

ORIGINAL ARTICLE

Validation and refinement of the Millennium Criteria for atopic dermatitis

Mandy E. SCHRAM,¹ Mariska M.G. LEEFLANG,² Jan Paul S. DEN OTTOLANDER,¹ Phyllis I. SPULS,¹ Jan D. BOS¹

Departments of ¹Dermatology and ²Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands

ABSTRACT

There is no gold standard for a definite diagnosis of atopic dermatitis. For the time being, several lists of diagnostic criteria have been proposed, some of them in actual use. The Millennium Criteria have been proposed to diagnose atopic dermatitis and to differentiate it from atopiform dermatitis. Our aim was to further refine the Millennium Criteria into a manageable set that can differentiate between atopic and atopiform dermatitis and other entities. The hereby refined Millennium Criteria will be compared with the UK Working Party Criteria and the Hanifin & Rajka Criteria. Data of 210 included patients were used. After multiple logistic regression, a minimum set of five criteria was identified as best discriminators: (i) typical morphology; (ii) early age of onset; (iii) Dennie–Morgan fold; (iv) historical and (v) actual flexural involvement. The refined Millennium Criteria were constituted from these criteria. When comparing the different list for validity in diagnosing atopic dermatitis, the refined Millennium Criteria showed a sensitivity of 81.8% and a specificity of 98.8% compared to a sensitivity of 97.7% and specificity of 72.9% of the UK Criteria and a sensitivity of 100% and specificity of 48.8% of the Hanifin & Rajka Criteria. This refinement and validity study shows that the refined Millennium Criteria are a valid tool to diagnose atopic and atopiform dermatitis in a hospital-based setting and therefore could be incorporated in clinical practice and trials.

Key words: atopic dermatitis, diagnosis, Hanifin & Rajka Criteria, Millennium Criteria, UK Working Party Criteria.

INTRODUCTION

Atopic dermatitis (AD) is a pruritic and chronic inflammatory skin disease that commonly presents during early infancy and childhood, but can persist or start in adulthood.^{1,2} During the last decades its prevalence has nearly tripled, for which there is no firm explanation.^{3–5}

There is no definite laboratory marker that is able to assure a diagnosis of AD. To ensure future pathogenesis and intervention studies to be comparable, realizing homogeneity of patient populations and allowing proper diagnosing of AD and atopiform dermatitis (AFD), an accurate and widely

accepted diagnostic tool is needed. For that and other purposes, several lists of diagnostic criteria have been introduced, validated and used in population and hospital-based studies over the last years, of which the Hanifin & Rajka Criteria (H&RC) and the UK Working Party Diagnostic Criteria (UKC) are the most commonly used.^{6–10} A recent systematic review revealed that the H&RC contain some unclear and undefined criteria and have been insufficiently validated.¹¹ The UKC, which were derived from the H&RC, showed to be manageable and particularly useful in population-based studies. In some studies, however, the UKC lacked sensitivity.^{12,13}

Correspondence: Mandy E. Schram, M.D., Department of Dermatology, A0-252, Academic Medical Center, University of Amsterdam, PO Box 22700, 1100 DD Amsterdam, the Netherlands. Email: m.e.schram@amc.uva.nl

Received 11 August 2010; accepted 14 December 2010.

Diagnostic criteria should be in concordance with the definition of the diagnosed disease. With new insights on disease definitions, it is inevitable that diagnostic criteria are adapted in order to stay accurate. In 2003, the World Allergy Organization (WAO) founded the Nomenclature Review Committee to update the nomenclature proposed by the European Academy of Allergology and Clinical Immunology in 2001.^{14,15} In this latest nomenclature, the term “atopy” and thereby AD, is based on immunoglobulin (Ig)E sensitization and thus cannot be reached without an IgE-antibody determination or skin test. To specify patients with phenotypic AD but without presence of allergen-specific IgE, the term AFD was introduced both to distinguish it as a different entity and to avoid the confusion with true atopy.^{16,17}

In order to comply with the new definition of AD and aiming for hospital-based populations, Bos *et al.*¹⁸ developed the Millennium Criteria (MC) in which presence of allergen-specific IgE is a mandatory criterion, followed by three principal criteria of which at least two should be positive to diagnose patients with AD. For circumstantial evidence, 21 additional criteria were added. Recently, Brenninkmeijer *et al.*¹⁹ identified the most discriminatory principal and additional criteria in a case-control study. However, a cohort design, which resembles the use of diagnostic criteria in daily practice, is a more sound instrument to validate diagnostic criteria accurately.²⁰

Therefore, we aim to show whether the results found in the case-control study will persist in a cohort setting and to further refine the MC in a manageable set that can differentiate between AD, AFD and other entities. The individual criteria were tested for interobserver agreement. After a possible new set of criteria was identified, the ensuing refined MC were compared with the UKC and H&RC.

