. One article validating the International Study of Asthma and Allergies in Childhood (ISAAC) criteria showed a positive and negative predictive value of 48·8% and 91·1%, respectively.

Conclusion With this systematic review of the existing sets of diagnostic criteria for AD a varying number of validation studies with varying methodological quality was found. The U.K. diagnostic criteria are the most extensively validated. However, improvement of methodological design for validation studies and uniformity in well-validated and applicable diagnostic criteria are needed to improve future intervention studies and to compare study results.

Atopic dermatitis (AD) has a wide spectrum of dermatological manifestations (e.g. presentation, severity and distribution) and there is disagreement about its definition. Nevertheless, results and reproducibility of genetic, aetiological, epidemiological, diagnostic and therapeutic studies depend on establishing reliable and valid diagnostic criteria. During the past decades various lists of diagnostic criteria for AD have been proposed (Table 1).[1–10]

Uniformity in the use of diagnostic criteria for AD is lacking. In 23% of the published clinical trials concerning AD the diagnostic criteria for the diagnosis of AD were not specified. The Hanifin and Rajka diagnostic criteria were used in 44% of the trials and the U.K. diagnostic criteria in 12%.[11]

The objective of this systematic review was to summarize the evidence concerning the validity of diagnostic criteria for AD.

Table 1 Diagnostic criteria for atopic dermatitis

Criteria list	Requirements (number of criteria)
Hanifin and Rajka diagnostic criteria, 1980	3 major + 3 minor (27)
Kang & Tian diagnostic criteria, 1989	1 basic + 3 minor (5)
Schultz-Larsen criteria, 1992	≥ 50 points (6)
Lillehammer criteria, 1994	Visible eczema + 4 minor (12)
U.K. diagnostic criteria, 1994	Pruritus + 3 minor (6)
ISAAC questionnaire, 1995	Score ≥ 3 (7)
Japanese Dermatology Association criteria, 1995	All 3 features (3)
Criteria of Diepgen, 1996	≥ 10 points (8)
Millennium diagnostic criteria, 1998	Allergen-specific IgE + 2 principal (4)
Danish Allergy Research Centre (DARC), 2005	3 features (3)

ISAAC, International Study of Asthma and Allergies in Childhood.

Methods

Inclusion and exclusion criteria

Randomized controlled trials, and case–control, cross-sectional and cohort studies that validated one or more of the diagnostic criteria for AD were assessed for eligibility. Included were studies that: (i) concerned the existing diagnostic criteria for AD (Table 1); (ii) were hospital or community based; (iii) were validation studies of translated criteria; (iv) reported sensitivity, specificity, positive predictive value (PPV) or negative predictive value (NPV) or had the possibility for calculating these outcome measures.

Excluded were studies in which: (i) the complete set of diagnostic criteria was not considered (e.g. only the major or the minor criteria of the Hanifin and Rajka criteria); (ii) the U.K. diagnostic criteria were positive if pruritus plus two, pruritus plus four, or pruritus plus five of the additional criteria were fulfilled, because the general recommended format by Williams *et al.*[8] for use in epidemiological studies requires pruritus plus three or more other features; (iii) parents qualified the presence of AD in their children by self-reporting questionnaires.

No restrictions were imposed with regard to the reference standards used. In addition, no restrictions were used for age, sex and skin type of the subjects. Language of the studies was not a limitation.

Literature search

A literature search was carried out between March and June 2007 on Medline, Embase and the Cochrane Library (CDSR, DARE and CENTRAL) databases (Table 2). Synonyms of atopic dermatitis yielded no additional, relevant articles and were therefore not mentioned in the search strategy. There was no limit to the date of publication. References cited in published articles were examined until no further study was identified. Additionally, articles written by the designers of the diagnostic criteria were screened for eligibility.

Study selection

All articles with titles and abstracts considering AD and diagnostic criteria were selected by one author (M.S.) for relevance. In case of doubt, an assessment by a second reviewer was performed. To determine eligibility, two reviewers (E.B. and M.S.) independently assessed the full texts of the articles. Disagreements were resolved by discussion.

Assessment of methodological quality

For the methodological quality assessment, we applied the Quality Assessment of Diagnostic Accuracy tool (QUADAS) (Table 3).[12] This tool uses predefined criteria based on elements of study design, conduct and analysis which are likely to have a direct relationship to bias in test accuracy studies. It was developed by a panel of nine experts in the field of diagnostic accuracy and consists of 14 validated questions (see Table 3 items 1–14).[13,14] A background document clarifies terms and indicates how the 14 items should be scored by Yes, No or Unclear. QUADAS does not incorporate a quality score.[15,16]

Comments on the application of QUADAS

Item 1 (Was the spectrum of patients representative of the patients who will receive the test in practice?) was scored with a Yes if patients with a dermatological disease other than AD were assigned to the control group. In case of healthy controls, a No was scored. The reference standard likely to classify AD correctly is the clinical diagnosis by an experienced dermatologist. Other reference standards were scored with No (item 3). The time interval between the reference test and the index test was considered short enough if it was shorter than 2 weeks (item 4). Because of lack of relevance, item 12 (Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?) was omitted. As the clinical diagnosis of AD as the reference standard and diagnostic criteria as the index test are inseparable it is impossible to blind the validation process. Therefore, item 7 (Was the reference standard independent of

