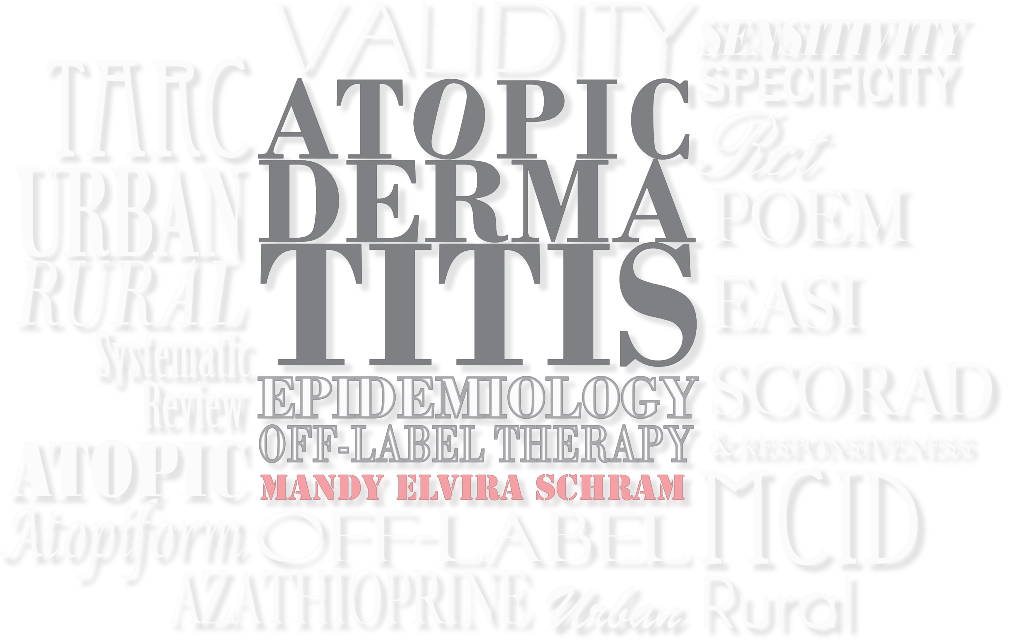
ATOPIC DERMATITIS

M.E. SCHRAM



ATOPIC DERMATITIS

EPIDEMIOLOGY
OFF-LABEL THERAPY

MANDY ELVIRA SCHRAM

Atopic dermatitis: epidemiology & off-label therapy

Mandy Elvira Schram

Atopic dermatitis: epidemiology and off-label therapy

ISBN: 978-94-6182-008-2

Layout and printing: Off Page, www.offpage.nl

Cover: Style doc. Studio, Merel Tempelman

Publication of this thesis was financially supported by:
Abbott BV, AMC afdeling Huidziekten, ALK Abelló BV, Actavis BV, Astellas Pharma BV, Padycare-BAP medical BV, Bauerfeind Benelux BV, Dos Medical BV, Eucerin, EuroTec BV, Huidstichting Chanfleury van IJsselstein, Fagron BV, Galderma SA, Janssen-Cilag BV, La Roche-Posay, Laser Skin Clinics, LEO Pharma BV, Louis Widmer NL, Meda Pharma BV, Medi Nederland BV, Mediq i.s.m. Pierre Fabre Dermo-Cosmétique, Medizorg Homecare & Technology, *mijn*huid huidtherapie Utrecht, MSD BV, Mönlnlycke Health Care BV, Nederlands Bijwerkingen Fonds, Novartis Pharma BV, Phadia BV, Pfizer BV, Raadsheren BV, Stichting EIS, Stiefel - a GSK company, TEVA Nederland BV, Vereniging voor Mensen met Constitutioneel Eczeem, Universiteit van Amsterdam, Van der Bend BV

Atopic dermatitis: epidemiology & off-label therapy

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties ingestelde
commissie, in het openbaar te verdedigen in de Agnietenkapel
op vrijdag 7 oktober 2011, te 12.00 uur

door

Mandy Elvira Schram
geboren te Purmerend

Promotiecommissie:

Promotor:	Prof. dr. J.D. Bos
Co-promotores:	Dr. Ph. I. Spuls Dr. M.M.G. Leeflang
Overige leden:	Prof. dr. C.A.F.M. Bruijnzeel-Koomen Prof. dr. P.J. Coenraads Prof. dr. M. Offringa Dr. I. Terreehorst Prof. dr. H.J.C. de Vries

Faculteit der Geneeskunde

CONTENTS

Part of this introduction has been published in a previous version:
Bos JD, Schram ME, Brenninkmeijer EE,
Middelkamp-Hup MA, Spuls PhI, Sillevis Smitt JH.
Atopic eczema or atopiform dermatitis.
Experimental Dermatology 2010;19:325-331

1

GENERAL INTRODUCTION AND AIMS OF THE THESIS

INTRODUCTION

The hallmark of atopic dermatitis/eczema (AD) is that it is a chronic and relapsing inflammatory skin disorder associated with IgE-mediated sensitization and impairment of the epidermal-barrier function. Pruritus, visible as excoriations and/or prurigo lesions, is the major symptom of AD. Pruritus is so dominant that AD is also known as 'the itch that rashes'. As AD is often a significant source of distress to those affected, the impact of AD on the quality of life is considerable.[1;2] Besides the associated symptoms, the appearance of AD affected skin can be stigmatizing.

AD is a very common disease which is known under many different names, such as constitutional eczema, neurodermitis and allergic eczema. The semantics have been thoroughly described in 'Histoire de la dermatite atopique' by Wallach, Taïeb and Tilles.[3] The word 'atopy' dates from 1922[4], quite well before IgE was recognized.[5] The word 'Allergie' ('allergy') is over 100 years old and was introduced by Von Pirquet.[6] Presence of IgE on cutaneous mast cells was identified and its binding to Langerhans cells in AD was first described in 1986.[7] Functional subpopulations of human T cells, with the principal role for type 2 T cells in central immune organ IgE production, were first revealed in 1990.[8;9] AD is part of the allergy syndrome, in which patients may develop rhinitis, conjunctivitis, food allergy, allergic contact urticaria, allergic asthma and AD in any order and in any combination over time.

There is no laboratory-test based diagnosis possible today. Histopathology is supportive but not definitive. For clinicians, it is difficult to give a precise definition of this disease that has so many different clinical phenotypes, is characterized by such a large variety in severity and is so unpredictable in its natural course in the individual patient. Thereby, until now, no consensus is achieved on major topics as diagnostic criteria, outcome assessment, pathofysiology and therapy.

CLINICAL PHENOTYPES

AD is characterized by highly pruritic erythematous squamous lesions often associated with pronounced lichenification and excoriations. Variations in localisations of eczema occur with age; infantile, juvenile and adult types are usually, but not always discerned.[10-12] In 45–60% of the children, onset of AD occurs during the first six months of life and this *'infantile AD'* runs until 2–3 years of age.[13] In this phase, the typical distribution pattern is that of a balaclava[14], with eczematous and highly pruritic lesions on the head and neck,

sparing the periorbital and perioral regions. In many cases, weeping and crusting occur. In *'juvenile AD'*, the flexural phenotype ensues. This childhood phase normally lasts from the age of 2–3 until puberty (13–15 years old). Typically, lesions are localized in the neck and in the elbow and knee folds. In *'adult AD'*, in addition to this juvenile flexural distribution, there is pronounced involvement of wrists and ankles, and facial and neck eczema become common locations. It must be emphasized that these are clinical observations. In addition, there are a substantial number of entities to be considered in a differential diagnosis (Table 1).

Table 1. Flexural and/or cheek eczema: differential diagnosis in patients with AD clinical phenotype

Chronic dermatoses	atopiform dermatitis prurigo seborrhoic eczema allergic contact dermatitis irritant contact eczema nummular eczema rosacea couperose essential teleangiectasia ulerythema ophryogenes keratosis pilaris juvenile acne psoriasis ichthyosis
Infections and infestations	scabies human immunodeficiency virus dermatophytosis erythema infectiosum other viral exanthema
Malignancies	cutaneous T-cell lymphoma Letterer-Siwe disease
Immunologic disorders	juvenile lupus erythematodes dermatitis herpetiformis graft-vs-host disease dermatomyositis
Immunodeficiencies	Wiskott-Aldrich syndrome severe combined immunodeficiency disease hyper-IgE syndrome DiGeorge syndrome
metabolic disorders	zinc, pyridoxine, or niacin deficiency phenylketonuria

(based in part on Leung *et al* 2004[74])

NOMENCLATURE

In general, the requirements of a disease definition are that it should result in an optimum discrimination; it should be easy to remember and to use, correspond to the current clinical concept of the disease and be acceptable as an accurate tool for studies.[15] The clinical phenotype of AD, but without allergen-specific IgE, is the focus of an ongoing debate. A major step was made by introducing the terms intrinsic and extrinsic AD.[16] The term 'Intrinsic AD' was probably first coined by Wüthrich for patients having the phenotype of AD but without detectable allergen-specific IgE.[17] In 'extrinsic AD', external allergens are assumed to sensitize the patient by penetration of a disturbed epidermal barrier. However, the term atopiform dermatitis (AFD) for 'intrinsic AD' as proposed by J.D. Bos, might be more clear since it indicates that it has the clinical phenotype of AD, but without atopy.[18] The percentage of AFD (intrinsic AD) in patients primarily diagnosed as AD varies from 6.9% to 55.6%.[19] A systematic review on this subject showed that the proportion of IgE sensitisation among phenotypic AD was more frequent in hospital than in population based studies.[20]

PATHOFYSIOLOGY

AD is a multitrait disorder. Two major hypotheses concerning the mechanisms of AD have been proposed; the inside (immunocentric) and the outside (corneocentric) paradigm. The first hypothesis postulates a primary defect in immunological processes (T-cells) that causes IgE sensitisation and a secondary epithelial-barrier dysfunction due to skin inflammation. In the second hypothesis, AD is primarily seen as a disorder of the skin barrier, characterized by malfunction of the corneal layer, enhanced penetration of allergens and secondary sensitisation.

A major development in understanding the pathofysiology was the recognition of functional subsets of T cells, which by their cytokine production profile, could be divided into Th1 cells (mainly producing IFN-γ) and Th2 cells (mainly producing IL-4, IL-5 and IL-13). Langerhans' cells of the skin are believed to contribute to Th2 polarization.[21;22] Th2 cells have been implicated in stimulating B cells to become IgE producing plasma cells. Atopy thus may be seen as a syndrome in which there is central immune dysregulation for yet unexplained reasons. Recent work indeed shows that myeloid and plasmacytoid dendritic cells isolated from peripheral blood of AD patients have an aberrant function compared to healthy controls, as they have a significantly decreased

capacity to produce cytokines (IL-12). This may favour a Th2 cell response (Figure 1).[23;24]

A concordance rate among monozygotic twins of 77% compared to 15% in dizygotic twins, suggests a strong genetic role in AD.[25] Genes encoding for proteins involved in skin barrier function and innate immunity are thought to be

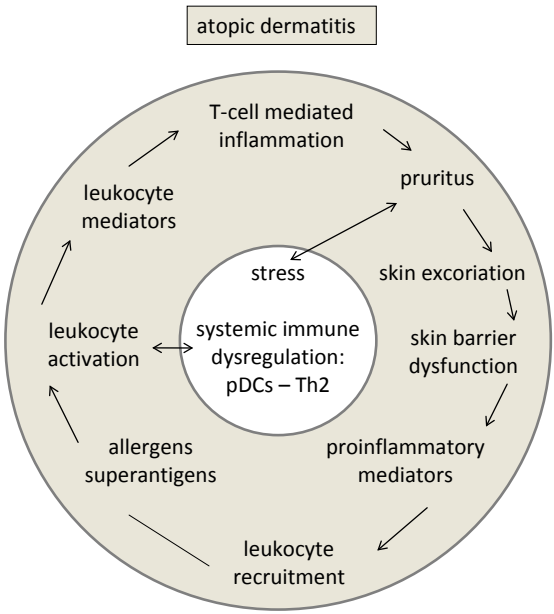


Figure 1. Pathogenesis. AD may be seen as a vicious circle of pathogenetic events in which it is as yet uncertain where it starts. Central is the immunological background of atopy in which a systemic abnormality may be present, perhaps at the level of plasmacytoid dendritic cells (pDCs) which by their abnormal cytokine production profile preferentially induce type 2 T cells (Th2) in the secondary immune organs such as lymph nodes. Pruritus, to begin with, leads to stress and vice versa, and the resulting skin excoriations add to an already existing skin barrier defect. The damage to epithelial cells leads to the production of pro-inflammatory mediators that recruit leukocytes, such as monocytes, eosinophils and T cells, into the lesions. These T cells are activated by dendritic cells that are supposed to present allergens and superantigens to them. The route by which these antigens reach the skin is as yet undetermined but many assume they directly enter the skin through the damaged skin barrier. The activation of type 2 T cells, for which there may be a role for abnormal pDCs, is of particular importance as their production of IL-31 seems to be a major mediator of the pruritus.

important in the genetic susceptibility for AD. Genome wide scans have revealed potential AD associated loci. Several candidate genes have been identified on chromosome 5q31-33.[26] Those genes encode for cytokines involved in the regulation of IgE synthesis and may contribute to the imbalance of Th1 and Th2 immune response.

Susceptibility loci (ATOD1-6) as well as single nucleotide polymorphisms (SNPs) have been described. Within the past 3 years, it has become apparent that of these susceptibility loci, ATOD2 is localized in the 'epidermal differentiation complex' on chromosome 1q21. It harbours the filaggrin gene for which null mutations have been found. Filaggrin has a role in keratin cytoskeleton aggregation during epidermal differentiation and corneocyte formation, when cells of the granular layer collapse into corneal layer scales. Mutations resulting in loss of filaggrin production, both rare and prevalent forms, have been identified in approximately 25% of AD patients in Western, mainly white Europeans.[26-35]

Till now, these mutations explain only a small proportion of the genetic heritability.

EPIDEMIOLOGY

Epidemiology can be defined as a scientific discipline that studies the (environmental) factors determining the causes, frequency and distribution of diseases in a community or specified population. Considerable efforts are being made to investigate the epidemiology of AD. In this introduction we focus on the prevalence, diagnostic criteria and outcome measures of AD.