METHODS

Patients

From April to June 2009, all new consecutive patients attending the dermatological outpatient clinic of the Academic Medical Center, University of Amsterdam, were screened for eligibility. Staff members and residents were instructed to inform the investigators if eligible new patients were seen at the outpatient clinic. Included were patients in whom the diagnosis

of AD was considered in the differential diagnosis. After explanation of the study and written consent was obtained, patients were enrolled in the study. When patients were younger than 10 years, parents answered the questions. The study population varied in age, sex and skin type. The study was approved by the local Medical Ethical Committee (Institutional Review Board) and was performed in accordance with the Declaration of Helsinki.

Methods

In order to further validate and refine the MC, we used the clinical diagnosis as gold standard for the phenotypic diagnosis of AD, which is generally accepted in the absence of a good alternative. Clinical diagnosis was determined by a panel of two dermatologists of our department, who diagnosed the participants blinded for each other's assessment directly after the regular consultation or on scheduled visit and were unaware of the study design. In case of disagreement, another dermatologist was consulted.

On the same day, directly after the clinical diagnosis was made, all patients were examined by a study investigator according to the MC, UKC and H&RC using the 1-year period prevalence of symptoms and signs. Study investigators involved were two research nurses and two medical doctors, who were blinded for the diagnosis made by the dermatologists. They used a data-extraction form in which all the individual criteria of the three diagnostic tools were listed randomly. A proportion of the patients was assessed twice by two investigators who were blinded to each other in order to determine interobserver agreement of the individual criteria. Additionally, when the clinical diagnosis of AD was highly suspected by the study investigators, the Severity Scoring of Atopic Dermatitis Index (SCORAD) was also measured by them. Investigators participating in this study were instructed by the first author on how to use the MC, UKC, H&RC and SCORAD by means of an oral presentation and written details.

Diagnostic criteria

All the individual criteria of the MC, UKC and H&RC were assessed during the examination by the study investigators with some exceptions. “White dermographism” was excluded because it was difficult to

interpret the results, especially in dark skin, “cradle cap” was excluded because a substantial portion of the adult population could not recall this event and “keratoconus” and “anterior subcapsular cataract” were excluded because they were considered ophthalmic diagnoses. Presence of allergen-specific IgE was identified by using the Phadiatop test (Phadiatop, Uppsala, Sweden) or skin prick test, historically or actually. Children under the age of four were tested using the Phadiatop Infant test (Phadiatop), in which also common food allergens were included. These tests have shown to be reliable in testing presence of allergen-specific IgE.^{21–24}

The criterion “typical distribution” was considered present if it followed an age-related pattern.² Therefore, subdivision into infantile, juvenile and adult types was made. In infancy, the age-related pattern is that of a balaclava, which is predominantly characterized by eczematous lesions of the head and neck, but sparing of the periorbital and perioral regions.²⁵ In juvenile AD, lesions are localized on the flexor surfaces, particularly the antecubital and popliteal fossae. In adult AD, pronounced involvement of wrists and ankles is seen, but also lesions in the head and neck region are common again. “Typical morphology” also follows an age-related pattern. In infantile AD, lesions are characterized by erythematous papules, patched and vesicles, which may have an oozing appearance. In juvenile AD, the lesions become less oozing, but more lichenified and excoriated due to chronic rubbing and scratching. Pronounced lichenified plaques with scaling and prurigo papules are common features in adult AD. “Typical distribution” and “typical morphology” were assessed as individual criteria and as a coupled criterion.

Power analysis

Given a prevalence of AD in the cohort equal to 0.4 and using the large sample normal approximation, a sample size of 200 people was considered appropriate to produce a 95% confidence interval for: (i) a single sensitivity value that will extend 0.048 from the observed sensitivity for an expected sensitivity of 0.95 (hence, ranging 0.902–0.998); and (ii) a single specificity value that will extend 0.064 from the observed specificity for an expected specificity of 0.85 (hence, ranging 0.786–0.914).

Statistical analysis

Of all the individual criteria (i.e. the MC, UKC and H&RC), we calculated the interobserver agreement. Interobserver agreement was expressed as absolute observed agreement and as (unweighted) kappa statistics. Kappa was calculated as the difference between observed and expected agreement, divided by one minus the expected agreement.