Table 2 Search strategy for Medline and Embase databases

MEDLINE	
#1	(''Dermatitis, Atopic''[mh] OR atopic dermatiti*[tw] OR atopic eczem*[tw] OR allergic dermatit*[tw] OR allergic eczem*[tw] OR ''intrinsic AD''[tw]) NOT ((animals[mh] OR dogs) NOT humans[mh])
#2	((U.K.[tiab] OR UK[tiab] OR united kingdom[tiab] OR millennium[tiab]) AND (working*[tiab] OR criteria*[tiab])) OR Hanifin*[tiab] OR Rajka*[tiab] OR william*[tiab] OR larsen*[tiab] OR schultz-larsen*[tiab] OR DARC[tiab] OR Tian[tiab] OR Kang[tiab] OR lillehammer OR ISAAC[tiab] OR ((danish[tw] OR denmark[tw] OR japanese[tw]) AND criteri*) OR (International stud*[tiab] AND Asthma[tiab] AND Allerg*[tiab] AND Child*[tiab])
#3	(diagnostic criteri*[tw] OR diagnostic feature*[tw] OR minor criteri*[tw] OR major criteri*[tw] OR minimum criteri*[tw] OR diagnostic feature*[tw] OR diagnostic score*[tw] OR classification criteri*[tw] OR clinical feature*[tw] OR clinical criteri*[tw] OR feature set*[tw] OR score criteri* OR basic feature* OR minor feature* OR major feature* OR clinical feature* OR cutaneous feature*) AND (validation studies[pt] OR validat*[tw] OR Sensitivity and Specificity[mh] OR accura* OR specificity[tiab] OR ''false negative''[tw] OR ''Predictive Value of Tests''[mh] OR Reference Standards[mh] OR prevalence OR logistic models[mh] OR Algorithms[mh] OR reproducibility[tw] OR significance OR diagnosis, differential[mh])
#4	(survey*[tw] OR interview*[tw] OR questionnair*[tw]) AND (diagnosis OR diagnostic OR prevalence) AND (validation studies[pt] OR validat*[tw] OR Sensitivity and Specificity[mh] OR accura* OR specificity[tiab] OR ''false negative''[tw] OR ''Predictive Value of Tests''[mh] OR hospital-based[tw] OR diagnostic outcome*)
#5	criteria[ti] OR validat*[ti]
#6	#2 OR #3 OR #4 OR #5
#7	#1 AND #6
#8	((classificat*[ti] OR diagnos*[ti] OR prevalenc*[ti] OR occurrence[ti] OR incidence[ti] OR epidemiol*[ti]) AND (atopic dermatiti*[ti] OR atopic eczem*[ti] OR allergic dermatit*[ti] OR allergic eczem*[ti] OR ''intrinsic AD''[ti])) NOT ((animals[mh] OR dogs) NOT humans[mh])
#9	#7 OR #8
EMBASE	
1.	Atopic Dermatitis/
2.	((atopic$ or intrins$ or allergic$) adj3 (dermatit$ or eczem$)).tw. 1
3.	1 or 2
4.	exp animal/not (exp animal/ and exp human/)
5.	3 not 4
6.	(((UK or united kingdom or millennium) and (working$ or criteria$)) or Hanifin$ or Rajka$ or larsen$ or schultz-larsen$ or DARC or Tian or Kang or lillehammer or ISAAC or ((danish or denmark or japanese) and criteri$) or (international adj4 stud$ adj4 Asthma adj4 Allerg$ adj4 Child$)).ti,ab.
7.	6 and 5
8.	((atopic$ or intrins$ or allergic$) adj1 (dermatit$ or eczem$)).tw.
9.	1 or 8
10.	9 not 4
11.	((diagnostic adj (criteri$ or feature$ or score$)) or ((minor or major or minimum or classification or clinical or score) adj criteri$) or ((clinical or set$ or basic or minor or major or cutaneous) adj1 feature$)).ti,ab.
12.	diagnostic accuracy/ or diagnostic value/ or differential diagnosis/ or prediction/ or reproducibility/ or exp reliability/ or probability/ or statistical model/ or FUNCTIONAL ASSESSMENT/ or CLINICAL ASSESSMENT/ or exp diagnostic procedure/ or exp standard/ or disease severity/
13.	(validat$ or accura$ or prevalence or specificity or reproducibility or significance or (false adj negative) or (predictive adj value)).mp.
14.	12 or 13
15.	10 and 11 and 14
16.	(criteria or validat$).ti.
17.	10 and 16
18.	(classificat$ or diagnos$ or prevalenc$ or occurence or incidence or epidemiol$).ti.
19.	(atopic dermatiti$ or atopic eczem$ or allergic dermatit$ or allergic eczem$ or ''intrinsic AD'').ti.
20.	18 and 19
21.	20 not 4
22.	(interview$ or questionnair$ or survey$).mp.
23.	(validat$ or accura$ or specificity or reproducibility or (false adj negative) or ((predictive adj value) or hospital-based or diagnostic outcome$)).mp.
24.	diagnostic value/ or diagnostic accuracy/
25.	prediction/ or reproducibility/ or exp reliability/ or probability/ or statistical model/ or FUNCTIONAL ASSESSMENT/ or CLINICAL ASSESSMENT/
26.	(diagnos$ or prevalence).mp.
27.	25 and 26
28.	23 or 24 or 27
29.	10 and 22 and 28
30.	7 or 15 or 17 or 21 or 29

the index test?) of the QUADAS tool was considered to be not applicable.

Data extraction and analysis

For each included study, data on study characteristics (both clinical and methodological) and on test accuracy (QUADAS) were independently extracted by the two first authors. For this purpose a data extraction form was designed. Disagreements about data extraction were resolved by discussion. Study characteristics included index test, reference standard, study design, setting, study population, number of participants, prevalence, country and data on translation. In case of translation of the criteria, we examined if the translation was verified

Table 3 The Quality Assessment of Diagnostic Accuracy tool (QUADAS) by Whiting *et al.* (2004)[12]

Item 1	Was the spectrum of patients representative of the patients who will receive the test in practice?
Item 2	Were the selection criteria clearly described?
Item 3	Is the reference standard likely to classify the target condition correctly?
Item 4	Is the period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two sets?
Item 5	Did the whole sample or a random selection of the sample, receive verification using a reference standard diagnosis?
Item 6	Did patients receive the same reference standard regardless of the index test result?
Item 7	Was the reference standard independent of the index test? (i.e. the index test did not form part of the reference standard?)
Item 8	Was the execution of the index test described in sufficient detail to permit replication of the test?
Item 9	Was the execution of the reference standard described in sufficient detail to permit its replication?
Item 10	Were the index test results interpreted without knowledge of the results of the reference standard?
Item 11	Were the reference standard results interpreted without knowledge of the results of the index test?
Item 12	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?
Item 13	Were uninterpretable/intermediate test results reported?
Item 14	Were withdrawals from the study explained?

by back translation. For studies carried out in nonnative-English speaking countries that did not describe their translations, we scored the translation as 'not reported'. Data with respect to the outcome measurements (sensitivity, specificity, positive and negative predictive value) were extracted. All calculations were verified by recalculation.

Results

Results of the literature search

Figure 1 summarizes the selection process for studies on diagnostic criteria for AD. An initial search retrieved 925 articles. After screening titles and abstracts for eligibility, 47 articles were selected. The excluded articles were primarily clinical trials in which diagnostic criteria were used to define AD. Reference searching yielded three additional publications.[7,17,18] The search for articles written by the designers of the diagnostic criteria yielded no additional articles. No additional articles were found on the Cochrane databases. Of the initial 47 selected articles 27 publications were excluded. Of these,

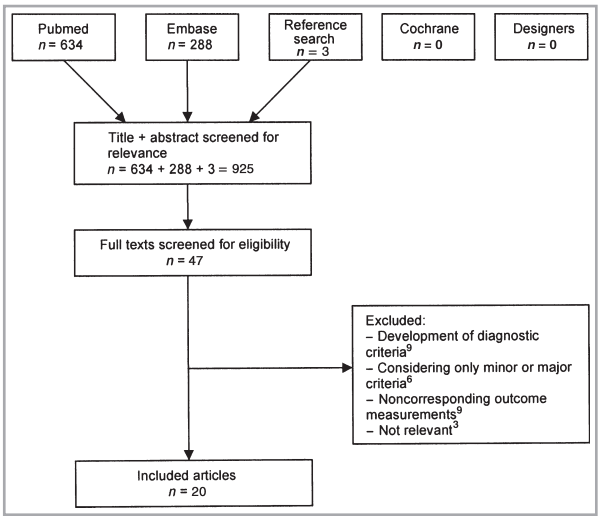

Fig 1. Flowchart summarizing the selection process for studies on diagnostic criteria for atopic dermatitis.