Prevalence

Prevalence studies have been conducted to both assess the magnitude of this health care problem as to try to identify the environmental factors associated with its development. In order to identify affected patients it is of vital importance to have a clear disease definition and to have reliable and validated diagnostics.

Prevalence can be established as lifetime prevalence, 1-year prevalence or point prevalence. As AD has a fluctuating course, the 1-year prevalence is often used.[36] In order to diagnose AD, diagnostic questionnaires, diagnostic criteria and/or clinical examination have been introduced and used. The choice for one of those diagnostic measures depends on the setting (population versus hospital-based), the available resources (finance, medical staff, time) and on the purpose (estimation of an international prevalence or to establish an accurate diagnosis for inclusion of a clinical trial).

The most commonly used and validated questionnaire is the International Study of Asthma and Allergies in Childhood (ISAAC). The ISAAC provides a standardized method for estimation of prevalences of all atopic diseases within and between countries over the world.[37] In 1998, the prevalence of AD in 56 countries was measured using the ISAAC questionnaire; prevalences ranged from 0.3% to 20.5%. Using ISAAC on large scale contributes to comparable prevalence outcomes worldwide.[38]

Table 2 shows a summary of 1-year period prevalences from different epidemiological studies in different age groups and countries worldwide. From these data, it may be concluded that AD is a very common disease in the infantile and juvenile age periods. Probably at least one to every 30 adults has one or more periods of AD. There is hardly any doubt that during the second half of the 20th century, there has been an increase of the prevalence of AD in western countries[39;40] for which various explanations have been put forward. As genetics alone cannot explain why AD has reached epidemic proportions, environmental factors clearly have a major role in the expression of the disease and thus, in the genotype-phenotype switch. A major theory explaining the increased prevalence and incidence of AD and atopy in general is the 'hygiene hypothesis'.[41-44] In summary, that view indicates that lack of microbial stimulation in the newborns' immune system due to hygienic environments in modern societies has resulted in lack of signals diverting type 2 responses to regulatory and type 1 T cell responses. Western lifestyle and urban over rural residency are thought to result in a rise in allergic diseases, e.g. by less exposure to parasites, better housing and diet. The increase of the prevalence of atopy has come to a standstill now in some western countries[45], while it is on the increase in developing nations[46]. Reasons for this need to be further explored (urban/rural residency, migration, pollution, etc.) for a better understanding of these factors. This could open up to new preventive and treatment strategies.

Table 2. Age period prevalence of AD in five continents over the last fifteen years, based on 1-year period prevalences

Age-group	Europe	North America	South America	Asia	Australia	Africa
AD patients (%)						
infant (0-4)	7.2[75] - 25.6[76]			6.9[77] - 20.3[78]	30.8[79]	1.8[80] - 4.4[80]
child (5-15)	3.4[81] – 26.0[82]	5.4[83] - 24.8[84]	3.7[85] – 8.2[86]	1.7[87] - 22.9[88]	16.3[89]	2.5[90] - 26.1[91]
adult (>16)	0.2[92] – 8.4[93]	7.1[94]		3.0[95]		

Diagnostic criteria

Various lists of diagnostic criteria for AD have been proposed during the last decades. A proportion of these criteria were validated against clinical examination. Clinical examination is regarded the gold standard for diagnosis of AD, although it is hard to standardize and accuracy will depend on the expertise of the examiner. Unluckily, uniformity in the use of diagnostic criteria for AD is lacking. Nevertheless, agreement about the definition of AD and widely accepted diagnostic criteria are needed to conduct future valuable studies and to be able to compare and if applicably pool the results of studies.

The most frequently used sets of clinical criteria for the diagnosis of AD are that of Hanifin and Rajka (H&R) and the UK Working Party criteria. The H&R criteria that were published after a meeting in 1980, have been used in genetic, biological, immunological, epidemiological and clinical studies ever since.[47] The UK Working Party has tried to further identify and validate sets of clinical criteria.[48-50] They went through the various stages of developing criteria, starting with a hospital based population study and ending with community oriented studies.[51;52]

In the Millennium Criteria (MC) proposed by J.D. Bos, there is one mandatory criterium which is biological and immunological in nature.[53] Presence of allergen-specific IgE in a given patient is a prerequisite for using the word atopy. A diagnosis of AD thus can only be made when there is allergen-specific IgE. The MC have been validated and refined in a case-control study.[54] Overall accuracy of the MC was 88.6%. Out of the MC, the best discriminatory principal and additional criteria of the Millennium Criteria were identified using tree analysis. However, the MC need to be further validated in a trial setting that better resembles clinical practice.

Outcome measurement

Outcome measurement can be performed for e.g. disease severity, symptoms, safety and quality of life. Severity outcome measures are needed for valid, reliable and applicable reporting in prospective studies and clinical practice and can serve different purposes: (1) to measure disease severity at a certain point of time and thereby discriminate between patients; (2) to measure changes in disease severity over a period of time (e.g. in an efficacy trial); and (3) to predict disease severity in the future (in prognostic models). To assess disease severity in AD mostly clinical severity outcomes are used. Clinical severity outcome measures can be physician or patients based or both and are designed to assess signs and/or symptoms of the target disease. Charman *et al*. reviewed all the randomized controlled trials (RCT's) of therapeutic intervention published

between 1994 and 2001 and showed that 91% included an assessment of clinical signs, of which only 27% used a clinical severity outcome measure that was published before.[55] Until now, 20 clinical severity outcome measures were proposed for AD, but only three of these measures were recommended in a recently published systematic review by Schmitt *et al*.: severity Scoring of Atopic Dermatitis (SCORAD), Eczema Area and Severity Index (EASI) and Patient-Oriented Eczema Measurement (POEM).[56]

To incorporate outcome measures in daily practice and clinical trials, outcome measures should achieve good clinimetric quality. Besides, validity and reliability, essential features for outcome measures are the responsiveness and the minimal clinically important difference. The responsiveness ensures the ability to detect meaningful changes in disease activity and the minimal clinically important difference can identify changes in clinical severity outcome measurements as clinically relevant or not.

The use of unvalidated outcome measures can provoke bias and inaccuracy. Thereby, the variety in outcome methodology is hindering comparison between RCT's. Currently, attempts are being made on an international level to Harmonise Outcome Measures for Eczema (HOME) by identifying core outcome domains, which are aimed to be widely accepted among dermatologists.[57]

In addition to clinical severity outcome measures, biomarkers are of growing interest for severity outcome assessment. The benefits of biomarkers are that objective read-out is guaranteed and treatment effect can be easily compared between studies. Multiple markers for AD have been proposed, but unfortunately a lot of these could not comply with the requirements for being a reliable biomarker. Requirements are 1) to be disease specific among other diseases of the allergy syndrome, 2) to correlate with other (clinical) severity measurements and 3) to have a sustained effect under different therapeutic strategies.

Thymus and activation-regulated chemokine (TARC), a chemokine that attracts CC chemokine receptor 4 or 8 -positive cells, was found to correlate with disease severity of AD and was disease specific among the allergy syndrome.[58;59]

THERAPY

Treatment of AD can target several aspects of the disease. The first target encompasses the skin-barrier-dysfunction and the decreased synthesis of skin proteins in AD patients. In this light, moisturization of the skin is a priority in restoring the skin barrier function and for this purpose emollients are widely used. Avoiding irritants is advised as AD patients are at risk for developing hand

eczema and irritant contact dermatitis. A new treatment strategy in this area is aimed at acidification of the skin. As neutralization of the stratum corneum adversely impacts epidermal functions including permeability and integrity of the skin, it is thought that acidification of stratum corneum improves these functions in inflammatory skin.[60]

The second target for treating AD is the Staphylococcal aureus colonization of the skin as AD patients have a higher rate of bacterial adhesion and decreased capacity to produce antimicrobial peptides.[61] Thereby, it was shown that Staphylococcal aureus colonization can promote inflammation through superantigen activation.[62;63] Nevertheless, a recent Cohrane review failed to show any evidence that antistaphylococcal intervention is clinically helpful in AD patients with clinically non-infected skin.[64] However, chronic use of dilute bleach baths decreased the clinical severity of AD in patients with clinical signs of secondary bacterial infections.[65] A study by Hata *et al.* showed that oral vitamin D significantly increased the levels of cathelicidin, a antimicrobial peptide.[66] The impact of increasing the production of antimicrobial peptides on clinical disease severity is a new field to explore.

Anti-inflammatory therapy is the third target for treatment of AD. Mainstay of anti-inflammatory therapy is the use of topical corticosteroids. However, while having an anti-inflammatory effect, topical steroids also decrease epidermal proliferation and differentiation and the synthesis of lipids. This is probably the main reason for a rebound effect after discontinuation.[67] Other topical drugs include tar products and the relatively new calcineurin inhibitors.[68;69] These therapies can be applied in the acute phase or as maintenance treatment.

For severe cases of AD, systemic anti-inflammatory treatment is indicated. Registered systemic treatment options for AD are cyclosporin and oral glucocorticosteroids, of which cyclosporin is first choice.[70;71] While proven to be effective, some of the patients have a contra-indication for or have to discontinue cyclosporin due to ineffectiveness or side effects. Moreover, long-term use of cyclosporin raises concerns over (nephro)toxicity. Systemic glucocorticosteroids are used frequently to suppress exacerbations, although clinical data are lacking. A recent RCT comparing short term cyclosporin versus prednisolone was interrupted because of an unsuspected high proportion of withdrawals due to severe exacerbations in the prednisolone group.[71] Long-term treatment with glucocorticosteroids is relatively contra-indicated due to the cumulative effect of the side effects.[72] This illustrates the need for novel medium-to-long term treatment options for patients with severe AD.

OFF-LABEL THERAPY

Currently, there are no new systemic drugs for AD to be expected. This results from the fact that therapeutics can only be registered by pharmaceutical companies after expensive development and clinical trials. It appears that the commercial interest is too small for these investments. In the light of increasing health-care costs however, awareness starts to come that there are valuable old and cheap drugs within reach. Those drugs are not registered for use in AD, thus treatment is off-label.[73] Many drugs in daily practice are prescribed off-label, approaching 50% of prescriptions in dermatology.[73]

The list of off-label tested and prescribed drugs for AD is long and includes the use of mycofenolate mofetil, azathioprine, methotrexate, interferon-γ, oral pimecrolimus and intravenous immunoglobulines. The strength of evidence of the first three is generally low but supports their use in AD.[72]

With the introduction of biologic treatment in dermatology, evidence appears about its efficacy in AD. In general, biologics have the property to deplete specific cells and mediators. Mainly case reports and pilot studies have been performed testing the effectiveness of biologics such as infliximab, efalizumab, adalimumab, omalizumab and etanercept. Though some of these biologics have shown to be effective in some patients, their immune modulating properties predispose to skin infection and the potential risk of malignancies in long-term use is of concern.

With admission of drugs in unregistered indications, there is an unknown balance between dose, efficacy and safety pattern. As international and national regulatory agencies are currently applying new rules and regulations concerning off-label drug use, guidelines are needed to provide evidence for or against a given therapy, to overcome legal issues involved and to indicate new research areas.

AIMS OF THE THESIS

Atopic dermatitis is one of the most common forms of dermatitis and much research is directed at its epidemiology and therapy. Nevertheless, there are aspects of AD that need some further attention.

As part of the 'hygiene hypothesis' it is thought that eczema is more common in urban than in rural communities. However, such a notion was never assessed systematically. **Chapter 2** is a systematic review that shows the available evidence by summarizing all the relevant primary studies on this subject.

Over the years many criteria were introduced to diagnose AD. However, not all of these diagnostic criteria are validated and show to have good diagnostic properties. **Chapter 3.1** is a systematic review that summarizes the evidence concerning the validity of diagnostic criteria for AD.

The Millennium Criteria are diagnostic criteria that were developed to accurately diagnose true atopic AD and are aimed to ensure homogenous populations to facilitate clinical research. In **Chapter 3.2** the Millennium Criteria are further validated and refined by conducting a hospital-based cohort study.

Severity outcome assessment is conditional for evaluating treatment response or to compare different therapeutic options for efficacy/effectiveness. Thereby, the possibility to pool results from different clinical trials is of vital importance to ensure meta-analysis and thereby to generate a high level of evidence in favour of or against the use of a given therapy.

Chapter 4 is a validation study in which two clinimetric properties of four severity outcome measures are analysed.

Since therapeutic options are scarce, management of patients with severe AD can be challenging. Cyclosporin and oral corticosteroids are the only systemic treatments that are officially registered for AD in the Netherlands. However treatment can fail or can be contra-indicated. **Chapter 5.1** is a RCT in which two off-label drugs, methotrexate and azathioprine, are compared for efficacy and safety in patients with severe AD. **Chapter 5.2 & 5.3** are systematic reviews that encompass the off-label use of azathioprine and efalizumab in dermatological patients.