We used multivariate logistic regression analysis to identify a minimum set of criteria that should be incorporated in the revised MC. As the main aim of the MC is to differentiate between AD, AFD and other diagnoses in clinical studies and hospital-based populations, we first developed a model that discriminates between phenotypic AD (AD or AFD) versus other diagnoses. The presence or absence of allergen-specific IgE was then used to further differentiate between AD and AFD.

Those criteria significantly associated with AD in a univariate analysis (P for entry: 0.05) and that fulfilled the criteria that were also used in the case-control study (diagnostic odds ratio [dOR] ≥ 3 and prevalence between 5% and 95%) were added to the multivariate model, using a forward selection.^{19,26} The accuracy of each individual criterion to diagnose phenotypic AD was expressed as sensitivity, specificity and dOR. To determine whether age would also contribute to a correct diagnosis, we added the variable “<18 years of age” to the starting set. If the -2 log likelihood significantly increased, forward regression was terminated. The goodness of fit was calculated using the Hosmer–Lemeshow test. To check whether the variable “typical morphology and distribution” should be divided into two variables or be kept as one single criterion, we repeated the forward selection process with a second starting set, equal to the first one but with “typical morphology” and “typical distribution” as two separate criteria. We did not include interaction terms in the model, as our aim was to find a minimum set of criteria that has a maximum diagnostic accuracy for diagnosing phenotypic AD in our patients and not to assess the causal relationship of the individual variables with AD.

To assess clinical relevance, the sensitivity, specificity and relative value were calculated for the combination of sets that were included in the final models. Relative value (also called Youden index) was calculated as sensitivity + specificity – 100%. After

identifying the set of criteria with the best diagnostic value for the phenotypic diagnosis of AD/AFD, we compared this new set, thus the refined MC (including and excluding the mandatory criterion for IgE sensitization) with the UKC and the H&RC to diagnose phenotypic AD, AD and AFD.

RESULTS

Patients

Of the approximately 650 new patients attending our outpatient clinic between April and June 2009, in 248 patients the diagnosis of AD was considered in the differential diagnosis. Of those patients, 30 refused to participate, seven patients failed to perform a Phadia-top and one patient was unable to answer the questions. In total, 210 patients were included in our study. Of the included patients, 44 were diagnosed as having AD, 19 AFD and 147 other diagnoses. Demographics of these groups are shown in Table 1. Of the patients having another diagnosis, 56 had other eczematous entities including dyshidrotic, seborrheic and nummular eczema, 18 had allergic reactions including allergic contact dermatitis, urticaria and toxicoderma, 16 had psoriasis, 13 had infectious skin diseases, 12 had pruritus, seven had acneiform dermatosis, five had lichen simplex chronicus and 22 of the patients suffered from other forms of dermatitis.

The dermatologists involved showed excellent agreement on establishing the clinical diagnosis. Disagreement about the presence or absence of phenotypic AD occurred in one case only (0.4% of total). In that case, a third dermatologist made the final diagnosis according to protocol.

Table 1. Demographics of included patients

	AD patients	AFD patients	Patients with other diagnosis
Number (%)	44 (21.0%)	19 (9.0%)	147 (70.0%)
Age (mean)	15.9 (SD 14.7)	12.9 (SD 15.7)	41.5 (SD 20.5)
Sex, <i>n</i> (%)			
Male	20 (45.5%)	10 (52.6%)	61 (41.5%)
Female	24 (54.5%)	9 (47.4%)	86 (58.5%)
Start age	4.8 (SD 10.5)	5.0 (SD 15.1)	34.3 (SD 22.5)
SCORAD, mean	26.2 (SD 20.2)	18.5 (SD 14.5)	–

AD, atopic dermatitis; AFD, atopiform dermatitis; SCORAD, Scoring Atopic Dermatitis scale; SD, standard deviation.

Refinement of the MC

Using univariate analysis, we identified 20 criteria with a dOR of 3 or more and prevalence between 5% and 95% (Table 2). All the criteria with a dOR 3 or more found in the previous case-control study also had a dOR of 3 or more in the present study, except for the criterion ichthyosis.¹⁹

The 20 identified criteria were used to form two start sets and analyzed in the multivariate regression model. The start set that separated “typical morphology” and “typical distribution” achieved better discrimination and was used for further analysis (data not shown). After forward selection, five criteria were identified: (i) early age of onset; (ii) Dennie–Morgan fold; (iii) history of flexural involvement; (iv) visible flexural eczema; and (v) typical morphology. “Typical morphology” was found to be of importance, which could indicate that the diagnostic value of the criterion “typical morphology and distribution” depends more on the typical morphology than on the typical distribution. The goodness of fit test of Hosmer and Lemeshow showed a χ^2 of 8.414, *df* = 6 and *P*-value of 0.209, which indicates that the found set of criteria has a good fit and is an accurate reflection of reality. The observed interobserver agreement of the selected criteria ranged 85–100%, corresponding with a kappa statistic of 57–100% (Fig. 1).