nine studies concerned the development of diagnostic criteria,[2,3,6,10,19–23] six studies considered only the minor or the major criteria of the Hanifin and Rajka criteria,[24–29] nine studies did not correspond with the outcome measurements[5,30–37] and three studies proved to be not relevant.[17,38,39]

Study description

Twenty articles, published between 1994 and 2007 were included in this review. Of the included articles, nine studies were hospital based,[3,8,40–46] 12 studies were population based[18,32,47–56] and one study was both hospital and population based.[57] Of all 27 validation studies, 18 were independent studies from centres with no conflict of interest. Reference standards used were the clinical diagnosis by a dermatologist, the Hanifin and Rajka diagnostic criteria, the U.K. diagnostic criteria and the Japanese Dermatology Association criteria. As a reference standard, the clinical diagnosis by a dermatologist was most frequently used. Study characteristics are shown in Table 4. Five studies were primarily prevalence studies that showed a subresearch on the validation of the diagnostic criteria used.[32,43,50,52,53] Question-only based formats used by parents on the U.K. diagnostic criteria were assessed in 10 studies.[43,47–50,52–56] The U.K. diagnostic criteria were validated in 17 studies,[18,40,42,44–57] Schultz-Larsen criteria and the Hanifin and Rajka diagnostic criteria were each validated in two studies,[40,43,45,57] and the Kang and Tian, the International Study of Asthma and Allergies in Childhood (ISAAC) and the criteria of Diepgen in one study each.[3,9,51] Only the U.K. diagnostic criteria are frequently validated both in hospital- and population-based settings. Williams *et al.* performed three validation studies in a hospital-based setting, all described in one article.[45] Seaki *et al.* performed two validation studies in one article[56] and Diepgen *et al.* validated three different sets of criteria.[3] The Millennium criteria, Danish Allergy Research Centre (DARC) criteria,

Table 4 Key characteristics of included validation studies

Criteria/reference	Reference standard	Study design	Setting	Age	Numbers (case/control)	AD prevalence	Country	Translation
U.K. diagnostic criteria								
Williams *et al.* (1994) I	Clinical diagnosis	Cross-sectional	Hos	≤ 10 years, > 10 years	200	–	England	–
Williams *et al.* (1994) II	Clinical diagnosis	Cross-sectional	Hos	< 16 years	114	–	England	–
Williams *et al.* (1994) III	Clinical diagnosis	Case–control	Hos	All ages	214 (116/98)	–	England	–
Williams *et al.* (1996)	Clinical diagnosis	Cross-sectional	Pop	3–11 years	695	8·5	England	–
Ortiz de Frutos *et al.* (1998)	Clinical diagnosis	Case–control	Hos	All ages	237 (102/135)	–	Spain	Spanish (verified)
Popescu *et al.* (1998)	Clinical diagnosis	Cross-sectional	Pop	6–12 years	1114	2·4	Romania	Romanian (verified)
Möhrenschlager *et al.* (1998)	Clinical diagnosis	Case–control	Pop	8–9 years	373 (43/330)	12·0	Germany	German (verified)
Firooz *et al.* (1999)	Clinical diagnosis	Cross-sectional	Hos	< 4, 4–10 years, > 10 years	416	–	Iran	Not reported
Marks *et al.* (1999)	Clinical diagnosis	Cross-sectional	Pop	4–18 years	2491	16·3	Australia	–
Gu *et al.*(2001)	H&R	Case–control	Hos	All ages	232 (111/121)	–	China	Not reported
Olesen *et al.* (2001)	Clinical diagnosis	Case–control	Pop	3–15 years	61 (31/30)	–	Denmark	Danish (verified)
Fleming *et al.* (2001)	U.K.	Case–control	Pop	1 year	118 (59/59)	–	Scotland	–
Ortiz *et al.* (2003)	Clinical diagnosis	Cross-sectional	Hos	3–17 years	874	7·1	Spain	Spanish (verified)
Girolomoni *et al.* (2003)	Clinical diagnosis	Cross-sectional	Pop	9 years	1331	5·8	Italy	Not reported
Hamada *et al.* (2005)	JDA	Cross-sectional	Pop	≤ 5 years	565	6·9	Japan	Japanese (not verified)
Haileamlak *et al.* (2005)	Clinical diagnosis	Case–control	Pop	1–5 years	(93/433) (140/592)	4·4	Southwest Ethiopia	Amharic (verified)
De *et al.* (2006)	Clinical diagnosis	Case–control	Hos	≤ 15 years	149 (101/48)	–	North India	Not reported
Chalmers *et al.* (2006)	Clinical diagnosis	Cross-sectional	Pop	3–11 years	3069	1·0	South Africa	Xhosa (verified)
Saeki *et al.* (2007)	JDA	Cross-sectional	Pop	6–7 years and 11–12 years	16 152, 3849	11·2, 10·4	Japan	Japanese (verified)
Hanifin and Rajka diagnostic criteria								
Williams *et al.* (1994) III	Clinical diagnosis	Case–control	Hos	All ages	214 (116/98)	–	England	–
De *et al.* (2006)	Clinical diagnosis	Case–control	Hos	≤ 15 years	149 (101/48)	–	North India	Not reported
Schulz-Larsen criteria								
Schultz Larsen *et al.* (1996)	H&R	Case–control	Hos/Pop	6–12 years	Unknown	15·6	North Europe	Not reported
Laughter *et al.* (2000)	Clinical diagnosis	Case–control	Hos	5–9 years	67 (18/49)	–	USA	–
Diepgen criteria								
Diepgen *et al.* (1996)	Clinical diagnosis	Case–control	Hos	10–55 years	329 (106/223)	–	Germany	Not reported
Kang and Tian criteria								
Gu *et al.*(2001)	H&R	Case–control	Hos	All ages	232 (111/121)	–	China	Not reported
ISAAC criteria								
Haileamlak *et al.* (2005)	Clinical diagnosis	Case–control	Pop	1–5 years	(93/433) (140/592)	4·4	Southwest Ethiopia	Amharic (verified)

AD, atopic dermatitis; U.K., U.K. diagnostic criteria; H&R, Hanifin and Rajka diagnostic criteria; JDA, Japanese Dermatology Association diagnostic criteria; Hos, hospital based; Pop, population based.