REFERENCE LIST

1. Carroll CL, Balkrishnan R, Feldman SR *et al*. The burden of atopic dermatitis: impact on the patient, family, and society. *Pediatr Dermatol* 2005; **22**: 192-9.
2. Chamlin SL, Frieden IJ, Williams ML *et al*. Effects of atopic dermatitis on young American children and their families. *Pediatrics* 2004; **114**: 607-11.
3. Wallach D, Taïb A, Tilles G. *Histoire de la dermatite atopique*. Issy les Moulineaux, 2004.
4. Coca AF, Cooke RA. On the Classification of the Phenomena of Hypersensitiveness. *J Immunol* 1923; **8**: 163-82.
5. Ishizaka K, Ishizaka T, Hornbrook MM. Physico-chemical properties of human reaginic antibody. IV. Presence of a unique immunoglobulin as a carrier of reaginic activity. *J Immunol* 1966; **97**: 75-85.
6. Von Pirquet CE. Allergie. *München Med Wochenschr* 1906; **Iiii**: 1457.
7. Bruijnzeel-Koomen CA, Mudde GC, Bruijnzeel PL. The presence of IgE molecules on epidermal Langerhans cells in atopic dermatitis and their significance for its pathogenesis. *Allerg Immunol (Paris)* 1989; **21**: 219-23.
8. Bos JD, Kapsenberg ML, Smitt JH. Pathogenesis of atopic eczema. *Lancet* 1994; **343**: 1338-41.
9. Wierenga EA, Snoek M, Jansen HM *et al*. Human atopen-specific types 1 and 2 T helper cell clones. *J Immunol* 1991; **147**: 2942-9.
10. Bos JD, Sillevis Smitt JH. Atopic dermatitis. *J Eur Acad Dermatol Venereol* 1996; **7**: 101-14.
11. Schudel P, Wuthrich B. [Clinical course of childhood atopic neurodermatitis. A catamnestic study of 121 cases]. *Z Hautkr* 1985; **60**: 479-86.
12. Yates VM, Kerr RE, MacKie RM. Early diagnosis of infantile seborrhoeic dermatitis and atopic dermatitis--clinical features. *Br J Dermatol* 1983; **108**: 633-8.
13. Rudikoff D, Lebwohl M. Atopic dermatitis. *Lancet* 1998; **351**: 1715-21.
14. 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**: 561-6.
15. Burton JL. The logic of dermatological diagnosis. Dowling oration 1980. *Clin Exp Dermatol* 1981; **6**: 1-21.
16. Schmid-Grendelmeier P, Simon D, Simon HU *et al*. Epidemiology, clinical features, and immunology of the "intrinsic" (non-IgE-mediated) type of atopic dermatitis (constitutional dermatitis). *Allergy* 2001; **56**: 841-9.
17. Wuthrich B. Neurodermatitis atopica sive constitutionalis. Ein pathogenetisches Modell aus der Sicht des Allergologen. *Akt Dermatol* 1983; **9**: 1-7.
18. Bos JD. Atopiform dermatitis. *Br J Dermatol* 2002; **147**: 426-9.
19. Brenninkmeijer EE, Spuls PI, Legierse CM *et al*. Clinical differences between atopic and atopiform dermatitis. *J Am Acad Dermatol* 2008; **58**: 407-14.
20. Flohr C, Johansson SG, Wahlgren CF *et al*. How atopic is atopic dermatitis? *J Allergy Clin Immunol* 2004; **114**: 150-8.
21. Kondo H, Ichikawa Y, Imokawa G. Percutaneous sensitization with allergens through barrier-disrupted skin elicits a Th2-dominant cytokine response. *Eur J Immunol* 1998; **28**: 769-79.
22. Strid J, Hourihane J, Kimber I *et al*. Disruption of the stratum corneum allows potent epicutaneous immunization with protein antigens

resulting in a dominant systemic Th2 response. *Eur J Immunol* 2004; **34**: 2100-9.

23. Lebre MC, van Capel TM, Bos JD *et al*. Aberrant function of peripheral blood myeloid and plasmacytoid dendritic cells in atopic dermatitis patients. *J Allergy Clin Immunol* 2008; **122**: 969-76.
24. Novak N, Valenta R, Bohle B *et al*. FcepsilonRI engagement of Langerhans cell-like dendritic cells and inflammatory dendritic epidermal cell-like dendritic cells induces chemotactic signals and different T-cell phenotypes in vitro. *J Allergy Clin Immunol* 2004; **113**: 949-57.
25. Schultz Larsen FV, Holm NV. Atopic dermatitis in a population based twin series. Concordance rates and heritability estimation. *Acta Derm Venereol Suppl (Stockh)* 1985; **114**: 159.
26. Morar N, Cookson WO, Harper JI *et al*. Filaggrin mutations in children with severe atopic dermatitis. *J Invest Dermatol* 2007; **127**: 1667-72.
27. Barker JN, Palmer CN, Zhao Y *et al*. Null mutations in the filaggrin gene (FLG) determine major susceptibility to early-onset atopic dermatitis that persists into adulthood. *J Invest Dermatol* 2007; **127**: 564-7.
28. Nomura T, Sandilands A, Akiyama M *et al*. Unique mutations in the filaggrin gene in Japanese patients with ichthyosis vulgaris and atopic dermatitis. *J Allergy Clin Immunol* 2007; **119**: 434-40.
29. Palmer CN, Irvine AD, Terron-Kwiatkowski A *et al*. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat Genet* 2006; **38**: 441-6.
30. Ruether A, Stoll M, Schwarz T *et al*. Filaggrin loss-of-function variant contributes to atopic dermatitis risk in the population of Northern Germany. *Br J Dermatol* 2006; **155**: 1093-4.
31. Sandilands A, Terron-Kwiatkowski A, Hull PR *et al*. Comprehensive analysis of the gene encoding filaggrin uncovers prevalent and rare mutations in ichthyosis vulgaris and atopic eczema. *Nat Genet* 2007; **39**: 650-4.
32. Weidinger S, Illig T, Baurecht H *et al*. Loss-of-function variations within the filaggrin gene predispose for atopic dermatitis with allergic sensitizations. *J Allergy Clin Immunol* 2006; **118**: 214-9.
33. Baurecht H, Irvine AD, Novak N *et al*. Toward a major risk factor for atopic eczema: meta-analysis of filaggrin polymorphism data. *J Allergy Clin Immunol* 2007; **120**: 1406-12.
34. Rodriguez E, Baurecht H, Herberich E *et al*. Meta-analysis of filaggrin polymorphisms in eczema and asthma: robust risk factors in atopic disease. *J Allergy Clin Immunol* 2009; **123**: 1361-70.
35. Rodriguez E, Baurecht H, Herberich E *et al*. Meta-analysis of filaggrin polymorphisms in eczema and asthma: robust risk factors in atopic disease. *J Allergy Clin Immunol* 2009; **123**: 1361-70.
36. Radulescu M, Diepghen T, Williams HC. What makes a good prevalence survey? In: *Evidence-based Dermatology* (Williams,HC, ed), 2 edn. Blackwell Publishing, 2008: 61-7.
37. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. The International Study of Asthma and Allergies in Childhood (ISAAC) Steering Committee. *Lancet* 1998; **351**: 1225-32.
38. Flohr C, Weinmayr G, Weiland Deceased SK *et al*. How well do questionnaires perform compared with physical examination in detecting flexural eczema? Findings from the

International Study of Asthma and Allergies in Childhood (ISAAC) Phase Two. *Br J Dermatol* 2009.

39. Ninan TK, Russell G. Respiratory symptoms and atopy in Aberdeen schoolchildren: evidence from two surveys 25 years apart. *BMJ* 1992; **304**: 873-5.
40. Taylor B, Wadsworth J, Wadsworth M *et al*. Changes in the reported prevalence of childhood eczema since the 1939-45 war. *Lancet* 1984; **2**: 1255-7.
41. Flohr C, Pascoe D, Williams HC. Atopic dermatitis and the 'hygiene hypothesis': too clean to be true? *Br J Dermatol* 2005; **152**: 202-16.
42. Holt PG. Parasites, atopy, and the hygiene hypothesis: resolution of a paradox? *Lancet* 2000; **356**: 1699-701.
43. McKeever TM, Lewis SA, Smith C *et al*. The importance of prenatal exposures on the development of allergic disease: a birth cohort study using the West Midlands General Practice Database. *Am J Respir Crit Care Med* 2002; **166**: 827-32.
44. Sherriff A, Golding J. Hygiene levels in a contemporary population cohort are associated with wheezing and atopic eczema in preschool infants. *Arch Dis Child* 2002; **87**: 26-9.
45. Eder W, Ege MJ, von Mutius ME. The asthma epidemic. *N Engl J Med* 2006; **355**: 2226-35.
46. Zar HJ, Ehrlich RI, Workman L *et al*. The changing prevalence of asthma, allergic rhinitis and atopic eczema in African adolescents from 1995 to 2002. *Pediatr Allergy Immunol* 2007; **18**: 560-5.
47. Hanifin JM, Rajka G. Diagnostic features of atopic dermatitis. *Acta Derm Venereol Suppl (Stockh)* 1980; **92**: 44-7.
48. Williams HC, Burney PG, Strachan D *et al*. 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**: 397-405.
49. 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.
50. 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.
51. Popescu CM, Popescu R, Williams H *et al*. Community validation of the United Kingdom diagnostic criteria for atopic dermatitis in Romanian schoolchildren. *Br J Dermatol* 1998; **138**: 436-42.
52. 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**: 12-7.
53. Bos JD, Van Leent EJ, Sillevis Smitt JH. The millennium criteria for the diagnosis of atopic dermatitis. *Exp Dermatol* 1998; **7**: 132-8.
54. Brenninkmeijer EE. Atopic and atopiform dermatitis. 52-69. 2009. Amsterdam, The Netherlands, Academic Medical Center. Ref Type: Thesis/Dissertation
55. Charman C, Chambers C, Williams H. Measuring atopic dermatitis severity in randomized controlled clinical trials: what exactly are we measuring? *J Invest Dermatol* 2003; **120**: 932-41.
56. Schmitt J, Langan S, Williams HC. What are the best outcome measurements for atopic eczema? A systematic review. *J Allergy Clin Immunol* 2007; **120**: 1389-98.
57. Schmitt J, Williams H. Harmonising Outcome Measures for Eczema

(HOME). Report from the First International Consensus Meeting (HOME 1), 24 July 2010, Munich, Germany. *Br J Dermatol* 2010; **163**: 1166-8.

58. Kakinuma T, Nakamura K, Wakugawa M *et al*. Thymus and activation-regulated chemokine in atopic dermatitis: Serum thymus and activation-regulated chemokine level is closely related with disease activity. *J Allergy Clin Immunol* 2001; **107**: 535-41.
59. Hijnen D, De Bruin-Weller M, Oosting B *et al*. Serum thymus and activation-regulated chemokine (TARC) and cutaneous T cell- attracting chemokine (CTACK) levels in allergic diseases: TARC and CTACK are disease-specific markers for atopic dermatitis. *J Allergy Clin Immunol* 2004; **113**: 334-40.
60. Hatano Y, Man MQ, Uchida Y *et al*. Maintenance of an acidic stratum corneum prevents emergence of murine atopic dermatitis. *J Invest Dermatol* 2009; **129**: 1824-35.
61. Gallo RL, Murakami M, Ohtake T *et al*. Biology and clinical relevance of naturally occurring antimicrobial peptides. *J Allergy Clin Immunol* 2002; **110**: 823-31.
62. Ardern-Jones MR, Black AP, Bateman EA *et al*. Bacterial superantigen facilitates epithelial presentation of allergen to T helper 2 cells. *Proc Natl Acad Sci U S A* 2007; **104**: 5557-62.
63. Marone G, Rossi FW, Detoraki A *et al*. Role of superallergens in allergic disorders. *Chem Immunol Allergy* 2007; **93**: 195-213.
64. Bath-Hextall FJ, Birnie AJ, Ravenscroft JC *et al*. Interventions to reduce Staphylococcus aureus in the management of atopic eczema: an updated Cochrane review. *Br J Dermatol* 2010; **163**: 12-26.
65. Huang JT, Abrams M, Tlougan B *et al*. Treatment of Staphylococcus aureus colonization in atopic dermatitis decreases disease severity. *Pediatrics* 2009; **123**: e808-e814.
66. Hata TR, Kotol P, Jackson M *et al*. Administration of oral vitamin D induces cathelicidin production in atopic individuals. *J Allergy Clin Immunol* 2008; **122**: 829-31.
67. Jensen JM, Pfeiffer S, Witt M *et al*. Different effects of pimecrolimus and betamethasone on the skin barrier in patients with atopic dermatitis. *J Allergy Clin Immunol* 2009; **123**: 1124-33.
68. McCollum AD, Paik A, Eichenfield LF. The Safety and Efficacy of Tacrolimus Ointment in Pediatric Patients with Atopic Dermatitis. *Pediatr Dermatol* 2010.
69. Paghdal KV, Schwartz RA. Topical tar: back to the future. *J Am Acad Dermatol* 2009; **61**: 294-302.
70. Schmitt J, Schmitt N, Meurer M. Cyclosporin in the treatment of patients with atopic eczema - a systematic review and meta-analysis. *J Eur Acad Dermatol Venereol* 2007; **21**: 606-19.
71. Schmitt J, Schakel K, Folster-Holst R *et al*. Prednisolone vs. ciclosporin for severe adult eczema. An investigator-initiated double-blind placebo-controlled multicentre trial. *Br J Dermatol* 2010; **162**: 661-8.
72. Schmitt J, Schakel K, Schmitt N *et al*. Systemic treatment of severe atopic eczema: a systematic review. *Acta Derm Venereol* 2007; **87**: 100-11.
73. Stafford RS. Regulating off-label drug use--rethinking the role of the FDA. *N Engl J Med* 2008; **358**: 1427-9.
74. Leung DY, Nicklas RA, Li JT *et al*. Disease management of atopic dermatitis: an updated practice parameter. Joint Task Force on Practice Parameters. *Ann Allergy Asthma Immunol* 2004; **93**: S1-21.

75. 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.

76. Wadonda-Kabondo N, Sterne JA, Golding J *et al*. A prospective study of the prevalence and incidence of atopic dermatitis in children aged 0-42 months. *Br J Dermatol* 2003; **149**: 1023-8.

77. 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.

78. Kawaguchi H, Takeuchi M, Tanaka Y *et al*. [Atopic dermatitis on infant physical examination]. [Japanese]. *Arerugi* 1999; **48**: 686-90.

79. Foley P, Zuo Y, Plunkett A *et al*. The frequency of common skin conditions in preschool-age children in Australia: atopic dermatitis. *Archives of Dermatology* 2001; **137**: 293-300.

80. 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.

81. Banac S, Tomulic KL, Ahel V *et al*. Prevalence of asthma and allergic diseases in Croatian children is increasing: survey study. *Croat Med J* 2004; **45**: 721-6.

82. Dotterud LK, Kvammen B, Lund E *et al*. Prevalence and some clinical aspects of atopic dermatitis in the community of Sor-Varanger. *Acta Dermato-Venereologica* 1995; **75**: 50-3.

83. Rodriguez Orozco AR, Nunez Tapia RM. Prevalence of atopic dermatitis in 6-14 year old children in Morelia, Michoacan, Mexico. *Rev Alerg Mex* 2007; **54**: 20-3.

84. Maymi MA, Somolinos AL, Nazario CM *et al*. The prevalence of atopic dermatitis in Puerto Rican school children. *P R Health Sci J* 2007; **26**: 127-33.

85. Camelo-Nunes IC, Wandalsen GF, Melo KC *et al*. Prevalence of atopic eczema and associated symptoms in school children. *J Pediatr (Rio J)* 2004; **80**: 60-4.