In our opinion, the criteria “typical morphology” should be a mandatory criterion for the diagnosis of AD/AFD and therefore had to be met by all cases. As typical morphology plus three of the four positive criteria achieved the best specificity in diagnosing the phenotypic AD and had a high relative value (RV), we have chosen to use this set in the refined MC (rMC).

With the present set of criteria it is now possible to determine if patients have phenotypic AD. To determine whether the patient suffers from AD or AFD, the presence of allergen-specific IgE is necessary. Patients with the presence of allergen-specific IgE have AD, patients without allergen-specific IgE suffer from AFD. Figure 2 presents a flowchart for the practical use of the rMC.

Comparison of the rMC with the UKC and H&RC against clinical diagnosis

When comparing the different lists with respect to diagnosing AD, the rMC had a sensitivity of 81.8% and a specificity of 98.8% (RV 0.81) (Table 3). The

Table 2. Prevalence rates, sensitivity and specificity rates, and overall accuracy (dOR) for individual criteria to predict phenotypic AD

	Total (n = 210)	AD (n = 44)	PhenAD (n = 63)	Other (n = 147)	Sens	Spec	dOR (95% CI)	P-value	Prev analysis	In analysis*
Presence of allergen-specific IgE	51%	100%	70%	43%	0.70	0.57	3.1 (1.6–5.8)	<0.001	NA	NA
Pruritus	87%	100%	98%	82%	0.98	0.18	14.0 (1.9–105)	0.011	–	+
Typical morphology and distribution	33%	89%	81%	12%	0.81	0.88	30.5 (13.7–67.7)	<0.001	+	+
Chronic or relapsing course	72%	98%	97%	72%	0.97	0.28	11.8 (2.8–50.5)	0.001	+	+
History of atopy	76%	98%	89%	70%	0.89	0.30	3.4 (1.4–8.1)	0.005	–	+
Xerosis	77%	93%	92%	70%	0.92	0.30	5.0 (1.9–13.2)	0.001	+	+
Ichthyosis	9%	16%	11%	8%	0.11	0.93	1.5 (0.57–4.2)	0.392	+	–
Early age of onset	35%	80%	79%	16%	0.79	0.84	19.7 (9.3–41.8)	<0.001	–	+
Skin infections	29%	46%	37%	25%	0.37	0.75	1.7 (0.91–3.2)	0.097	–	–
Tendency to non-specific hand/foot eczema	32%	32%	37%	30%	0.37	0.70	1.3 (0.72–2.5)	0.350	–	–
Nipple eczema	8%	16%	16%	5%	0.16	0.95	3.8 (1.4–10.4)	0.010	+	+
Cheilitis	38%	59%	56%	31%	0.56	0.69	2.8 (1.5–5.2)	0.001	–	–
Conjunctivitis	40%	50%	43%	39%	0.43	0.61	1.2 (0.65–2.2)	0.580	–	–
Dennie–Morgan fold	40%	84%	78%	24%	0.78	0.76	11.2 (5.5–22.7)	<0.001	+	+
Orbital darkening	28%	55%	49%	19%	0.49	0.81	4.1 (2.2–7.8)	<0.001	–	+
Facial pallor	51%	73%	67%	44%	0.67	0.56	2.5 (1.4–4.7)	0.003	–	–
Pityriasis alba	25%	41%	44%	17%	0.44	0.83	3.9 (2.0–7.5)	<0.001	+	+
Anterior neck folds	20%	30%	27%	16%	0.27	0.84	1.9 (0.93–3.8)	0.077	–	–
Itch when sweating	31%	52%	49%	22%	0.49	0.78	3.3 (1.8–6.3)	<0.001	+	+
Food intolerance	17%	41%	35%	9%	0.35	0.91	4.8 (2.2–10.2)	<0.001	–	+
Environment/emotion	47%	73%	68%	38%	0.68	0.62	3.5 (1.9–6.5)	<0.001	–	+
Intolerance to wool and solvents	43%	64%	54%	38%	0.54	0.62	1.9 (1.0–3.5)	0.034	–	–
Perifollicular accentuation	25%	34%	35%	21%	0.35	0.79	2.0 (1.0–3.9)	0.036	–	–
Palmar hyperlinearity	16%	18%	19%	15%	0.19	0.85	1.3 (0.62–2.9)	0.463	–	–
Keratosis pilaris	23%	48%	44%	14%	0.44	0.86	4.8 (2.4–9.5)	<0.001	–	+
Perleche	25%	41%	37%	20%	0.37	0.80	2.3 (1.2–4.5)	0.011	–	–
Auricular rhagades	20%	41%	37%	12%	0.37	0.88	4.1 (2.0–8.4)	<0.001	+	+
Hertoghe sign	8%	16%	14%	5%	0.14	0.95	2.9 (1.1–7.9)	0.038	–	–
Photophobia	27%	11%	13%	33%	0.13	0.67	0.29 (0.13–0.66)	0.003	–	+
History of flexural involvement	43%	98%	94%	21%	0.94	0.79	41.7 (15.4–113)	<0.001	–	+
Rash <2 years of age	21%	52%	51%	8%	0.51	0.92	11.6 (5.4–25.1)	<0.001	–	+
Visible flexural dermatitis	28%	75%	73%	8%	0.73	0.92	25.8 (11.7–56.8)	<0.001	–	+
Typical morphology	51%	96%	94%	33%	0.94	0.67	29.5 (10.1–85.9)	<0.001	–	+