Table 5 Qualitative outcomes of validated diagnostic criteria by the Quality Assessment of Diagnostic Accuracy tool (QUADAS)

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 8	Item 9	Item 10	Item 11	Item 13	Item 14
U.K. diagnostic criteria												
Williams *et al.* (1994) I	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y
Williams *et al.* (1994) II	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	Y
Williams *et al.* (1994) III	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y
Williams *et al.* (1996)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Ortiz de Frutos *et al.* (1998)	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	N
Popescu *et al.* (1998)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y
Möhrenschlager *et al.* (1998)	Y	Y	Y	N	N	Y	Y	Y	Y	Y	N	N
Firooz *et al.* (1999)	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	N
Marks *et al.* (1999)	Y	Y	Y	Y	Y	Y	Y	Y	U	U	N	Y
Gu *et al.*(2001)	Y	Y	N	U	Y	Y	Y	N	U	U	N	N
Olesen *et al.* (2001)	N	Y	Y	U	Y	Y	Y	Y	U	U	N	N
Fleming *et al.* (2001)	Y	N	N	N	Y	Y	Y	Y	Y	Y	N	Y
Ortiz *et al.* (2003)	Y	Y	Y	Y	Y	Y	N	N	Y	Y	N	Y
Girolomoni *et al.* (2003)	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N
Hamada *et al.* (2005)	Y	Y	Y	U	Y	Y	Y	Y	N	U	Y	Y
Haileamlak *et al.* (2005)	Y	Y	Y	N	Y	Y	Y	Y	U	Y	N	N
De *et al.* (2006)	Y	Y	Y	U	Y	Y	Y	Y	U	U	N	N
Chalmers *et al.* (2006)	Y	Y	Y	N	Y	Y	Y	Y	U	Y	N	N
Saeki *et al.* (2007)	Y	N	N	Y	Y	Y	Y	Y	Y	Y	N	N
Hanifin and Rajka criteria												
Williams *et al.* (1994) III	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y
De *et al.* (2006)	Y	Y	Y	U	Y	Y	Y	Y	U	U	N	N
Schultz-Larsen criteria												
Schultz Larsen *et al.* (1996)	N	Y	N	U	U	U	Y	Y	U	U	N	N
Laughter *et al.* (2000)	N	Y	Y	N	N	Y	Y	Y	Y	Y	N	N
Diepgen criteria												
Diepgen *et al.* (1996)	Y	Y	Y	U	Y	Y	Y	N	U	U	Y	N
ISAAC criteria												
Haileamlak *et al.* (2005)	Y	Y	Y	N	Y	Y	Y	Y	U	Y	N	N
Kang and Tian criteria												
Gu *et al.*(2001)	Y	Y	N	U	Y	Y	Y	N	U	U	N	N

Y, yes; N, no; U, unclear.

Lillehammer criteria and Japanese Dermatology Association criteria were not validated.[2,5,7]

The methodological quality assessed with the QUADAS tool is illustrated in Table 5. Item 6 (Did patients receive the same reference standard regardless of the index test result?) was scored once with Unknown. Only one study scored a No for item 8 (Was the execution of the index test described in sufficient detail to permit replication of the test?). Item 9 (Was the execution of the reference standard described in sufficient detail to permit its replication?) was scored with a No in only three studies. Only three studies scored a Yes for item 13 (Were uninterpretable/intermediate test results reported?). Total percentages of Yes per item of rest of the items fluctuated from 41% to 86%.

Results of validation studies

Two studies validated the Hanifin and Rajka diagnostic criteria using the clinical diagnosis as reference standard. Their sensitivity ranged from 87·9% to 96·0%, specificity from 77·6% to 93·8%.[45,57] With respect to the hospital-based studies validating the U.K. diagnostic criteria the sensitivity ranged from 10% to 95·5% and specificity from 90·4% to 98·3%.[8,41,42,44–46] The hospital-based study by Firooz *et al.* showed a remarkably low sensitivity of 10%.[41] For studies validating the U.K. diagnostic criteria in a population-based setting sensitivity ranged from 42·8% to 100% and specificity from 89·3% to 99·1%.[18,47–56] Besides the corresponding specificity, four population-based studies showed a relatively low sensitivity. Reference standards used in these studies were: the clinical diagnosis in 15 studies, the Japanese Dermatology Association criteria in two studies and the Hanifin and Rajka diagnostic criteria in one study. When the Schultz-Larsen diagnostic criteria were evaluated for hospital- and population-based settings in two studies, the sensitivity and specificity ranged from 88% to 94·4% and from 77·6% to 95·9%, respectively.[43,57] Reference standards used in these studies were the clinical diagnosis and the Hanifin and Rajka diagnostic criteria. The Kang and Tian criteria resulted in 95·5% sensitivity and 100% specificity when compared with the Hanifin and Rajka diagnostic criteria.[9] The ISAAC questionnaire showed a PPV of 48·8% and a NPV of 91·1%.[51] Sensitivity and specificity of the three

validated sets of diagnostic criteria of Diepgen ranged from 83·0% to 87·7% and from 83·9% to 87·0%, respectively.[3] Detailed information on the outcomes is presented in Tables 6 and 7. An overview of the varying sensitivity and specificity of the various diagnostic criteria are presented in ROC plots (Figs 2 and 3). Due to the heterogeneity (e.g. variability in study populations and the settings) of the included studies, we considered a meta-analysis to generate summary estimates to be inappropriate.

Discussion

With this systematic review we have systematically collected and analysed validation studies of various sets of diagnostic criteria for AD. Overall, sensitivity and specificity ranged from 10% to 95% and from 77·6% to 100%, respectively, in hospital-based studies and from 42·8% to 100% and from 44·7% to 96·6%, respectively, in the population-based studies.

The first diagnostic criteria were introduced in 1980 by Hanifin and Rajka to delineate the clinical population in the absence of a clear definition of AD, mainly in order to conduct investigative studies.[4] The four major and 23 minor criteria were based on consensus between experienced dermatologists without objective clinical validation. As many criteria are involved and some clear definitions of items are missing, studies choose to exclude minor features (e.g. immediate skin-test reactivity, impaired cell-mediated immunity, keratoconus) from assessment.[40,42] In conclusion, the list is time consuming and not manageable. It is therefore unsuitable for population-based studies.[6] However, the Hanifin and Rajka criteria are often used in clinical trials, and the question remains whether or not they are applied in an appropriate way. Their suitability in hospital-based studies has not been guaranteed. No validation studies for these criteria were found between 1980 and 1993, and only two validations were published in 1994 and 2006. Although indicating varying specificities, the validity in these two hospital-based studies showed good outcomes.[40,45]

Later, in 1989, Kang and Tian developed a new set of criteria specially designed for the Chinese population.[9] By evaluating the Kang and Tian criteria with the Hanifin and Rajka diagnostic criteria as the reference standard, Gu *et al.* created a gold standard bias as the Kang and Tian are partly based on the Hanifin and Rajka criteria.[42]

ISAAC was founded to maximize the value of epidemiological research into AD and other allergic diseases, by facilitating international collaboration in 1991.[1] As the ISAAC questionnaire is primarily used to assess prevalences, studies validating the ISAAC questionnaire often use a prevalence estimate as the outcome measure and were therefore excluded. Only Haileamlak *et al.* conducted a case-controlled validation study.[51] Unfortunately, due to the sampling method of the cases and controls, specificity and sensitivity could not be calculated.