86. Sole D, Camelo-Nunes IC, Wandalsen GF *et al*. Prevalence of atopic eczema and related symptoms in Brazilian schoolchildren: results from the International Study of Asthma and Allergies in Childhood (ISAAC) phase 3. *Journal of Investigational Allergology & Clinical Immunology* 2006; **16**: 367-76.

87. Yang YC, Cheng YW, Lai CS *et al*. Prevalence of childhood acne, ephelides, warts, atopic dermatitis, psoriasis, alopecia areata and keloid in Kaohsiung County, Taiwan: a community-based clinical survey. *J Eur Acad Dermatol Venereol* 2007; **21**: 643-9.

88. Yura A, Shimizu T. Trends in the prevalence of atopic dermatitis in school children: longitudinal study in Osaka Prefecture, Japan, from 1985 to 1997. *British Journal of Dermatology* 2001; **145**: 966-73.

89. Marks R, Kilkenny M, Plunkett A *et al*. The prevalence of common skin conditions in Australian school students: 2. Atopic dermatitis. *British Journal of Dermatology* 1999; **140**: 468-73.

90. 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. *British Journal of Dermatology* 2007; **156**: 111-6.

91. Bouayad Z, Aichane A, Afif A *et al*. Prevalence and trend of self-reported asthma and other allergic disease symptoms in Morocco: ISAAC phase I and III. *Int J Tuberc Lung Dis* 2006; **10**: 371-7.

92. Herd RM, Tidman MJ, Prescott RJ *et al*. Prevalence of atopic eczema in the community: the Lothian Atopic Dermatitis study. *British Journal of Dermatology* 1996; **135**: 18-9.

93. Worm M, Forschner K, Lee HH *et al*. Frequency of atopic dermatitis and relevance of food allergy in adults in Germany. *Acta Dermato-Venereologica* 2006; **86**: 119-22.

94. Hanifin JM, Reed ML. A population-based survey of eczema prevalence in the United States. *Dermatitis* 2007; **18**: 82-91.

95. Muto T, Hsieh SD, Sakurai Y *et al*. Prevalence of atopic dermatitis in Japanese adults. *British Journal of Dermatology* 2003; **148**: 117-21.

ME Schram
AM Tedja
R Spijker
JD Bos
HC Williams
PhI Spuls

Is there a rural/urban gradient in the prevalence of eczema?
A systematic review
British journal of Dermatology 2010;162:964-73

2

IS THERE A RURAL/URBAN GRADIENT IN THE PREVALENCE OF ECZEMA? A SYSTEMATIC REVIEW

SUMMARY

Rationale: Eczema affects approximately 10% of all schoolchildren in the western world and has shown an increase over the past decades in 'developing' countries. Numerous factors have been suggested that might contribute to the increasing prevalence of eczema. A plausible explanation is the role of environmental factors. As part of the 'hygiene hypothesis' it has been thought that eczema is more common in urban than in rural communities, but such a notion has never been assessed systematically.

Objective: Our aim was to assess whether there is a rural/urban gradient for the prevalence of eczema and, if so, to what extent.

Methods: All data sources were identified through a search in MEDLINE and EMBASE. All primary studies comparing the prevalence rate of eczema between urban and rural populations were assessed for eligibility. Included articles were reviewed for methodological quality and a relative risk was calculated to indicate the risk of eczema in urban over rural areas.

Results: Twenty-six articles were included for analysis. Nineteen showed a higher risk for eczema in an urbanized area, of which 11 were significant. Six studies showed a lower risk of eczema in an urbanized area, of which one was statistically significant. One study had a relative risk of 1.00. Results were more homogeneous among studies of good methodological quality. A pooled relative risk could have been calculated but was not because of heterogeneity.

Conclusion: There is some evidence of a higher risk for eczema in urban compared with rural areas, suggesting that place of residence may have a role in the pathogenesis of eczema. Future reviews on environmental circumstances should be carried out to reveal the factors associated with a higher prevalence of eczema in urban areas and the association with other allergic diseases.

INTRODUCTION

Eczema affects approximately 10% of all schoolchildren in the western world and has shown a rapid increase over the past decades in 'developing' countries.[1-4] Numerous factors have been suggested that might contribute to the increasing prevalence of eczema and other atopic diseases. A genetic predisposition is likely to play an important role for both skin barrier dysfunction and inflammatory responses, but such factors cannot explain the rapid increase in disease prevalence. Changes in public and professional awareness of eczema and diagnostic labelling are also thought to have had their effect. Another plausible explanation however is the role of environmental factors.[5] A major theory explaining the increased prevalence and incidence of eczema and atopy in general is the 'hygiene hypothesis'.[6-9] This view indicates that hygienic environments in modern society result in insufficient microbial stimulation in the immune system of newborns. This leads to lack of signals diverting type 2 responses to regulatory and type 1 T-cell responses and thereby induces eczema. Over the past 50 years industrialization, urbanization and improvement of housing and hygiene have been prominent in the western world, whereas 'developing' countries are currently undergoing these steps. Indications for an increasing prevalence in 'developing' countries have already been found.[10]

As part of the 'hygiene hypothesis' it was thought that eczema is more common in urban than in rural communities, but such a notion has never been assessed systematically.

Our aim was to assess whether there is a rural/urban gradient for the prevalence of eczema and if so, to what extent. If a strong rural/urban gradient in populations that are genetically similar is found, this suggests that area of residence has a role in the pathogenesis of eczema.

METHODS

Literature search

In February 2009, a literature search in MEDLINE and EMBASE was performed. As the main search strategy, 'eczema' 'atopic dermatitis', 'epidemiology', 'prevalence', 'urban', 'rural' and all their synonyms were used and combined (Table 1). References to all relevant articles found were checked for eligible articles.

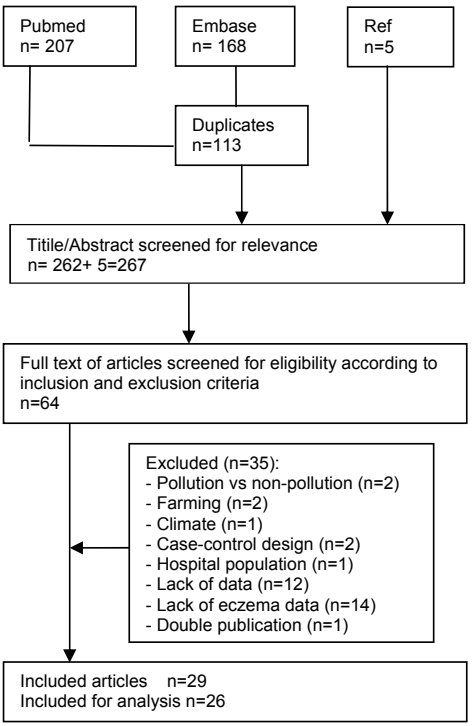


Figure 1. Flowchart of selection process

Inclusion and exclusion criteria

All primary studies comparing the prevalence rate of eczema between urban and rural populations were assessed for eligibility. The phenotypic appearance of eczema and/or questions leading to the presence of that phenotypic appearance (e.g. in questionnaires) were considered sufficient for the diagnosis of eczema. Although the terms 'atopic dermatitis' or 'atopic eczema' are commonly used in epidemiological studies, we use the term 'eczema' throughout in accordance with the World Allergy Nomenclature Committee's recommendation.[11] The demonstration of allergen-specific IgE sensitization of patients with eczema was not required. Included were all definitions of 'urban' and 'rural' described in articles; if no definition was given, but the prevalence of eczema was assessed between areas with different population densities (e.g. villages vs. cities),

Table 1. Search strategy for MEDLINE and EMBASE

1. exp dermatitis, atopic/
2. ((atopic$ or intrins$ or allergic$) adj3 (dermatit$ or eczem$)).tw.
3. 1 or 2
4. exp animals/ not (exp animal/ and exp humans/)
5. 3 not 4
6. exp Urban Rural Difference/ or exp Urban Area/ or exp Urban Population/ or exp Urban Hygiene/
7. exp Rural Health Care/ or exp Rural Population/ or exp Rural Area/ or exp Rural Hygiene/ or exp Rural Health Nursing/
8. 6 or 7
9. rural.ti,ab.
10. bucolic.ti,ab.
11. country-side.ti,ab.
12. rustic.ti,ab.
13. provincial.ti,ab.
14. agrarian.ti,ab.
15. urban*.ti,ab.
16. metropolitan.ti,ab.
17. municipal.ti,ab.
18. nonagricultural.ti,ab.
19. nonfarm.ti,ab.
20. inner-city.ti,ab.
21. farming.ti,ab.
22. ((rural or remote or nonmetropolitan) adj (communit$ or area? or region? or province?)).tw.
23. 11 or 21 or 9 or 17 or 12 or 20 or 15 or 14 or 22 or 18 or 19 or 10 or 13 or 16
24. 8 or 23
25. 24 and 5

articles were included. Most important is that the same strategies and definitions were used to assess the prevalence of eczema in both the rural (less dense populations) and the urban setting (more dense populations).

Articles comparing farming with non-farming communities, polluted with non-polluted areas, hospital populations, case-control studies, unpublished articles, reviews or abstracts, double publications or articles reporting only descriptive information were excluded. Articles that lacked essential data for calculating a relative risk (RR) were also excluded from analysis. No restrictions were imposed regarding sample size, age, sex and skin type of the subjects and no language restriction was applied.

Study selection and data extraction

All articles with a title and abstract considering a comparison of the prevalence rate of eczema, asthma and rhinitis between urban and rural populations were selected for relevance by the reviewers M.E.S. and A.M.T. To determine eligibility, the full text of the selected articles was screened. Data were extracted independently and disagreements about the selection process and data extraction were solved by discussion between the two reviewers.

If studies reported measured prevalences in rural and urban areas, but only showed combined data, the authors were contacted with a request to provide data per group.

Assessment of methodological quality

Measuring the methodological quality of prevalence studies was done according to the guidelines for the critical appraising of studies of prevalence or incidence of a health problem as proposed by Loney *et al.* and adapted by M. Radulescu *et al.*[12-13] Each article is scored according to seven criteria: (1) target population – this item was considered adequate if the prevalence surveys gave a definition of the target population, information on geographical area, age and sex; (2) sampling methods - these were considered adequate if a whole or an entirely random sample of the population was used; (3) sample size – adequate if sample size was > 810 subjects [assuming prevalence of eczema was 5% and an error rate of ± 1.5% at the 95% confidence interval (CI) is accepted];[12-14] (4) response rate – a response rate of > 70% was considered adequate;[15,16] (5) information on non-responders – accepted if any attempt was made to obtain information about reasons for non-participation and characteristics of the group of non-responders; (6) use of valid and repeatable disease definitions – this item was considered adequate if a generally accepted disease definition was presented and diagnosis was defined by validated criteria or clinical observation;[17] and (7) efforts to reduce observer bias – considered adequate when attempts were made to reduce observer bias by training, teaching or presenting interobserver variability in case of surveys based on clinical observation or interviews; when one observer is responsible for all examinations, a conflict of interest should be excluded.

When assessing these points, special attention was given to detecting differences in data collection processes between the urban and rural areas. If such a difference was found, the particular criterion was considered inadequate.

Good methodological quality of an article was defined as not having any limitations on the above-mentioned criteria. If only 'reduction of observer bias' was lacking and/or there was no information given on non-responders, we also considered the articles to be of good quality. Articles with limitations in the sampling methods, sample size, response rate and use of valid and repeatable disease definition were considered to be of low quality.

Statistical analysis

We used RRs and 95% CIs to explore the risk of eczema in an urban over a rural residential area. Countries were classified as being 'developed/western'

and 'developing' according to the classification of the United Nations and International Monetary Fund (http://www.imf.org/external/pubs/ft/weo/2009/01/weodata/groups.htm) at the time the study was performed.

The results of individual studies were compiled into The Cochrane Collaboration Review Manager 5 (http://www.cc-ims.net/revman) and analysed using Metaview 5. The χ^2 test was used to calculate heterogeneity between studies. The pooled RR and 95% CI were estimated using a random-effect model, as heterogeneity was evident.

RESULTS

Literature search

Figure 1 summarizes the selection process for studies comparing urban and rural prevalences of eczema. An initial search retrieved 267 articles of which five were found during the additional reference search. After screening titles and abstracts for eligibility, 64 articles were selected. Of the initial 64 selected articles, 35 were excluded after screening the full texts of the articles. Three primarily eligible studies were excluded from the analysis because they did not report essential data from which a RR could be calculated.[18-20]

Study description

Twenty-nine articles, published between 1982 and 2009 were included in this review (Table 2).[18-46] All articles were prospective cohort studies, with the exception of one retrospective cohort study.[22] Study populations consisted primarily of children (n = 22); four studies included subjects of all ages and three studies only adults. Seventeen studies were conducted in 'developed/western' countries and 12 in 'developing' countries.

The definitions of 'urban' and 'rural' employed were often not reported and only predefined areas were given (n=17). In most cases urban and rural residency was defined by the number of inhabitants. However, the cut-off values varied from 200 to 100 000 inhabitants between studies. Also, inhabitants per square metre,[25] distance from the city centre, physical features of the school building involved and water supply,[35] and National Health Service classification of the family practice were used.[30] In the study of Van der Ven *et al.*[41] the level of urbanization was judged by the patients themselves. The definition of 'eczema' and diagnostic or epidemiological criteria for diagnosis also varied considerably from study to study or was not given at all.