AD, atopic dermatitis; dOR, diagnostic odds ratio; NA; not applicable; PhenAD, phenotypic AD; prev, previous analysis by Breninkmeijer *et al.*;¹⁹ sens, sensitivity; spec, specificity. *Odds ratio ≥ 3 .

sensitivity of the rMC was 83.3% in patients with a SCORAD of 14–27 and 100% in patients with a SCORAD above 27. The UKC showed a sensitivity of 97.7%, a specificity of 72.9% and a RV of 0.71 and the H&RC showed 100% sensitivity, 48.8% specificity and 0.49 on RV. When adding the mandatory criterion of presence of allergen-specific IgE to the UKC and the H&RC, the RV were 0.85 and 0.72, respectively, for diagnosing AD. This increase in RV was due to an increased specificity.

For diagnosing phenotypic AD, Table 3 shows that the rMC (without the criteria for allergen-specific IgE) had a higher specificity and RV (sensitivity 79.4%, specificity 98.0% and RV 0.77) than the UKC (sensitivity 92.1% specificity 79.6% and RV 0.72) and H&RC (sensitivity 96.8%, specificity 53.7% and RV 0.51), but a lower sensitivity.

For diagnosing AFD, the rMC showed a sensitivity of 73.7%, a specificity of 99.5% and a RV of 0.73. The UKC had a sensitivity of 78.9%, a specificity of

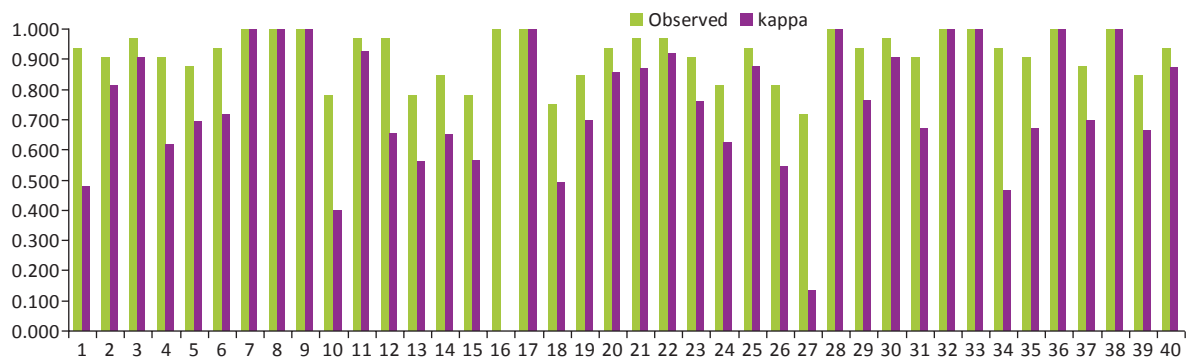
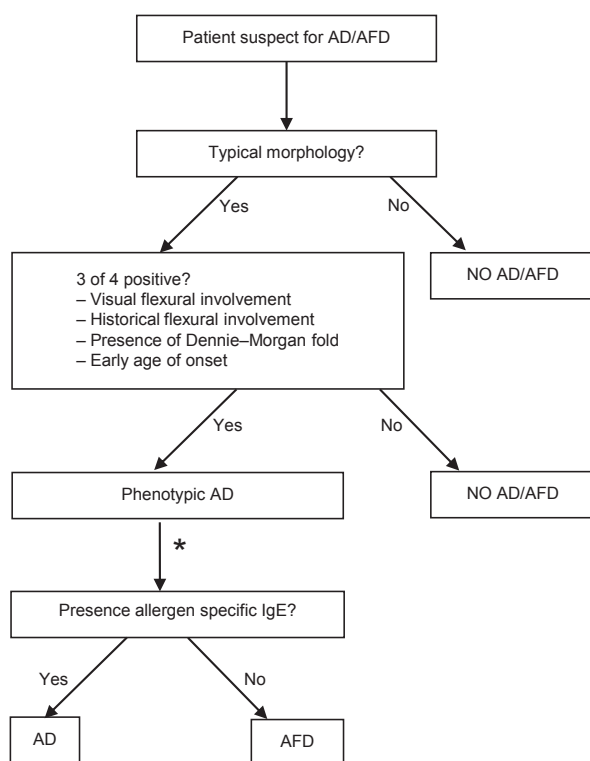