In 1992, a special task force was introduced by the board of the Japanese Dermatological Association to create new criteria on the diagnosis of AD by means of discussion: the Japanese Dermatological Association criteria.[10] The diagnostic criteria consist of only mandatory features: pruritus, typical morphology and chronic or chronically relapsing course. No studies were found that validated these criteria. However, these criteria were used as a reference standard in several studies.[52,56]

The Schultz-Larsen criteria were introduced by Schultz Larsen and Hanifin.[6] These consist of statements and questions, each of which is assigned a certain point value. After slightly modifying the Schultz-Larsen criteria, Laughter *et al.* compared those criteria with the clinical diagnosis in a case–control study.[43] They excluded 10 patients (13·2%) with possible AD and therefore the results might be too optimistic. Kuhnyar *et al.* mentioned that they validated these criteria, but did not publish data.[32] In 1994, the Lillehammer criteria were proposed by Schultz Larsen, Diepgen and Svensson.[7] No studies were found validating or using these criteria.

Around the same time, Diepgen developed and validated another three lists of diagnostic criteria: an objective model (without subjective features), a simplistic model (without laboratory measures and subjective features) and a model without constraints.[3] These lists showed corresponding results in a validation study, but none of the lists were subsequently used in published studies as far as we have been able to detect in the literature.

In 1997, the U.K. diagnostic criteria were introduced by Williams *et al.* as a refinement of Hanifin and Rajka's diagnostic criteria for AD.[8] These criteria consist of one mandatory and five major criteria. The criteria are all noninvasive and were designed for clinical and epidemiological studies as illustrated on http://www.nottingham.ac.uk/dermatology/eczema/section5-1.html (accessed 16 November 2007). A slight modification of the criteria is needed when infants are assessed. Although fulfilment of the criteria by itch plus three criteria is recommended, studies also validated cut-off points of two, four or five criteria.[40,41,44,45,48,55] Seven of the 19 U.K. validation studies were independent studies from centres with no conflicts of interest. Unlike the five other hospital-based studies with corresponding validity, the independent Iranian study of Firooz *et al.* showed a remarkably low sensitivity (10%).[41] This might be due to international differences in clinical phenotype, environmental factors and observation bias.[58] Although specificity showed uniformity, sensitivity fluctuated in the population-based studies. The lower sensitivities found by Hamada *et al.* and Saeki *et al.* might be due to some incomprehensibility in the Japanese translation and to insufficient parent cooperation.[52,56] Heterogeneity of diverse cultural, socioeconomic and language settings might explain the low sensitivity of 43·7% shown by Chalmers *et al.*[48] In addition, questions about personal or familial atopy may result in poorer performance.[53]

The Danish Allergy Research Centre (DARC) criteria are primarily used to diagnose AD in infants and were specially developed for the study of Johnke *et al.* in 2005.[5] They compared the U.K., Hanifin and Rajka, Schultz-Larsen and DARC

Table 6 Results for hospital-based studies

	Sensitivity		Specificity		PPV		NPV	
Criteria/reference	Absolute numbers	%	Absolute numbers	%	Absolute numbers	%	Absolute numbers	%
U.K. diagnostic criteria								
Williams *et al.* (1994) I	25/36	69·4	157/164	95·7	25/32[b]	78·1	157/168[b]	93·4
Williams *et al.* (1994) II	33/39	84·6	72/75	96·0	33/36[b]	91·7[b]	72/78[b]	92·3[b]
Williams *et al.* (1994) III	102/116	87·9	91/98	92·8	102/109[b]	93·6[b]	91/105[b]	86·7[b]
Ortiz de Frutos *et al.* (1998)		76·5		90·4		85·7		83·6
Firooz *et al.* (1999)	6/60[a]	10·0	350/365	98·3	6/21[b]	28·6[b]	350/404[b]	86·6[b]
Gu *et al.* (2001)	106/111	95·5	118/121	97·5	106/109	97·3	118/123	95·9
Ortiz *et al.* (2003)		63·6		96·7		63·6		96·7
De *et al.* (2006)	87/101	86·1	46/48	95·8	87/89	97·8	46/60	76·7
Hanifin and Rajka diagnostic criteria								
Williams *et al.* (1994) III	108/116	93·1	76/98	77·6	108/130[b]	83·1[b]	76/84[b]	90·5[b]
De *et al.* (2006)	97/101	96·0	45/48	93·8	97/100	97	45/49	91·8
Schulz-Larsen criteria								
Schultz Larsen *et al.* (1996)		88		89				
Laughter *et al.* (2000)	17/18 SLI	94·4	38/49	77·6	17/28	60·7	38/39	97·4
	16/18 SLII	88·9	47/49	95·9	16/18	88·9	47/49	95·9
Diepgen criteria								
Diepgen *et al.* (1996)		87·7 DgI		83·9				
		83·0 DgII		87·0				
		87·7 DgIII		84·3				
Kang and Tian diagnostic criteria								
Gu *et al.* (2001)	106/111	95·5	121/121	100	106/106	100	121/126	96·0

[a]Adjusted for 1-year period prevalence; [b]number calculated. If the studies subdivided their population by age, only the total results of the whole group were described. PPV, positive predictive value; NPV, negative predictive value; SLI, criteria Schultz-Larsen sumscore 50+; SLII, Schultz-Larsen criteria sumscore 80+; DgI, criteria of Diepgen without restrictions; DgII, criteria of Diepgen simplicity model; DgIII, criteria of Diepgen objective model.

Table 7 Results for population-based studies

Criteria/reference	Sensitivity		Specificity		PPV		NPV	
	Absolute numbers	%	Absolute numbers	%	Absolute numbers	%	Absolute numbers	%
U.K. diagnostic criteria								
Williams *et al.* (1996)	41/59	69·5	590/636	92·8	41/87	47·1	590/608	97·0
	70/88	79·5[a]	590/607	97·2[a]	70/87	80·0[a]	590/608	97·0[a]
Popescu *et al.* (1998)	20/27	74·0	1075/1087	98·9	20/32	62·5	1075/1082	99·3
Möhrenschlager *et al.* (1998)	38/43	88·4	326/330	98·8	38/42[c]	90·5[c]	326/331[c]	98·5[c]
Marks *et al.* (1999)	177/414	42·8[c]	1974/2077	95·0[c]	177/280	63·6[c]	1974/2211	89·3[c]
Olesen *et al.* (2001)	28/31	90·3[b]	29/30	96·7[b]	28/29[c]	96·6[b,c]	29/32[c]	90·6[b,c]
Fleming *et al.* (2001)	49/49[c]	100[c]	55/59[c]	93·2[c]	49/53[c]	92·4[c]	55/55[c]	100[c]
Girolomoni *et al.* (2003)	66/85[c]	77·6	1235/1246[c]	99·1	66/77[c]	85·7	1235/1254[c]	98·5
Hamada *et al.* (2005)	23/39	59·0	498/526	94·7	23/51	45·1[c]	498/514	96·9[c]
Haileamlak *et al.* (2005)					50/90	55·5	393/436	90·1
Chalmers *et al.* (2006)		43·7[a]		97·9[a]		18·4[a]		99·4[a]
Saeki *et al.* I (2007)	1250/1742	71·8	12866/14410	89·3	1250/2794	44·7	12866/13358	96·3
Saeki *et al.* II (2007)	236/401	58·9	3290/3448	95·4	236/394	59·9	3290/3455	95·2
	292/401	72·8[a]	3148/3448	91·3[a]	292/592	49·3[a]	3148/3257	96·6[a]
Schultz-Larsen criteria								
Schultz Larsen *et al.* (1996)		88		89				
ISAAC criteria								
Haileamlak *et al.* (2005)					62/127	48·8	388/426	91·1