Table 2. Overview of included articles

Author (year of publication, country)	Age (y)	Diagnosis of eczema	Prev estimate	Definition of urban and rural	Prev Urban	Prev Rural
Bouayad *et al.*[21] (2006, Morocco)	13-14	ISAAC	1-year	None	20.3%[a]	13.3%
Bråbäck *et al.*[22] (2004, Sweden)	17-20	Qnaire, Clin Exam	Point	Urban: home located in settlement with at least 200 inhabitants	3.3%[b]	2.7%[b]
Chalmers *et al.*[23] (2007, South Africa)	3-11	UK, Clin Exam	Point	None	2.1%[b]	0.3%[b]
Du Prel *et al.*[24] (2006, Germany)	6	Clin Exam	Lifetime	None	16.1%	15.4%
Dutau *et al.*[18] (1997, France)	5-6	Qnaire	Lifetime	Urban: areas with industrial impact Rural: areas without heavy industrial impact	19.4%	17.1%
Galassi *et al.*[25] (2006, Italy)	13-14	ISAAC	1-year	Urban: >500 000 Other areas: population density of <1000 inhabitants /km^2	8.6%	8.4%
Gniazdowska *et al.*[26] (1990, Poland)	10-15	Qnaire	Unk	None	2.1%	0.4%
Graif *et al.*[19] (2004, Israel)	13-14	Modified ISAAC	Point	Urban: area of residence > 2000 inhabitants Rural: <2000	7.7%	8.6%
Haileamlak *et al.*[27] (2005, Ethiopia)	1-5	ISAAC	Lifetime	None	3.8%[a]	4.0%[a]
Hanifin *et al.*[28] (2007, United States)	All	Qnaire	1-year	Rural: <100 000 size of population centre Urban: >100 000	6%	6%
Heinrich *et al.*[29] (2001, Germany)	5-14	Qnaire, Clin Exam	Lifetime	Urban: 16000-35000 inhabitants Rural: <2500 inhabitants	11.3%	7.7%
Iversen *et al.*[30] (2005, Scotland)	>16	Qnaire	Lifetime	Rural: by National Health Service classification of the family practice - practice that received rural practice payment for more than one-third of the patients	18.3%	13.7%
Kim *et al.* [31] (2000, South Korea)	6-8 10-12 16-18	Qnaire, Clin Exam	Lifetime	None. Pre-defined areas plus industrialised area (excluded)	7.8%	6.6%

Kuhlich *et al.* [32] (2000, Germany)	5-11	ISAAC. Qnaire	1-year	None. Pre-defined areas. Inhabitants per square meter given	17.1%	9.9%
Laughter *et al.* [33] (2000, United States)	5-9	Qnaire	Lifetime	None	18.6%	13.9%
Lynch *et al.* [34] (1984, Venezuela)	All	Interview, Clin Exam	Unk	None	3.2%	3.6%
Mavale-Manuel *et al.* [35] (2007, Mozambique)	6-7 13-14	ISAAC	1-year	Urban: physical features building, pavement and, water supply, distance town centre Semi-rural: water supply by wells, no electricity or telephone connections	4.0% 10.9%	4.8% 10.9%
Maymi *et al.* [36] (2007, Puerto Rico)	6-7	Laughter	Lifetime	None	25.8%	23.5%
Nilsson *et al.* [20] (1999, Sweden)	13-14	ISAAC	Lifetime	Rural: area with fewer than 50 households within distance of 200 m from each other and with fewer than 200 residents in total Urban:>10 000 inhabitants in total	28.0%[a, c] 28.5%[a, d]	24.0%[a, c] 22.5%[a, d]
Padegimas *et al.* [37] (1982, Lithuanian)	all	Qnaire	Unk	None	2.9%[a]	1.9%
Saeki *et al.* [38] (2005, Japan)	6-7 11-12	Clin Exam	Point	None	10.9%	11.55
Selcuk *et al.* [39] (1997, Turkey)	7-12	Qnaire	1-year	None	2.1%	2.5%
Solé *et al.* [40] (2007, Brazil)	13-14	ISAAC	Lifetime	Brazilian Institute of geography. Urban: towns, villages, isolated urban areas Rural: out of limits of urban	14.1%[e] 9.8%[f]	14.2%[e] 11.8%[f]
Van de Ven *et al.* [41] (2006, Netherlands)	12-14	ISAAC	1-year	Urban/rural: Item in questionnaire, judged by subjects	9.7%	8.3%
Vedanthan *et al.* [42] (2006, India)	6-16	Qnaire	Lifetime	None	0%	20%
Wolkewitz *et al.* [43] (2007, Germany)	50-74	Qnaire	Lifetime	Village: <10 000 inhabitants Small town 10 000-100 000 inhabitants Large city: > 100 000 inhabitants	4.7%[g]	4.1%[g]
Yemaneberhan *et al.* [44] (2004, Ethiopia)	All	Qnaire	1-year	None	0.8%	0.2%

Table 2. *Continued*

Author (year of publication, country)	Age (y)	Diagnosis of eczema	Prev estimate	Definition of Urban and rural	Prev Urban	Prev Rural
Yu *et al.* [45] (2005, Taiwan)	7-15	Qnaire	Unk	None	3.4%	3.3%
Zeng *et al.* [46] (2006, China)	0-6	ISAAC	1-year	None	3.5%	2.4%

Qnaire; Questionnaire, Clin Exam; Clinical Examination, H&R criteria; Hanifin and Rajka diagnostic criteria, Laughter; Laughter questionnaire, Pop; population Prev; prevalence, unk; unknown, UK; U.K. Working party criteria

[a], data extracted from figure or calculated.

[b], from personal communication with authors (H Williams, L. Bråbäck)

[c], place of residence in first year of life

[d], place of residence in second year of life

[e], Area of Caruaru

[f], Area of Santa Maria

[g], place of residence in first 18 years of life

With respect to assigning the degree of urbanization to subjects, all but two articles used the current place of residence: Nilsson *et al.*[20] used the place of residence of subjects in their first and second year of life and Wolkewitz *et al.*[43] used place of residence before 18 years of age.

Methodological quality

The methodological quality varied considerably between studies (Table 3). One study did not define the target population.[34] The sampling methods, if employed, were adequate in most cases, except for one[42] and the sampling method was not reported in another.[37] Graif *et al.*[19] excluded some ethnic groups from their analysis. Two studies did not meet our criterion of > 810.[34,42] Six studies did not report a response rate[18,27,34,37,40,43] and three of the reported rates were not adequate.[30,33,36] Information on non-responders was given in only two studies.[19,30] Valid and repeatable disease definitions were used in most of the studies, although eight studies failed to provide any definition.[18,26,30,34,37,43-45] Two studies used invalid diagnostic criteria.[39,41] Efforts to reduce observer bias (when clinical examination or interviewing was used to determine diagnosis) were employed in only one of the seven studies.[29]

Thirteen studies could be indicated as good quality studies, nine of which were conducted in 'developed' countries.

Prevalence studies

Prevailing prevalence rates for urban and rural residency are shown in Table 2. In 20 studies the prevalence of eczema was higher in urban areas than in rural areas, in eight studies lower and equal in one study. Twenty-six studies were involved in the risk analysis as three studies were excluded because lack of data meant it was not possible to calculate RRs. Figure 2 shows the RR of eczema in urban vs. rural areas. Nineteen of the analysed studies showed a higher risk of having eczema in an urbanized area, 11 of which were statistically significant. Six articles showed a lower risk of having eczema in an urbanized area, of which only one was significant. One study had a RR of 1.00. When dividing the studies according to level of development of the countries, it is shown that in 'developed/western' countries only two of the 14 studies yielded a lower risk of eczema in urban areas; in 'developing' countries it was four out of 12 studies. Kim *et al.*[31] showed no statistical difference in the total group of patients; however, they did find a statistically significant result in the first-graders group.

Nilsson *et al.*[20] compared prevalences in 13- and 14-year olds between place of residence in the first and second year of life and found that there

Table 3. Methodological quality of included articles

Author	Target population	Sampling methods	Sample size	Response rate	Information on non-responders	valid disease definitions	observer bias	Overall quality
Bouayad *et al.*	+	+	+	+	-	+	na	Good
Bråbäck *et al.*	+	+	+	+	-	+	-	Good
Chalmers *et al.*	+	+	+	+	-	+	-	Good
Du Prel *et al.*	+	+	+	+	-	+	-	Good
Dutau *et al.*	+	-	+	unk	-	-	na	Low
Galassi *et al.*	+	+	+	+	-	+	na	Good
Gniazdowska *et al.*	+	+	+	+	-	-	na	Low
Graif *et al.*	-	-	+	+	+	+	na	Low
Haileamlak *et al.*	+	+	+	unk	-	+	na	Low
Hanifin *et al.*	+	+	+	+	-	+	na	Good
Heinrich *et al.*	+	+	+	+	-	+	+	Good
Iversen *et al.*	+	+	+	-	+	-	na	Low
Kim *et al.*	+	+	+	+	-	+	-	Good
Kuhlich *et al.*	+	+	+	+	-	+	na	Good
Laughter *et al.*	+	+	+	-	-	+	na	Low
Lynch *et al.*	-	-	-	unk	-	-	-	Low
Mavale-Manuel *et al.*	+	+	+	+	-	+	na	Good
Maymi *et al.*	+	-	+	-	-	+	na	Low
Nilsson *et al.*	+	+	+	+	-	+	na	Good
Padegimas *et al.*	+	unk	+	unk	+	-	na	Low
Saeki *et al.*	+	+	+	+	-	+	-	Good
Selcuk *et al.*	+	+	+	+	-	-	na	Low
Solé *et al.*	+	+	+	unk	-	+	na	Low
Van de Ven *et al.*	+	+	+	+	-	-	na	Low
Vedanthan *et al.*	+	-	-	+	-	+	na	Low
Wolkewitz *et al.*	+	+	+	unk	-	-	na	Low
Yemaneberhan *et al.*	+	+	+	+	-	-	na	Low
Yu *et al.*	+	+	+	+	-	-	na	Low
Zeng *et al.*	+	+	+	+	-	+	na	Good

na; not applicable, unk; unknown, +; adequate, -; not adequate. See methods section for details on the criteria.

was a significant (P<0.001) difference between urban and rural residency. Also supported by the results of Kim *et al.*,[31] they found a statistical difference between urban and rural residency in the first-graders group exclusively. Grouping the included studies into studies conducted with children or adults did not show a difference in overall RRs.

Some studies were not consistent with an overall tendency favouring urban prevalence over rural areas. A clear example of this was the study by Vedanthan *et al.*[42] with a RR of 0.05 (95% CI 0.00-0.79). This was probably due to sampling variation from the very small and inconclusive sample size. Lynch *et al.*[34] and Haileamlak *et al.*[27] also showed a lower prevalence (although statistically not significant) in urban over rural areas, of which the first study was characterized by a very low overall methodological quality. Of the 'developed' countries, two studies also showed an increased prevalence in rural areas, but this was less prominent and not significant.[38,39] Of these two studies, one used no valid criteria, so methodological quality was doubtful.[39] An outlier in favour of an increased

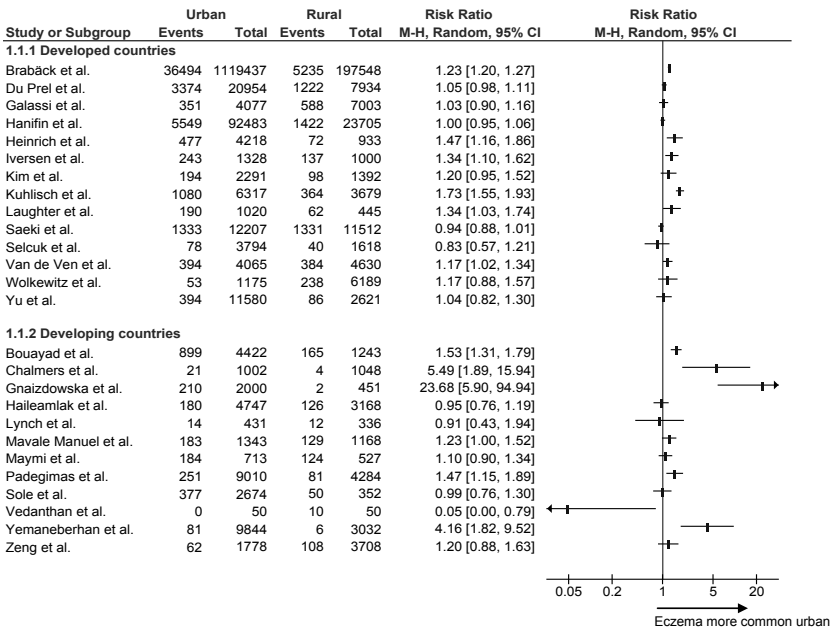


Figure 2. Relative risks for eczema in an urban area compared with rural area. CI, confidence interval.

urban prevalence was the study of Gnaizdowska and Jefimow[26] with a RR of 23.68 (95% CI 5.90-94.94). This study was prone to bias because disease definition was not given and criteria were not validated. Two other (less prominent) outliers were by Chalmers *et al.*[23] and Yemaneberhan *et al*.[44] In contrast with the other study, Chalmers *et al.*[23] was a methodologically sound study.

When stratifying the studies by study quality, a slightly more stable outcome can be seen among the studies of good quality.

When using a random-effect model to estimate the cumulative magnitude of the rural/urban gradient over the developed countries, developing countries and the total group, a high statistical heterogeneity was found ($I^2 > 60$) for all groups. This was due to the richness of the data and precluded us showing a cumulative RR. When focusing on the high-quality articles exclusively, the statistical heterogeneity was still too high for data pooling. Also, grouping a selection of studies, for instance by country or age of the subjects, did not result in an acceptable heterogeneity for pooling. A factor that has led to a high statistical heterogeneity was the small coincidence intervals due to large patient populations used. Clinical and methodological heterogeneity were also contributing factors.

DISCUSSION

Main findings

This systematic review provides some evidence for a higher risk of eczema in urban over rural residential areas. This effect was even more evident in the 'developed/western' countries compared with 'developing' countries.

Overall this might indicate that exposure to environmental factors is of importance and adds to the idea that early sensitization is associated with later atopic manifestations.[47,48]

Quality of the included studies

The methodological quality of the included studies was good in approximately 50%. The older studies were more frequently among those of poor quality. None of the included studies was without limitations. In a substantial number of studies, a valid and repeatable disease definition was lacking. Most often, data on response rate and non-responders of surveys were not reported. Noticeable is that in cases of clinical diagnosis or interviews to determine diagnosis, only

one of the seven studies made efforts to reduce observer bias by training, briefings, etc.

The two largest outliers in this study (Vendanthan *et al.*[42] and Gnaizdowska and Jefimow[26]) scored particularly disappointing on methodological quality. Results of both studies are therefore also likely to be biased and their results might be questioned. Stratification by study quality showed slightly less heterogeneous results among the studies of good methodological quality.

The awareness among researchers of the need for sound methodological design of studies is increasing but we still want to emphasize the importance of adequate design and reporting. Radulescu *et al.*[13] offer clear guidelines for assessing methodological quality in prevalence studies.

Why could eczema be more prevalent in urban areas?