Figure 1. Interobserver agreement.



*This step might be postponed in infantile AD.

Figure 2. Refined Millennium Criteria; flowchart. AD, atopic dermatitis; AFD, atopiform dermatitis; IgE, immunoglobulin E.

95.3% and an RV of 0.74 and the H&RC 89.5%, 95.3% and 0.78, respectively.

DISCUSSION

Main findings

This validation study provides evidence for the use of the rMC to diagnose AD as well as AFD. Currently,

this is unique among the existing diagnostic criteria. Although the rMC were simplified from the initial MC, the rMC achieved good discrimination to differentiate between AD and AD-like dermatoses in a hospital-based setting. Diagnosing AFD was also accurately done. Thereby, the criteria included in the rMC scored well on interobserver agreement.

The rMC have a lower sensitivity (82% vs 98–100%) as compared to all other lists, but a higher specificity (99% vs 49–87%). The consequence of this is that the rMC may miss patients who do have AD, but that using these criteria, less patients will be assigned to have AD while in fact they have not. As the sensitivity of the rMC was high in patients with a SCORAD above 14, moderate to severe AD is unlikely to be misdiagnosed. Hereby, the chance to miss suitable patients for clinical trials is small. It can be considered to use the H&R Criteria to distinguish the presumed non-AD population in reverse to avoid false-negatives.

Whereas diagnostic tools for population-based prevalence research should ensure high sensitivity to prevent underregistration of a health-care problem, diagnostic tools aimed for hospital-based use and thus mostly used in clinical trials should ensure a high specificity to generate homogenous study populations.

Presence of allergen-specific IgE was added as a criterion that differentiates between AD and AFD. Because the Phadiatop and the skin prick test are considered reliable tests to measure atopy, these tests were chosen to detect allergen-specific IgE. However, they only test common allergens. The possibility remains that allergen-specific IgE against uncommon (or even as yet unidentified) allergens

Table 3. Sensitivity and specificity

	Phenotypic AD			AD			AFD		
	Sens %	Spec %	RV	Sens %	Spec %	RV	Sens %	Spec %	RV
rMC – IgE	79.4 (67.8–85.5)	98.0 (94.2–99.3)	0.77	NA	NA	NA	NA	NA	NA
UKC	92.1 (82.7–96.6)	79.6 (72.4–85.3)	0.72	97.7 (88.2–99.6)	72.9 (65.7–79.1)	0.71	NA	NA	NA
H&RC	96.8 (89.1–99.1)	53.7 (45.7–61.6)	0.51	100.0 (92.0–100)	48.8 (41.3–56)	0.49	NA	NA	NA
rMC	NA	NA	NA	81.8 (68.0–90.5)	98.8 (96.7–99.7)	0.81	73.7 (51.2–88.2)	99.5 (97.1–99.9)	0.73
UKC + IgE	NA	NA	NA	97.7 (88.2–99.6)	87.3 (81.4–91.6)	0.85	78.9 (56.7–91.5)	95.3 (91.3–97.5)	0.74
H&RC + IgE	NA	NA	NA	100.0 (92.0–100)	72.3 (65.7–79.2)	0.72	89.5 (68.6–97.1)	88.5 (83.2–92.3)	0.78

AD, atopic dermatitis; AFD, atopiform dermatitis; H&RC, Hanifin & Rajka Criteria; MC, Millennium Criteria; UKC, UK Working Party Criteria; rMC, refined Millennium Criteria; sens, sensitivity; spec, specificity; RV, relative value. +/- immunoglobulin (Ig)E: taking into account the criteria for the presence of allergen-specific IgE.

are missed and patients are thus misdiagnosed. Nonetheless, this will have minimal therapeutic consequences because avoiding common allergens is also not indicated for this patient group. Adding the mandatory criterion for presence of allergen-specific IgE to the UKC and the H&RC resulted in an increase in specificity while it had no effect on sensitivity. This indicates that presence of allergen-specific IgE is also an important criterion to rule out false-positives.