[a]Adjusted for 1-year period prevalence; [b]based on lifetime prevalence; [c]numbers were calculated. PPV, positive predictive value; NPV, negative predictive value.

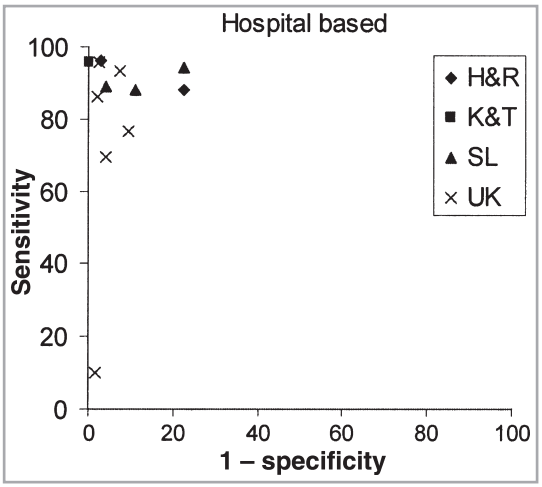

Fig 2. Receiver operating characteristics plot for hospital-based studies. H&R, Hanifin and Rajka diagnostic criteria; K&T, Kang & Tian diagnostic criteria; SL, Schultz-Larsen criteria; U.K., U.K. diagnostic criteria.

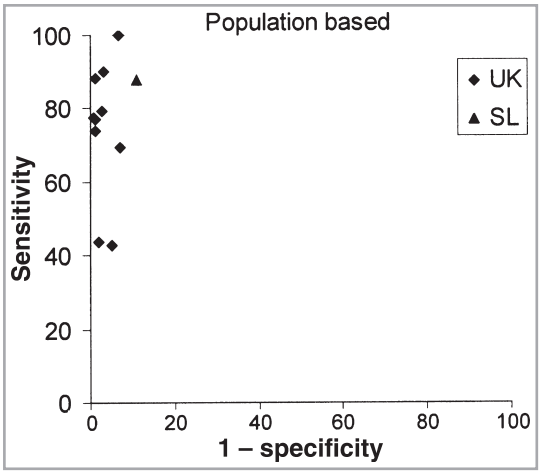

Fig 3. Receiver operating characteristics plot for population-based studies. U.K., U.K. diagnostic criteria; SL, Schultz-Larsen criteria.

diagnostic criteria with a prevalence estimate, without the use of a reference standard. No further evidence was found on the validity of the DARC criteria.

Exceptional are the Millennium criteria, proposed by Bos *et al.* in 1998.[2] In this list the presence of allergen-specific IgE is mandatory for the diagnosis of AD. These criteria were developed in response to the new knowledge about the pathogenesis of atopy in which the presence of allergen-specific IgE is essential. To satisfy the Millennium criteria, the mandatory criterion and two of the three principal criteria have to be fulfilled. No study has yet validated these criteria.

We collected all relevant articles concerning the validation of diagnostic criteria by an extensive systematic search. However, studies presented as prevalence studies might have involved a validation substudy, which was not reported in the abstract. If present, those studies might have been missed. Single-used, nonvalidated personal definitions of AD were

applied in prevalence studies, which were not taken into account. Of the validation studies included, only the U.K. diagnostic criteria were tested for repeatability. Assuming that a high accuracy corresponds with a good repeatability, evaluation of repeatability was not taken into account.

There are at least eight possible explanations why the results of the validation studies show considerable variability. (i) Differences in study characteristics lead to inconsistent study outcomes. In particular, differences in age might be an important issue, as the study designs varied in age groups. The study populations varied in culture, skin type and the settings of the studies range from urban to rural environments and from industrial to developing countries. (ii) Different reference standards were used. In general, the clinical diagnosis made by an experienced dermatologist is considered gold standard. However, to use the clinical diagnosis as a gold standard might be a point of discussion, as uniformity is not guaranteed in such a diagnostic process of AD. If the clinical diagnosis relies on one person, the validation study might be affected. With a panel of experts diagnosing AD, the problem could be partly solved. (iii) Most studies used point prevalence estimates to establish the diagnosis of AD, while others used a 1-year period or lifetime prevalence estimate. Studies using a 1-year period or a lifetime to validate diagnostic criteria showed more optimistic outcomes as illustrated by Williams *et al.*[55] and Saeki *et al.*[56] Applying diagnostic criteria over a 1-year period or a lifetime might have reduced false negatives due to decreased disease activity during a single examination. (iv) As scabies mimics the symptoms of AD, false positives in scabies endemic areas were numerous. (v) Some studies employed question-only-based formats on the U.K. diagnostic criteria.[4,47,51–53] By excluding the criterion of 'visible flexural dermatitis', the results can be questioned. (vi) Diagnostic criteria were often translated to conduct studies in nonnative-English speaking countries.[18,44,46–48,51,52,54,56] Due to these translations, inconsistencies might have evolved. To ensure nothing is lost in translation, translated diagnostic criteria must be retranslated into the original language. If not, results are less reliable. Six studies did not report any data on translation.[3,40–42,50,57] In addition to these translation issues, cultural issues such as the interpretation of pruritus may be important to explain differences in validity. (vii) Diagnostic criteria did not perform well on PPV, when applied in regions where the prevalence of AD is low. (viii) The studies investigated show differing methodological strengths as illustrated by the QUADAS tool. Interpreting methodological quality by using the QUADAS tool raised several issues. Although selection criteria, reference standards and index tests were well reported in most studies, withdrawals and intermediate results were not commonly stated. The period between the reference standard and the index test was often not reported or was considered too long to be reasonably sure that the course of AD did not fluctuate. Approximately half the studies showed a lack of blinding. As the current diagnostic criteria are based on clinical experience and the gold standard is the clinical diagnosis, the results of the index test are used in establishing the final diagnosis. With this, an incorporation bias is inevitable, which may lead to overestimation of sensitivity and specificity.[59,60]

With regard to all the included validation studies, the U.K. diagnostic criteria have been validated the most, both in hospital- and in population-based settings. Unlike the other criteria, this scientifically derived, minimum list of criteria has been shown to be applicable and repeatable across all ages and in a wide range of ethnic groups. However, the Hanifin and Rajka diagnostic criteria are most mentioned in investigational studies, but they have not been investigated enough to consider them applicable for epidemiological as well as clinical trials. Unlike the Schulz-Larzen, Diepgen, Kang and Tian and ISAAC criteria, which have been validated only once or twice, other existing criteria such as the Lillehammer, Japanese Dermatology Association, Millennium and DARC have not yet been validated. In addition, independence of validation studies is an important issue. Only the U.K. criteria have been validated more than once, independently, without conflict of interest. Validation studies of the other diagnostic criteria should be performed independently to give a reliable value judgment.