There are several factors that vary between rural and urban areas that might contribute to the differences found in prevalence. Examples of possible factors that contribute to the effect are differences in family size, exposure to animals, maternal age, overcrowding (in a house), differences in food (e.g. processed vs. fresh) and water intake (spring vs. chlorinated water), socioeconomic factors and time spent indoors.[22,48-50] In addition, the amount of traffic in urban areas is notably higher than in rural areas. In contrast, rural areas are prone to have harsher climatic conditions, and these could also play a role in prevalence or severity of eczema.[51]

Pollution, which is usually higher in urban areas, could also be of influence. Dotterud *et al.*[52] found a significantly higher prevalence of eczema in polluted vs. non-polluted areas: RR 3.0 (2.5–3.5) and also Sriyaraj *et al.*[53] found a significantly higher prevalence of eczema in air-polluted areas when comparing urban and semi-urban areas selected for their degree of air pollution. Nilsson *et al.*[20] found that the largest confounding factor was the presence of bronchial asthma followed by parental history of allergy, passive smoking, indoor pets and dampness in the home.

In addition, another explanatory factor that might have contributed is whether the participants were truly atopic or not, as it is possible that IgE sensitization is the predominant factor in explaining the rural/urban gradient for atopic eczema.[54] In none of the included studies was IgE sensitization mandatory for the diagnosis of eczema and the percentage of patients with eczema with IgE sensitization was not given in any study. A possible explanation for the more profound effect found in 'developed/western' countries compared with developing countries could lie in the differences in urban city life. In 'developed' cities, there is a lower exposure to animals, better housing and/or less crowding

compared with cities in 'developing' countries and urban living is therefore more in contrast with rural living. This is also in agreement with the 'hygiene hypothesis'.

Overall, it is difficult to differentiate between factors that contribute to the differences found in prevalences and factors that are confounding.

Strengths and weakness of study

This review was performed following methodologically sound guidelines for systematic reviews made by the Cochrane collaboration. We performed a thorough search and the selection and data-extraction process was done objectively and by two researchers independently.

We assigned countries a label of being 'developed/western' or 'developing'. We think this division is appropriate as environmental conditions between rural and urban areas vary considerably depending on the standard of living in a country. The primary differences between the two groups were that the study populations in 'developed/western' countries were bigger and the RRs more consistent.

We have made the broad assumption that genetic backgrounds will be more or less the same in a given country. As urban locations may be subject to a large influx of migrants and rural locations in some countries may include tribal peoples who are not often found in cities, it is unlikely that the genetic composition of the rural and urban populations within each country is the same.

As some study populations were not just divided into urban and rural areas, but also into semiurban, semirural or industrialized, we had to exclude some comparative areas. In two studies semiurban groups were excluded.[23,43] Padegimas and Dauksiene[37] divided their urban population into two groups: a recently built district and a residential district built 25–30 years ago. We combined both groups into one urban group. In the study of Mavale-Manuel *et al.*,[35] the semirural area was considered rural. The industrialized area of Ulsan was excluded in the analysis of the study by Kim *et al.*[31]

Several relevant studies were excluded for lack of quantitative data, although they did report information on urban vs. rural prevalences of eczema. Kramer *et al.*[55] showed an odds ratio of 0.85 (0.3–2.0) in favour of a lower ratio of eczema in rural communities in Germany, which was in contrast with another study conducted in Germany that stated that there was no significant association between the frequency of eczema and density of urbanization.[56] Poysa *et al.*[57] did not show any comparative data but noted a lower prevalence of eczema in the rural communities of Finland. Also in Finland, Kilpeläinen *et al.*[58] compared urban, rural farming and rural non-farming residency in childhood and stated that there were no significant differences. Furthermore, Aberg *et al.*[59] showed

that there was no significant variation in the presence of eczema between urban and rural Swedish areas.

During this review we encountered data heterogeneity on different levels. Clinical heterogeneity was caused by differences in the characteristics of the populations, age and sample sizes (ranging from 100 to 1 316 985), the definition of eczema employed, diagnostics and prevalence estimates used. The definitions of 'atopic dermatitis/eczema', 'urban' and 'rural' regarded as acceptable for inclusion in the study were very flexible. For instance, what would be defined as being 'urban' according to the definition in one article could be defined as 'rural' in another. Thereby, over the years the differences between urban and rural circumstances have changed. All together, this leads to a considerable degree of heterogeneity. However, by ensuring that comparisons between rural and urban centres in the same study were done using the same method of disease ascertainment this will not lead to distortion of the results. This was the subject of attention while assessing methodological quality. Studies that do not score well on the selection of subjects, response rate or criteria employed are prone to this form of bias.

Only two studies used place of residence at a certain age. All the other studies used current place of residence. As there are indications that exposure to environmental factors early in life is of importance, studies using current place of residence of older subjects could show bias in results to some extent. Patients could have moved from urban to rural areas or vice versa.

Analysis of data was further challenged by statistical heterogeneity. Although 'eye-balling' the outcomes of the studies conducted in 'developed/western' countries suggests low statistical heterogeneity, heterogeneity was high (> 60%). This was primarily due to the large populations involved and thereby small 95% CIs.

For all these reasons, we were not able to pool the collected data and give sum scores or an rural/urban gradient. However, although a cumulative gradient could not be measured, this study does support the idea that urbanization might be a key risk factor for eczema.

We are aware of the presence of new and as yet unpublished data on this matter and we encourage those authors to send us the unpublished work, when it becomes available. The same accounts for any future work or data we missed. This would enable us to maintain an accurate overview in an online database.

Research implications

Future, methodologically sound reviews on environmental circumstances should be done to reveal the factors associated with a higher prevalence of eczema in

urban areas and the association with other allergic diseases. Associating factors could be microbial load, exposure to endotoxins and timing of allergen exposure. If an explanation for the rural/urban gradient could be found, this would help us to unravel the complex aetiology of eczema and enable us to make evidence-based public health decisions on the prevention and even treatment of eczema.

LITERATURE

1. Schultz LF. The epidemiology of atopic dermatitis. *Monogr Allergy* 1993; **31**: 9-28.
2. Williams H, Robertson C, Stewart A *et al*. Worldwide variations in the prevalence of symptoms of atopic eczema in the International Study of Asthma and Allergies in Childhood. *J Allergy Clin Immunol* 1999; **103**: 125-38.
3. Williams H, Stewart A, von Mutius E *et al*. Is eczema really on the increase worldwide? *J Allergy Clin Immunol* 2008; **121**: 947-54.
4. Williams HC. Is the prevalence of atopic dermatitis increasing? *Clin Exp Dermatol* 1992; **17**: 385-91.
5. Williams HC. Atopic eczema. *BMJ* 1995; **311**: 1241-2.
6. Flohr C, Pascoe D, Williams HC. Atopic dermatitis and the 'hygiene hypothesis': too clean to be true? *Br J Dermatol* 2005; **152**: 202-16.
7. Holt PG. Parasites, atopy, and the hygiene hypothesis: resolution of a paradox? *Lancet* 2000; **356**: 1699-701.
8. McKeever TM, Lewis SA, Smith C *et al*. The importance of prenatal exposures on the development of allergic disease: a birth cohort study using the West Midlands General Practice Database. *Am J Respir Crit Care Med* 2002; **166**: 827-32.
9. Sherriff A, Golding J. Hygiene levels in a contemporary population cohort are associated with wheezing and atopic eczema in preschool infants. *Arch Dis Child* 2002; **87**: 26-9.
10. Zar HJ, Ehrlich RI, Workman L *et al*. The changing prevalence of asthma, allergic rhinitis and atopic eczema in African adolescents from 1995 to 2002. *Pediatr Allergy Immunol* 2007; **18**: 560-5.
11. 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.
12. Loney PL, Chambers LW, Bennett KJ *et al*. Critical appraisal of the health research literature: prevalence or incidence of a health problem. *Chronic Dis Can* 1998; **19**: 170-6.
13. Radulescu M, Diepghen T, Williams HC. What makes a good prevalence survey? In: Evidence-Based Dermatology (Williams HC, ed), 2n edn. Oxford: Blackwell Publishing, 2008; 61–7.
14. Jolley DJ. Chance. In: *The Challenge of Dermato-Epidemiology* (Williams,HC, Strachan,DP, eds). Boca Raton: CRC Press, 1997: 61-74.
15. Bobak M. Bias-the silent menace. In: *The Challenge of Dermato-Epidemiology* (Williams,HC, Strachan,DP, eds). Boca raton: CRC Press, 1997: 87-98.
16. Marshall V. Factors affecting response and completion rates in some canadian studies. *Can J Aging* 1987; **6**: 217-27.

17. Williams HC. Defining cases. In: *The Challenge of Dermato-Epidemiology* (Williams,HC, Strachan,DP, eds). Boca raton; CRC Press, 1997: 13-24.
18. Dutau G, Juchet A, Rance F *et al*. Epidemiological survey of asthma and allergies in five-to- six-year-olds in the Midi-Pyrenees region of France. *Annales de Pediatrie* 1997; **44**: 299-306.
19. Graif Y, Garty BZ, Livne I *et al*. Prevalence and risk factors for allergic rhinitis and atopic eczema among schoolchildren in Israel: results from a national study. *Ann Allergy Asthma Immunol* 2004; **92**: 245-9.
20. Nilsson L, Castor O, Lofman O *et al*. Allergic disease in teenagers in relation to urban or rural residence at various stages of childhood. *Allergy: European Journal of Allergy and Clinical Immunology* 1999; **54**: 716-21.
21. Bouayad Z, Aichane A, Afif A *et al*. Prevalence and trend of self-reported asthma and other allergic disease symptoms in Morocco: ISAAC phase I and III. *Int J Tuberc Lung Dis* 2006; **10**: 371-7.
22. Bråbäck L, Hjern A, Rasmussen F. Trends in asthma, allergic rhinitis and eczema among Swedish conscripts from farming and non-farming environments. A nationwide study over three decades. *Clin Exp Allergy* 2004; **34**: 38-43.
23. 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**: 111-6.
24. du Prel X, Krämer U, Behrendt H *et al*. Preschool children's health and its association with parental education and individual living conditions in East and West Germany. *BMC Public Health* 2006; **6**: 312.
25. Galassi C, De SM, Biggeri A *et al*. Changes in prevalence of asthma and allergies among children and adolescents in Italy: 1994-2002. *Pediatrics* 2006; **117**: 34-42.
26. Gniazdowska B, Jefimow A. Epidemiologic studies on allergic diseases among rural and urban school children in Poland. [in Polish]. *Pol Tyg Lek* 1990; **45**: 855-60.
27. Haileamlak A, Dagoye D, Williams H *et al*. Early life risk factors for atopic dermatitis in Ethiopian children. *J Allergy Clin Immunol* 2005; **115**: 370-6.
28. Hanifin JM, Reed ML, Eczema Prevalence and Impact Working Group. A population-based survey of eczema prevalence in the United States. *Dermatitis* 2007; **18**: 82-91.
29. Heinrich J, Frye C, Holscher B *et al*. Prevalence of hay fever and asthma in children is associated with the size of the residential community. *Allergo Journal* 2001; **10**: 89-94.
30. Iversen L, Hannaford PC, Price DB *et al*. Is living in a rural area good for your respiratory health? Results from a cross-sectional study in Scotland. *Chest* 2005; **128**: 2059-67.
31. Kim CW, Park CJ, Kim JW *et al*. Prevalence of atopic dermatitis in Korea. *Acta Derm Venereol* 2000; **80**: 353-6.
32. Kuhlisch W, Gebert P, Vorwergk S *et al*. Respiratory diseases and allergies in a rural and an urban community in Saxony. *Allergologie* 2000; **23**: 327-33.
33. Laughter D, Istvan JA, Tofte SJ *et al*. The prevalence of atopic dermatitis in Oregon schoolchildren. *J Am Acad Dermatol* 2000; **43**: 649-55.
34. Lynch NR, Medouze L, Di Prisco-Fuenmayor MC *et al*. Incidence of atopic disease in a tropical environment: partial independence from intestinal helminthiasis. *J Allergy Clin Immunol* 1984; **73**: 229-33.
35. Mavale-Manuel S, Joaquim O, Macome C *et al*. Asthma and allergies

in schoolchildren of Maputo. *Allergy* 2007; **62**: 265-71.

36. Maymi MA, Somolinos AL, Nazario CM *et al*. The prevalence of atopic dermatitis in Puerto Rican school children. *P R Health Sci J* 2007; **26**: 127-33.
37. Padegimas B, Dauksiene L. Allergic disorders in urban and rural populations. *Allergol Immunopathol (Madr)* 1982; **10**: 121-4.
38. Saeki H, Iizuka H, Mori Y *et al*. Prevalence of atopic dermatitis in Japanese elementary schoolchildren. *Br J Dermatol* 2005; **152**: 110-4.
39. Selcuk ZT, Caglar T, Enunlu T *et al*. The prevalence of allergic diseases in primary school children in Edirne, Turkey. *Clin Exp Allergy* 1997; **27**: 262-9.
40. Solé D, Cassol VE, Silva AR *et al*. Prevalence of symptoms of asthma, rhinitis, and atopic eczema among adolescents living in urban and rural areas in different regions of Brazil. *Allergol Immunopathol (Madr)* 2007; **35**: 248-53.
41. Van de Ven MOM, van den Eijnden RJ, Engels RC. Atopic diseases and related risk factors among Dutch adolescents. *Eur J Public Health* 2006; **16**: 549-58.
42. Vedanthan PK, Mahesh PA, Vedanthan R *et al*. Effect of animal contact and microbial exposures on the prevalence of atopy and asthma in urban vs rural children in India. *Ann Allergy Asthma Immunol* 2006; **96**: 571-8.
43. Wolkewitz M, Rothenbacher D, Low M *et al*. Lifetime prevalence of self-reported atopic diseases in a population-based sample of elderly subjects: results of the ESTHER study. *Br J Dermatol* 2007; **156**: 693-7.
44. Yemaneberhan H, Flohr C, Lewis SA *et al*. Prevalence and associated factors of atopic dermatitis symptoms in rural and urban Ethiopia. *Clin Exp Allergy* 2004; **34**: 779-85.
45. Yu JH, Lue KH, Lu KH *et al*. The relationship of air pollution to the prevalence of allergic diseases in Taichung and Chu-Shan in 2002. *J Microbiol Immunol Infect* 2005; **38**: 123-6.
46. Zeng SW, Coenraads PJ, Tang NJ *et al*. Survey on prevalence of atopic dermatitis in 0-6 year-old children in Taijin. Chin J Dermatol 39(4), 184. 2006.
47. Hattevig G, Kjellman B, Bjorksten B. Appearance of IgE antibodies to ingested and inhaled allergens during the first 12 years of life in atopic and non-atopic children. *Pediatr Allergy Immunol* 1993; **4**: 182-6.
48. Nilsson L, Bjorksten B, Hattevig G *et al*. Season of birth as predictor of atopic manifestations. *Arch Dis Child* 1997; **76**: 341-4.
49. Berglund M, Braback L, Bylin G *et al*. Personal NO2 exposure monitoring shows high exposure among ice-skating schoolchildren. *Arch Environ Health* 1994; **49**: 17-24.
50. Schafer T, Vieluf D, Behrendt H *et al*. Atopic eczema and other manifestations of atopy: results of a study in East and West Germany. *Allergy* 1996; **51**: 532-9.
51. Guiote-Dominguez MV, Munoz-Hoyos A, Gutierrez-Salmeron MT. [Prevalence of atopic dermatitis in schoolchildren in granada, Spain.]. [Spanish]. *Actas Dermosifiliogr* 2008; **99**: 628-38.
52. Dotterud LK, Odland JO, Falk ES. Atopic dermatitis and respiratory symptoms in Russian and northern Norwegian school children: a comparison study in two arctic areas and the impact of environmental factors. *J Eur Acad Dermatol Venereol* 2004; **18**: 131-6.
53. Sriyaraj K, Priest N, Shutes B. Environmental factors influencing the prevalence of respiratory diseases and allergies among schoolchildren

in Chiang Mai, Thailand. *Int J Environ Health Res* 2008; **18**: 129-48.