Strengths and weaknesses of the study

Important quality criteria for diagnostic studies are selection of patients, verification of the results and blinding. The standards of the QUADAS tool to control for methodological quality were used.^{27,28} We think that our reference standard/gold standard is the best available method to assess whether a person has AD or not, but we could not prevent incorporation bias. Incorporation bias occurs when the test under evaluation is part of the reference standard. In our case, the criteria overlap with the clinical diagnosis (gold standard) almost by definition. Overestimation of the sensitivity and specificity might be the result.^{29,30} However, because all evaluations of criteria for AD suffer from this same form of bias, we expect that the results of this study are still comparable with previous ones. Furthermore, we compared the accuracy of three different tools in one study, which will not be influenced by incorporation bias.

We did not limit ourselves to the major and minor criteria that were proposed in the former MC, but we

included all criteria of the three diagnostic criteria sets in order to find the best discriminatory ones and thus the best possible set. This resulted in incorporating two UK criteria (historical or visible flexural involvement) and one H&R criterion (early age of onset). When we individually assessed each feature of the criterion “typical morphology and distribution”, we found that “typical morphology” alone was more discriminative than in combination with distribution. Probably, adding the criteria for historical and visual flexural involvement to the set could be the explanation that excluding “typical distribution” from the criterion could be done without resulting in a lower validity.

Thereby, we used 1-year period to assess the validity of the MC, UKC and H&RC, to overcome misclassification when using a point-prevalence as AD is a fluctuating disease that might be temporarily absent during examination.

Because an invasive test is included in the rMC, it can be argued that the rMC is both inconvenient and time-consuming. But because these diagnostic criteria were developed for hospital-based populations, which are relatively small compared to populations in nationwide surveys, and the outcome of this test has implications for diagnosis, management and patient information, we reckon the application of this invasive test outweighs the disadvantages.

Unfortunately, ease of use of the rMC could not be assessed due to the study design.

In our analysis, we found that age of the subjects (<18 years) was not a major discriminatory factor for the risk of having phenotypic AD as it was not detected as such during the multivariate regression analysis. This can be explained by the fact that the used criteria set (including “early age of onset”) for the rMC overcomes the different age-related features of disease expression.

Because the population of patients aged less than 1 year was very small ($n = 3$), we are not confident about our results with respect to very young children. Also, it is questionable whether the Phadiatop test should be performed on those children as sensitization to allergens might not be detectable yet. Titers of allergen-specific IgE will rise with age. Besides aeroallergens, the Infant Phadiatop detects allergen-specific IgE against common food allergens. In our study, presence of allergen-specific IgE against common food allergens was also used to fulfill the criterion for presence of allergen-specific IgE. In a recently performed study on associated factors for AD in Japan, however, a clear relation between the presence of food allergy and AD was not established.³¹

For all these reasons, when handling infants, it might be considered to postpone the last step in the diagnostic work-up (thus performing the Phadiatop test) and rather diagnose patients as having phenotypic AD or not.

Clinical research and clinical practice implications

Diagnostic criteria can be used for different purposes. They can be aimed to provide a clear and accurate diagnosis in experimental research as well as in clinical practice, or they can be aimed to make an inventory on national or international prevalence of a disease. Each aim requires different features of a diagnostic tool. Whereas tools for population-based surveys have to be easily applicable, inexpensive and low in false-negatives to avoid underestimation, tools for diagnosing in clinical practice and trials should avoid false-positives and should therefore be very precise. For population-based surveys, the UKC are a suitable and validated tool. Although the rMC should be further validated, we consider them to be legitimate as diagnostic criteria for clinical practice and research. The rMC can assure homogenous (sub)populations, but can also make a clear distinc-

tion between AD and AFD which can be used for sub-analysis. This would be relevant for adding weight to supporting or refuting the hypothesis that AD and AFD are indeed separate entities.

ACKNOWLEDGMENTS

We thank Miranda Roskam for her assistance in designing the database and we thank Ahu Karatli for assistance with the study planning and performance.