Besides the various sets of criteria that have been proposed, the nomenclature of AD has been changed and updated over the years.[61–63] There is an increasing need for consensus in nomenclature and reconsideration of the diagnostic criteria of AD.

In conclusion, in this systematic review, six validated sets of diagnostic criteria for AD and a total of 27 validation studies were found. In the included studies, the methodological quality varied substantially. For future validation studies improvement of the methodological design is recommended following a clear guideline such as QUADAS. As the most extensively validated are the U.K. diagnostic criteria, this set of criteria for AD should be recommended in future intervention studies. However, the ideal set of diagnostic criteria still has to be established. In addition, uniformity in nomenclature and in up-to-date, well-validated, applicable diagnostic criteria is needed to improve future intervention studies and to compare study results.

Acknowledgments

We thank C.E.J.M. Limpens, PhD, for her assistance in the literature search strategy.

References

1 Asher MI, Keil U, Anderson HR *et al.* International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. *Eur Respir J* 1995; **8**:483–91.
2 Bos JD, Van Leent EJ, Sillevis Smitt JH. The millennium criteria for the diagnosis of atopic dermatitis. *Exp Dermatol* 1998; **7**:132–8.
3 Diepgen TL, Sauerbrei W, Fartasch M. Development and validation of diagnostic scores for atopic dermatitis incorporating criteria of data quality and practical usefulness. *J Clin Epidemiol* 1996; **49**:1031–8.
4 Hanifin JM, Rajka G. Diagnostic features of atopic dermatitis. *Acta Derm Venereol Suppl (Stockh)* 1980; **92**:44–7.

5 Johnke H, Vach W, Norberg LA *et al.* A comparison between criteria for diagnosing atopic eczema in infants. *Br J Dermatol* 2005; **153**:352–8.
6 Schultz Larsen F, Hanifin JM. Secular change in the occurrence of atopic dermatitis. *Acta Derm Venereol Suppl (Stockh)* 1992; **176**:7–12.
7 Schultz Larsen F, Diepgen T, Svensson A. Clinical criteria in diagnosing atopic dermatitis: the Lillehammer criteria 1994. *Acta Derm Venereol Suppl (Stockh)* 1996; **96**:115–19.
8 Williams HC, Burney PG, Hay RJ *et al.* The U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis. I. Derivation of a minimum set of discriminators for atopic dermatitis. *Br J Dermatol* 1994; **131**:383–96.
9 Kang KF, Tian RM. Criteria for atopic dermatitis in a Chinese population. *Acta Derm Venereol Suppl (Stockh)* 1989; **144**:26–7.
10 Tagami H. Japanese Dermatology Association criteria for the diagnosis of atopic dermatitis. *J Dermatol* 1995; **22**:966–7.
11 Roguedas AM, Machet L, Fontes V *et al.* Atopic dermatitis: which are the diagnostic criteria used in medical literature? [In French] *Ann Dermatol Venereol* 2004; **131**:161–4.
12 Whiting P, Rutjes AW, Reitsma JB *et al.* The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; **3**:25.
13 Whiting P, Rutjes AW, Dinnes J *et al.* Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess* 2004; **8**:iii, 1–234.
14 Whiting PF, Weswood ME, Rutjes AW *et al.* Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol* 2006; **6**:9.
15 Juni P, Witschi A, Bloch R *et al.* The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; **282**:1054–60.
16 Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Med Res Methodol* 2005; **5**:19.
17 Foley P, Zuo Y, Plunkett A *et al.* The frequency of common skin conditions in preschool-age children in Australia: atopic dermatitis. *Arch Dermatol* 2001; **137**:293–300.
18 Möhrenschlager M, Schäfer H, Williams H *et al.* Übersetzung and Validiering der britischen Diagnosekriterien des atopischen Ekzem bei acht- und neunjärigen Schulkindren in Ausburg 1998. *Allergo J* 1998; **7**:373–7.
19 Kang YC, Lee S, Ahn SK *et al.* Clinical manifestations of hand eczema compared by etiologic classification and irritation reactivity to SLS. *J Dermatol* 2002; **29**:477–83.
20 Svensson A, Edman B, Moller H. A diagnostic tool for atopic dermatitis based on clinical criteria. *Acta Derm Venereol Suppl (Stockh)* 1985; **114**:33–40.
21 Wisuthsarewong W, Viravan S. Diagnostic criteria for atopic dermatitis in Thai children. *J Med Assoc Thai* 2004; **87**:1496–500.
22 Dotterud LK, Falk ES. Evaluation of a self-administered questionnaire on atopic diseases: discrepancy between self-reported symptoms and objective signs. *Eur J Public Health* 2000; **10**:105–7.
23 Kang KF, Tian RM. Atopic dermatitis. An evaluation of clinical and laboratory findings. *Int J Dermatol* 1987; **26**:27–32.
24 Bohme M, Svensson A, Kull I *et al.* Hanifin's and Rajka's minor criteria for atopic dermatitis: which do 2-year-olds exhibit? *J Am Acad Dermatol* 2000; **43**:785–92.
25 Chan HH, Pei A, Van KC *et al.* Validation of the Chinese translated version of ISAAC core questions for atopic eczema. *Clin Exp Allergy* 2001; **31**:903–7.
26 Kanwar AJ, Dhar S. How specific are major criteria for the diagnosis of atopic dermatitis? *Dermatology* 1994; **189**:102.
27 Mevorah B, Frenk E, Wietlisbach V *et al.* Minor clinical features of atopic dermatitis. Evaluation of their diagnostic significance. *Dermatologica* 1988; **177**:360–4.
28 Nagaraja, Kanwar AJ, Dhar S *et al.* Frequency and significance of minor clinical features in various age-related subgroups of atopic dermatitis in children. *Pediatr Dermatol* 1996; **13**:10–13.
29 Rudzki E, Samochocki Z, Rebandel P *et al.* Frequency and significance of the major and minor features of Hanifin and Rajka among patients with atopic dermatitis. *Dermatology* 1994; **189**:41–6.
30 Diepgen TL, Fartasch M. Statistical analysis of diagnostic criteria of atopic eczema. *Allergologie* 1991; **14**:301–6.
31 Kramer U, Schafer T, Behrendt H *et al.* The influence of cultural and educational factors on the validity of symptom and diagnosis questions for atopic eczema. *Br J Dermatol* 1998; **139**:1040–6.
32 Kuhnyar A, Egyud K, Szabo I *et al.* Prevalence of atopic dermatitis among children under 19 in an East-Hungarian agricultural county. *Clin Dev Immunol* 2006; **13**:395–9.
33 Ozkaya E. Adult-onset atopic dermatitis. *J Am Acad Dermatol* 2005; **52**:579–82.
34 Saeki H, Iizuka H, Mori Y *et al.* Prevalence of atopic dermatitis in Japanese elementary school children. *Br J Dermatol* 2005; **152**:110–14.
35 Saeki H, Tsunemi Y, Fujita H *et al.* Prevalence of atopic dermatitis determined by clinical examination in Japanese adults. *J Dermatol* 2006; **33**:817–19.
36 Wordemann M, Polman K, Diaz RJ *et al.* The challenge of diagnosing atopic diseases: outcomes in Cuban children depend on definition and methodology. *Allergy* 2006; **61**:1125–31.
37 Yamada E, Vanna AT, Naspitz CK *et al.* International Study of Asthma and Allergies in Childhood (ISAAC): validation of the written questionnaire (eczema component) and prevalence of atopic eczema among Brazilian children. *J Investig Allergol Clin Immunol* 2002; **12**:34–41.
38 Benn CS, Benfeldt E, Andersen PK *et al.* Atopic dermatitis in young children: diagnostic criteria for use in epidemiological studies based on telephone interviews. *Acta Derm Venereol* 2003; **83**:347–50.
39 Muto T, Hsieh SD, Sakurai Y *et al.* Prevalence of atopic dermatitis in Japanese adults. *Br J Dermatol* 2003; **148**:117–21.
40 De D, Kanwar AJ, Handa S. Comparative efficacy of Hanifin and Rajka's criteria and the UK working party's diagnostic criteria in diagnosis of atopic dermatitis in a hospital setting in North India. *J Eur Acad Dermatol Venereol* 2006; **20**:853–9.
41 Firooz A, Davoudi SM, Farahmand AN *et al.* Validation of the diagnostic criteria for atopic dermatitis. *Arch Dermatol* 1999; **135**:514–16.
42 Gu H, Chen XS, Chen K *et al.* Evaluation of diagnostic criteria for atopic dermatitis: validity of the criteria of Williams *et al.* in a hospital-based setting. *Br J Dermatol* 2001; **145**:428–33.
43 Laughter D, Istvan JA, Tofte SJ *et al.* The prevalence of atopic dermatitis in Oregon school children. *J Am Acad Dermatol* 2000; **43**:649–55.
44 Ortiz de Frutos FJ, Guerra TA, Gomez de la CA *et al.* Validation of the Spanish version of the diagnostic questionnaire of the Working Group on Atopic Dermatitis of the United Kingdom. [In Spanish] *Rev Clin Esp* 1998; **198**:424–8.
45 Williams HC, Burney PG, Pembroke AC *et al.* The U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis. III. Independent hospital validation. *Br J Dermatol* 1994; **131**:406–16.
46 Javier Ortiz F, Guerra A, Zarco C. Measuring the prevalence of atopic dermatitis in a school-age population in Madrid. *Actas Dermosifiliogr* 2003; **94**:458–63.
47 Braae Olesen A, Bang K, Juul S, Thestrup-Pedersen K. Development and validation of a questionnaire for diagnosing atopic dermatitis. *Acta Derm Venereol* 2001; **81**:277–81.