54. Flohr C, Weiland SK, Weinmayr G *et al*. The role of atopic sensitization in flexural eczema: findings from the International Study of Asthma and Allergies in Childhood Phase Two. *J Allergy Clin Immunol* 2008; **121**: 141-7.
55. 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.
56. Werner S, Buser K, Kapp A *et al*. The incidence of atopic dermatitis in school entrants is associated with individual life-style factors but not with local environmental factors in Hannover, Germany. *Br J Dermatol* 2002; **147**: 95-104.
57. Poysa L, Korppi M, Pietikainen M *et al*. Asthma, allergic rhinitis and atopic eczema in Finnish children and adolescents. *Allergy* 1991; **46**: 161-5.
58. Kilpelainen M, Terho EO, Helenius H *et al*. Farm environment in childhood prevents the development of allergies. [see comment]. *Clin Exp Allergy* 2000; **30**: 201-8.
59. Aberg N, Engstrom I, Lindberg U. Allergic diseases in Swedish school children. *Acta Paediatr Scand* 1989; **78**: 246-52.

EEA Brenninkmeijer

ME Schram

MMG Leeflang

JD Bos

PhI Spuls

Diagnostic criteria for atopic dermatitis: a systematic review.
British journal Dermatology 2008;158:754-765

3.1

DIAGNOSTIC CRITERIA FOR ATOPIC DERMATITIS: A SYSTEMATIC REVIEW

SUMMARY

Rationale: Atopic dermatitis/eczema (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.

INTRODUCTION

Atopic dermatitis/eczema (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.[11] 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)
Diepgen criteria, 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, 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 population-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 and date of publication was not a limitation.

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 occurence[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

Continued on next page

Table 2. *Continued*

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

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 AD yielded no additional, relevant articles and were therefore not mentioned in the search strategy. 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]

Table 3. The Quality Assessment of Diagnostic Accuracy tool (QUADAS) by Whriting 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?

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

RESULTS

Results of the literature search

An initial search retrieved 925 articles. After screening titles and abstracts for eligibility, 47 articles were selected. Figure 1 summarizes the selection process for studies on diagnostic criteria for AD. 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

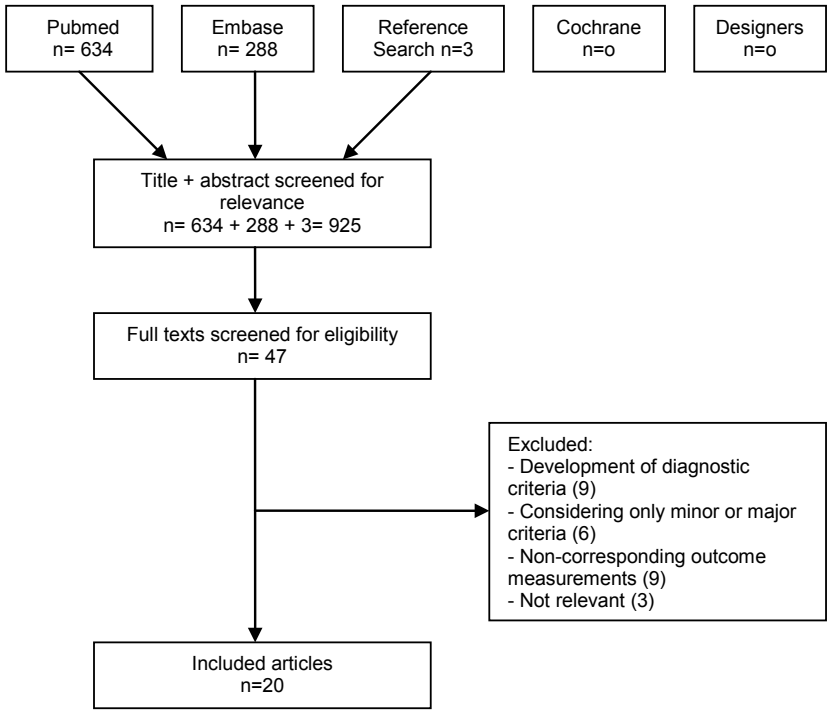


Figure 1. Flowchart summarizing the selection process for studies on diagnostic criteria for atopic dermatitis.

designers of the diagnostic criteria yielded no additional articles. No additional articles were found in the Cochrane databases. Of the initial 47 selected articles, 27 publications were excluded. Of these, 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 measures[5,30–37] and three studies proved not to be relevant.[17,38,39]

Study description

Twenty articles describing 27 validation studies, 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, 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 93.1% 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

Table 4. Key Characteristics of included studies

Criteria /reference	Reference standard	Study design	Setting	Age (year)	Numbers (case/ control)	AD prev	Country	Translation
U.K. diagnostic criteria								
Williams *et al.* (1994) I	Clinical diagnosis	Cross-sectional	Hos	≤ 10, > 10	200	–	England	–
Williams *et al.* (1994) II	Clinical diagnosis	Cross-sectional	Hos	< 16	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	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	1114	2.4	Romania	Romanian (verified)
Mohrenschlager *et al.* (1998)	Clinical diagnosis	Case–control	Pop	8–9	373 (43/330)	12.0	Germany	German (verified)
Firooz *et al.* (1999)	Clinical diagnosis	Cross-sectional	Hos	< 4, 4–10, > 10	416	–	Iran	Not reported
Marks *et al.* (1999)	Clinical diagnosis	Cross-sectional	Pop	4–18	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	61 (31/30)	–	Denmark	Danish (verified)
Fleming *et al.* (2001)	U.K.	Case–control	Pop	1	118 (59/59)	–	Scotland	–
Ortiz *et al.* (2003)	Clinical diagnosis	Cross-sectional	Hos	3–17	874	7.1	Spain	Spanish (verified)
Girolomoni *et al.* (2003)	Clinical diagnosis	Cross-sectional	Pop	9	1331	5.8	Italy	Not reported
Hamada *et al.* (2005)	JDA	Cross-sectional	Pop	≤ 5	565	6.9	Japan	Japanese (not verified)
Haileamlak *et al.* (2005)	Clinical diagnosis	Case–control	Pop	1–5	(93/433) (140/592)	4.4	Southwest Ethiopia	Amharic (verified)

De *et al.* (2006)	Clinical diagnosis	Case–control	Hos	≤15	149 (101/48)	–	North India	Not reported
Chalmers *et al.* (2006)	Clinical diagnosis	Cross-sectional	Pop	3–11	3069	1.0	South Africa	Xhosa (verified)
Saeki *et al.* (2007)	JDA	Cross-sectional	Pop	6–7 and 11–12	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	149 (101/48)	–	North India	Not reported
Schulz-Larsen criteria								
Schultz Larsen *et al.* (1996)	H&R	Case–control	Hos/ Pop	6–12	Unknown	15.6	North Europe	Not reported
Laughter *et al.* (2000)	Clinical diagnosis	Case–control	Hos	5–9	67 (18/49)	–	USA	–
Diepgen criteria								
Diepgen *et al.* (1996)	Clinical diagnosis	Case–control	Hos	10–55	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	(93/433) (140/592)	4.4	Southwest Ethiopia	Amharic (verified)

AD prev; prevalence of atopic dermatitis, JDA; Japanese Dermatology Association diagnostic criteria, H&R, Hanifin and Rajka diagnostic criteria, UK, U.K. 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.

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 (Figures 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 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 78% to 100%, respectively, in hospital-based studies and from 43% to 100% and from 45% to 97%, 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

Table 6. Results for hospital-based studies

	Sensitivity		Specificity		PPV		NPV	
Criteria / reference	**Absolute numbers**	**%**	**Absolute numbers**	**%**	**Absolute numbers**	**%**	**Absolute numbers**	**%**
U.K. 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	6/21b	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 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	97.0	45/49	91.8
Schultz-Larsen criteria								
Schultz Larsen et al. (1996)		88.0		89.0				
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 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

	Sensitivity		Specificity		PPV		NPV	
Criteria / reference	**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.

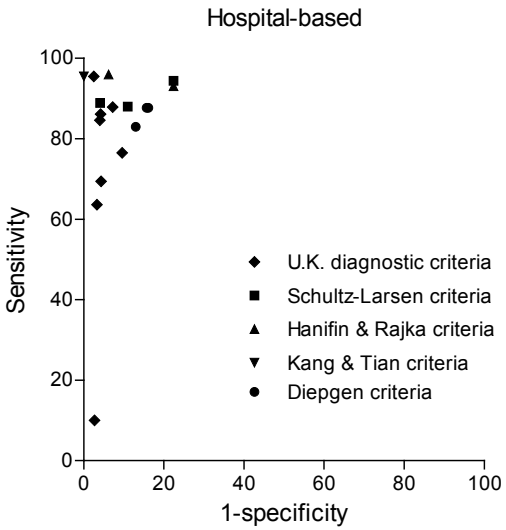


Figure 2. Receiver operating characteristics plot for hospital-based studies.

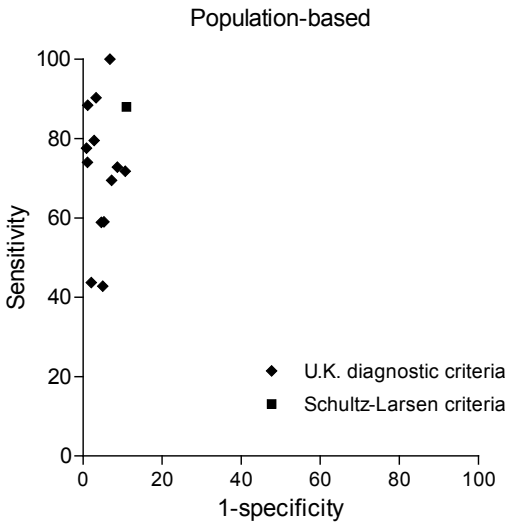


Figure 3. Receiver operating characteristics plot for population-based studies.

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 diagnostic criteria for AD.[8] These criteria consist of one mandatory and five major criteria. The criteria are all non-invasive 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 fulfillment 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 centers 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 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, non-validated 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. 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 could be 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 non-native 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 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-Larsen, 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.

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? 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. Mohrenschlager M, Schafer H, Williams H et al. Ubersetzung and Validiering der britischen Diagnosekriterien des atopischen Ekzem bei acht- und neunja¨rigen 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. 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.

ME Schram
MMG Leeflang
JS den Ottolander
PhI Spuls
JD Bos

Validation and refinement of the millennium criteria for atopic dermatitis
Journal of Dermatology 2011, in press

3.2 VALIDATION AND REFINEMENT OF THE MILLENNIUM CRITERIA FOR ATOPIC DERMATITIS

SUMMARY

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 are 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.

INTRODUCTION

Atopic dermatitis/eczema (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] Chapter 3.1 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]

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 IgE 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. The dermatologists 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 (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, except for a few, individual criteria of the MC, UKC and H&RC were assessed during the examination by the study investigators. "White dermographism" was excluded because it was difficult to interpret, 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 had a diagnostic odds ratio [dOR] $\geq$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

Approximately 650 new patients attended 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 Phadiatop 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)
Gender, *N* (%)			
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 dermatitos, AFD; atopiform dermatitis, N; number, SCORAD; severity scoring of atopic dermatitis, 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 were 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% (data not shown).

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 1 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 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

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 allergic 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	-	+

Table 2. *Continued*

	Total (n=210)	AD (n=44)	PhenAD (n=63)	Other (n=147)	Sens	Spec	dOR (95% CI)	P-value	Prev Analysis	In Analysis*
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 Brenninkmeijer *et al*[19]., sens; sensitivity, spec; specificity
*Odds ratio≥3

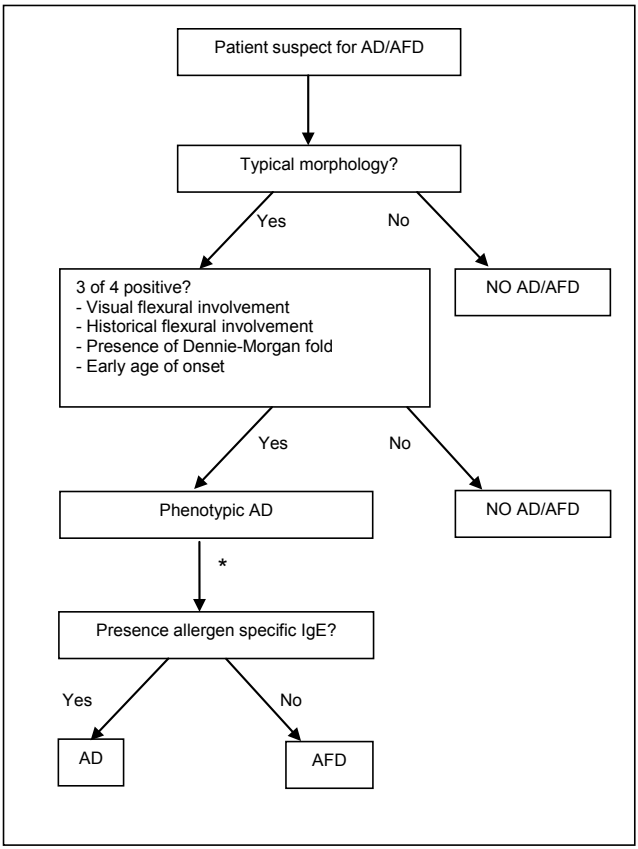


Figure 1. The refined Millennium Criteria; flowchart.
* This step might be postponed in infantile AD.