REFERENCES

- 1 Bieber T. Atopic dermatitis. *N Engl J Med* 2008; **358**(14): 1483–1494.
- 2 Bos JD, Brenninkmeijer EE, Schram ME, Middelkamp-Hup MA, Spuls PI, Smitt JH. Atopic eczema or atopiform dermatitis. *Exp Dermatol* 2010; **22**: 19.
- 3 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**(6): 1526–1530.
- 4 Levy RM, Gelfand JM, Yan AC. The epidemiology of atopic dermatitis. *Clin Dermatol* 2003; **21**(2): 109–115.
- 5 Williams H, Flohr C. How epidemiology has challenged 3 prevailing concepts about atopic dermatitis. *J Allergy Clin Immunol* 2006; **118**(1): 209–213.
- 6 Hanifin JM, Rajka G. Diagnostic features of atopic dermatitis. *Acta Derm Venereol Suppl (Stockh)* 1980; **92**: 44–47.
- 7 Williams HC, Burney PG, Pembroke AC, Hay RJ. The U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis. III. Independent hospital validation. *Br J Dermatol* 1994; **131**(3): 406–416.
- 8 Williams HC, Burney PG, Strachan D, Hay RJ. The U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis. II. Observer variation of clinical diagnosis and signs of atopic dermatitis. *Br J Dermatol* 1994; **131**(3): 397–405.
- 9 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**(3): 383–396.
- 10 Williams HC, Burney PG, Pembroke AC, Hay RJ. 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**(1): 12–17.
- 11 Brenninkmeijer EE, Schram ME, Leeflang MM, Bos JD, Spuls PI. Diagnostic criteria for atopic dermatitis: a systematic review. *Br J Dermatol* 2008; **158**(4): 754–765.
- 12 Firooz A, Davoudi SM, Farahmand AN, Majdzadeh R, Kashani N, Dowlati Y. Validation of the diagnostic criteria

- for atopic dermatitis. *Arch Dermatol* 1999; **135**(5): 514–516.
- 13 Ortiz de Frutos FJ, Guerra TA, Gomez dIC, de la Cruz BJ, Alvarez Fernandez JG, de la Mano OD. [Validation of the Spanish version of the diagnostic questionnaire of the Working Group on Atopic Dermatitis of the United Kingdom]. *Rev Clin Esp* 1998; **198**(7): 424–428.
- 14 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**(9): 813–824.
- 15 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**(5): 832–836.
- 16 Bos JD. Atopiform dermatitis. *Br J Dermatol* 2002; **147**(3): 426–429.
- 17 Brenninkmeijer EE, Spuls PI, Legierse CM, Lindeboom R, Smitt JH, Bos JD. Clinical differences between atopic and atopiform dermatitis. *J Am Acad Dermatol* 2008; **58**(3): 407–414.
- 18 Bos JD, Van Leent EJ, Sillevs Smitt JH. The millennium criteria for the diagnosis of atopic dermatitis. *Exp Dermatol* 1998; **7**(4): 132–138.
- 19 Brenninkmeijer EE. *Atopic and Atopiform Dermatitis*. Academic Medical Center, Amsterdam, The Netherlands 2009.
- 20 Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Little Brown, Boston 1991.
- 21 Garcia-Marcos L, Sanchez-Solis M, Martinez-Torres AE, Lucas Moreno JM, Hernando Sv. Phadiatop compared to skin-prick test as a tool for diagnosing atopy in epidemiological studies in schoolchildren. *Pediatr Allergy Immunol* 2007; **18**(3): 240–244.
- 22 Gustafsson D, Danielsson D. In vitro diagnosis of atopic allergy in children. A comparison between total IgE, conventional RAST and a new multi RAST (Phadiatop). *Allergy* 1988; **43**(2): 105–108.
- 23 Merrett J, Merrett TG. Phadiatop – a novel IgE antibody screening test. *Clin Allergy* 1987; **17**(5): 409–416.
- 24 Vidal C, Gude F, Boquete O *et al.* Evaluation of the phadiatop test in the diagnosis of allergic sensitization in a general adult population. *J Investig Allergol Clin Immunol* 2005; **15**(2): 124–130.
- 25 Halkjaer LB, Loland L, Buchvald FF *et al.* Development of atopic dermatitis during the first 3 years of life: the Copenhagen prospective study on asthma in childhood cohort study in high-risk children. *Arch Dermatol* 2006; **142**(5): 561–566.
- 26 Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med* 2000; **19**(8): 1059–1079.
- 27 Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. 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.
- 28 Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol* 2006; **6**: 9.
- 29 Rutjes AW, Reitsma JB, Di NM, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006; **174**(4): 469–476.
- 30 Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004; **140**(3): 189–202.
- 31 Fukiwake N, Furusyo N, Takeoka H *et al.* Association factors for atopic dermatitis in nursery school children in Ishigaki islands – Kyushu University Ishigaki Atopic Dermatitis Study (KIDS). *Eur J Dermatol* 2008; **18**(5): 571–574.