48 Chalmers DA, Todd G, Saxe N *et al.* Validation of the U.K. Working Party diagnostic criteria for atopic eczema in a Xhosa-speaking African population. *Br J Dermatol* 2007; **156**:111–16.
49 Fleming S, Bodner C, Devereux G *et al.* An application of the United Kingdom Working Party diagnostic criteria for atopic dermatitis in Scottish infants. *J Invest Dermatol* 2001; **117**:1526–30.
50 Girolomoni G, Abeni D, Masini C *et al.* The epidemiology of atopic dermatitis in Italian schoolchildren. *Allergy* 2003; **58**:420–5.
51 Haileamlak A, Lewis SA, Britton J *et al.* Validation of the International Study of Asthma and Allergies in Children (ISAAC) and U.K. criteria for atopic eczema in Ethiopian children. *Br J Dermatol* 2005; **152**:735–41.
52 Hamada M, Furusyo N, Urabe K *et al.* Prevalence of atopic dermatitis and serum IgE values in nursery school children in Ishigaki Island, Okinawa, Japan. *J Dermatol* 2005; **32**:248–55.
53 Marks R, Kilkenny M, Plunkett A *et al.* The prevalence of common skin conditions in Australian school students: 2. Atopic dermatitis. *Br J Dermatol* 1999; **140**:468–73.
54 Popescu CM, Popescu R, Williams H *et al.* Community validation of the United Kingdom diagnostic criteria for atopic dermatitis in Romanian school children. *Br J Dermatol* 1998; **138**:436–42.
55 Williams HC, Burney PG, Pembroke AC *et al.* Validation of the U.K. diagnostic criteria for atopic dermatitis in a population setting. U.K. Diagnostic Criteria for Atopic Dermatitis Working Party. *Br J Dermatol* 1996; **135**:12–17.
56 Saeki H, Iizuka H, Mori Y *et al.* Community validation of the U.K. diagnostic criteria for atopic dermatitis in Japanese elementary schoolchildren. *J Dermatol Sci* 2007; **47**:227–31.
57 Schultz Larsen F, Diepgen T, Svensson A. The occurrence of atopic dermatitis in north Europe: an international questionnaire study. *J Am Acad Dermatol* 1996; **34**:760–4.
58 Williams HC. Diagnostic criteria for atopic dermatitis: where do we go from here? *Arch Dermatol* 1999; **135**:583–6.
59 Rutjes AW, Reitsma JB, Di Nisio M *et al.* Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006; **174**:469–76.
60 Whiting P, Rutjes AW, Reitsma JB *et al.* Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004; **140**:189–202.
61 Johansson SG, Hourihane JO, Bousquet J *et al.* A revised nomenclature for allergy. An EAACI position statement from the EAACI nomenclature task force. *Allergy* 2001; **56**:813–24.
62 Johansson SG, Bieber T, Dahl R *et al.* Revised nomenclature for allergy for global use: Report of the Nomenclature Review Committee of the World Allergy Organization, October 2003. *J Allergy Clin Immunol* 2004; **113**:832–6.
63 Bos JD. Atopiform dermatitis. *Br J Dermatol* 2002; **147**:426–9.