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.

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, NA; not applicable, rMC; refined Millennium criteria, RV; relative value, sens; sensitivity, spec; specificity, UKC; U.K. working party criteria
+/- IgE; With or without the criteria for the presence of allergen specific IgE

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 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. On the other hand, 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&RC 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 are missed and patients are 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.

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 distinction 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.

ACKNOWLEDGEMENTS

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**: 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**: 1526–1530.
4. Levy RM, Gelfand JM, Yan AC. The epidemiology of atopic dermatitis. Clin Dermatol 2003; **21**: 109–115.
5. Williams H, Flohr C. How epidemiology has challenged 3 prevailing concepts about atopic dermatitis. J Allergy Clin Immunol 2006; **118**: 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**: 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 dlC, 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, Sillevis 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 thediagnosis 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.

ME Schram

PhI Spuls

MMG Leeflang

R Lindeboom

JD Bos

J Schmitt

(Objective) SCORAD, EASI & POEM for atopic dermatitis: responsiveness and minimal clinially important difference
Allergy 2011, in press

4

(OBJECTIVE) SCORAD, EASI & POEM FOR ATOPIC DERMATITIS: RESPONSIVENESS AND MINIMAL CLINICALLY IMPORTANT DIFFERENCE

SUMMARY

Rationale: Demonstration of adequate reliability and validity is sufficient for concluding that an instrument is applicable for descriptive and predictive purposes, but before we can confidently use an outcome measure in clinical trials the responsiveness (synonymous with sensitivity to change) and minimal clinically important difference should be known.

Objective: With this study we aimed to assess responsiveness and minimal clinically important difference of four outcome measures used in atopic dermatitis; the Severity Scoring of Atopic Dermatitis (SCORAD), the objective SCORAD, Eczema Area and Severity Index (EASI) and the Patient-Oriented Eczema Measure (POEM).

Methods: Data of three randomized controlled trials were used. To demonstrate responsiveness, we performed receiver operating characteristic (ROC) curves. Minimal clinically important difference was estimated using mean change scores of patients that showed a relevant improvement. Bland & Altman methods were used to quantify the limits of agreement.

Results: Area under the ROC-curve for the SCORAD was 0.70 (95% confidence interval (CI) 0.61-0.78), for the objective SCORAD 0.73 (95%CI 0.70-0.77), for the EASI 0.67 (95%CI 0.60-0.76) and for the POEM 0.67 (95%CI 0.59-0.75). Scores above 0.70 represent a fair responsiveness. The minimal clinically important difference was 8.7 points for the SCORAD, 8.2 for the objective SCORAD, 6.6 for the EASI and 3.4 for the POEM.

Conclusion: The objective SCORAD and SCORAD showed a fair responsiveness. The minimal clinically important differences are established. They are an important prerequisite for the interpretation of published eczema trials and for the planning / sample size estimation of future trials.

BACKGROUND AND RATIONALE

As reliable laboratory tests to assess and monitor disease severity of atopic dermatitis/eczema (AD) are not available yet, clinical outcome measures were proposed. Until now, twenty clinical outcome measures have been inaugurated to measure the signs and symptoms of AD. However, a recently published systematic review[1] indicated that only three of these measures have shown adequate validity and reliability; the Severity Scoring of Atopic Dermatitis (SCORAD), of which there is a variant that is called the objective SCORAD, the Eczema Area and Severity Index (EASI) and the Patient-Oriented Eczema Measure (POEM).[2-4]

One significant limitation of these outcome measures, however, is the lack of evidence concerning responsiveness. Also, it is still unclear what change in scores represents a clinical relevant change. Demonstration of adequate validity and reliability is sufficient for concluding that an instrument is applicable for descriptive and predictive purposes, but before we can confidently use an outcome measure in clinical trials the responsiveness and minimal clinically important difference (MCID) should be known.[5;6]

Responsiveness has been demonstrated in one study for the EASI and SCORAD, but not for the objective SCORAD and POEM.[7] MCID is unknown for all the outcome measures. Therefore, we aimed to assess and compare the responsiveness of the (objective) SCORAD, EASI and POEM with data of three recently performed clinical trials and to estimate their MCID.

PATIENTS & METHODS

Data

For this study we used the clinical outcome data of three investigator-initiated randomized controlled trials (RCT's) on the efficacy of treatments for AD.

The first RCT (further specified as MAcAD) compared methotrexate versus azathioprine in adult patients with severe AD and was performed in the Academic Medical Center, University of Amsterdam, the Netherlands.[8] It concerned a single blind, parallel-group trial in which 42 subjects were recruited and randomized in a 1:1 ratio. The total study duration was 24 weeks: 12 weeks active treatment and 12 weeks follow up in which responders could continue or stop their study medication. At week 0, 2, 4, 8, 12 and 24 of the trial, (objective) SCORAD, EASI, POEM, investigator global assessment (IGA) and patients global assessment (PGA) were assessed.

The second group of subjects (n=38) was recruited for a parallel-group double blinded RCT (PROVE) comparing prednisolone with cyclosporin for severe AD conducted at the Carl Gustav Carus Medical Faculty, Technical University of Dresden, Germany and three other centres in Germany (Kiel, Münster, Hamburg).[9] In the prednisolone group patients were treated for three weeks with prednisolone and then four weeks with placebo. In the cyclosporin group patients were treated for six weeks. Duration of follow-up was 12 weeks in which no medication was given. At week 0, 2, 4, 6, 10, 14 and 18 the objective SCORAD, POEM, IGA and PGA were assessed.

Main inclusion criteria for both trials were age ≥ 18 years, state of general good health, severe eczema (objective SCORAD ≥ 40 or severe according to Rajka and Langeland criteria).[10]

In the third RCT (TASCO), which was performed at the Carl Gustav Carus Medical Faculty, Technical University of Dresden, Germany, an evidence-based treatment algorithm was compared with individualized symptom-oriented treatment in 63 subjects with AD aged two years or older, who had at least moderate AD (SCORAD > 20).[11] Duration of the study was one year in which the objective SCORAD and IGA were assessed every four weeks. Treatment regimens included topical therapy (corticosteroids and calcineurine inhibitors), phototherapy and systemic therapy (corticosteroids and cyclosporin).

All patients were diagnosed with AD by an experienced dermatologist and by the use of the U.K. Working Party Diagnostic criteria.[12] Approval by the ethics committees was obtained for all studies and all patients signed a written informed consent prior to study related procedures.

Outcome measures

The SCORAD combines both 'objective' items as affected area and intensity of the lesions, and 'subjective' items as extent of pruritus and sleep loss. The 'objective' part of the SCORAD is referred to as 'objective SCORAD' and is often used alone. To calculate the SCORAD, affected area is scored by applying *the rule of nine* after drawing the lesions on an evaluation form. The intensity is determined by grading each intensity item (erythema, oedema/papulation, oozing/crusts, excoriation, lichenification and dryness) on a scale of 0 to 3. Pruritus and sleeploss were rated by the patient on a visual analogue scale. The total score is the sum of extent/5 + 7 x intensity/2 + sum of symptoms. Range of the scores is from 0 to 103 for the SCORAD and from 0 to 83 for the objective SCORAD. Over the years, SCORAD has been validated in several studies.[1]

The EASI combines affected area and intensity of the lesions.[13;14] For affected area, the EASI assigns proportionate body surface areas to the head (10%),

trunk (30%), upper extremities (20%), and lower extremities (40%). Intensity of the lesions is determined by 4 items (erythema, infiltration/population, excoriation and lichenification) in each particular region (head, trunk, upper and lower extremities). Each item is assessed on a scale from 0 to 3 for each region. The region score is a product of the total intensity score for that region and the proportional area score. The total score is a sum of each region score and can range from 0 to 72. To date, the EASI has been validated in several studies, but adequate evidence concerning sensitivity to change is missing.[1]

Both SCORAD and EASI showed adequate convergent and divergent construct validity according to those validation studies. Thereby, it should be noted that the (objective) SCORAD and EASI measure extent in different ways.[15]

The POEM was inaugurated in 2004 by Charman *et al.* to enable patient-oriented assessment of disease severity.[3] The POEM questionnaire consists of seven questions regarding the amount of days during the last week that patient experienced itch, sleep loss, bleeding, cracking weeping, flaking of the skin and dry skin. The number of days per item can represent a score from 0 to 4. The total score is a sum of all individual scores. Maximum score is 28 points. From the measures investigated here, the POEM is the only outcome measurement that has been documented to show adequate internal consistency.[3]

IGA and PGA were scored using a six-point Likert scale; 0= clear, 1= almost clear, 2= mild disease, 3= moderate disease, 4= severe disease and 5= very severe disease. The IGA has a close correlation with both the EASI and the SCORAD.[7;16] The POEM correlated well with the PGA.[3] In our study the correlation between IGA and EASI was 0.83, between IGA and (objective) SCORAD 0.85 and between PGA and POEM 0.65.

The POEM questionnaire was translated into Dutch and German. The process involved forward and back-translation per standard protocol. Since all investigators had a thorough command of English language, the other outcome measures were not translated.

Assessments

Outcome assessors of the included RCTs were instructed to use the outcome measures prior to the studies by written details and/or oral presentation. In the MAcAD and PROVE trial, outcome assessors were nurses and trial doctors who were blinded for treatment allocation. Evaluation of patients was conducted under standardized conditions in each study. All patients in the different groups were instructed not to apply any creams or ointments to their eczema directly before assessment. The sum scores were calculated in SPSS 18.0.

Statistical analysis

For measuring the MCID and responsiveness of the EASI and SCORAD, data from the MAcAD trial were used. For the POEM, data from the PROVE and MAcAD trial were used. The MCID and responsiveness of the objective SCORAD was measured by pooled data from the MAcAD and PROVE trial and by data from the TASCO trial concerning adults and children (<18 years).

PGA and IGA were used as anchors/global ratings of change in our study. A clinically relevant change was predefined as an 'improvement' or 'decline' of $\geq$ 1 in PGA and IGA scores since it would be highly likely that a clinical relevant difference occurs in health status if the patients changes one point on the global assessment.

Responsiveness (synonym: sensitivity to change) is the ability of an outcome measure to detect change over time in the construct to be measured and can be seen as a form of longitudinal validity.[17] Firstly, the mean scores of the global assessments (reference test) were examined in relation to the mean scores of the outcome measure of interest (index test) per time-point in each treatment group, thereby globally illustrating responsiveness. Secondly, the receiver operating characteristic (ROC) curves - area under the curve (AUC) was used to estimate responsiveness.[18;19] An ROC curve is a graph that plots sensitivity against 1- specificity. The AUC illustrates how well score changes in the outcome measures can discriminate between patients that did or did not change, defined as a 1-point change in global assessment score between time-points. We calculated only improvements versus no improvement, since data on decline were too scarce. If area under the ROC curve is at least 0.70, the responsiveness is considered fair, 0.8 indicates good and 0.9 excellent responsiveness.[20]

The MCID is the smallest change in an outcome measure that represents a clinically relevant outcome. We estimated MCID using three complementary approaches. Primarily, we analyzed the absolute changes observed within individuals during treatment using the global assessment as anchors for clinical change.[21] Anchors for relevant change were PGA for the POEM and IGA for the (objective) SCORAD and EASI. We calculated the mean change score on the outcome measures when an improvement of one point on global assessment was seen. Sensitivity analysis was done by calculating the cut-off point of the ROC-curve at which correct classification (improved versus not improved on global assessment) was optimal i.e. the cut-off with the highest sum of true positive and true negative classifications.

Finally, to assess the expected amount of variability that can be expected in stable patients, we compared the scores within measurement occasions of patients that did not change on the global assessment using the method

of Bland & Altman.[22] The limits of agreement were calculated to judge the absolute magnitude of the measurement error i.e. any score deviation larger than the expected zero. Limits of agreement were calculated as +/- 1.96 x standard deviation of the measurement error. When the measurement error is random, 95% of the deviation scores are expected to fall within 1.96 SD. In the absence of bias, the mean measurement error should approximate zero and have equal variance over the total range of possible scores.

Missing data were minimal (< 2%). Missing observations were excluded from analysis.

Data was analysed by the first author using SPSS 18.0.

RESULTS

Table 1 shows the baseline characteristics of the included patients of the three trials. In total, data of 143 patients were analyzed. Mean age of the patients was approximately 40 years in the MAcAD trial and 30 years in the PROVE trial. In the TASCO trial, 25 children (<18 years) and 38 adults participated with a mean age of 9.4 and 30.9 years respectively. Sex was generally equally distributed among the treatment groups in all trials. In the MAcAD trial, 39 (93%) patients showed a gradual improvement over 6 months. Three (7%) patients showed no change. The PROVE trial was terminated preliminary due to exacerbations after the termination of prednisolone.[9] In the prednisolone group, most patients experiencing rapid improvement within the first 3 weeks and tended to relapse thereafter. In the cyclosporin group, there was improvement until week 6. Overall, 20 (53%) patients improved, 12 (32%) patients showed no change and 6 (16%) patients declined. In the TASCO trial improvement was seen in the first 4 weeks in both groups, after which a stable period was seen. In general, 47 (75%) of the patients improved, 1 (2%) declined and 15 (24%) showed no change.

For the assessment of responsiveness and the MCID of the SCORAD 239 observations were used, for the objective SCORAD 1067, for the EASI 239 and for the POEM 385.

Responsiveness

To illustrate responsiveness, figure 1 shows the mean scores of the outcome measures in relation to either IGA or PGA score during the trials. It is exemplified that all outcome measures reacted to improvements and declines in global severity measure.

Table 1. Demographics and baseline characteristics of the included patients

	MAcAD		PROVE		TASCO	
	Methotrexate (n=20)	**Azathioprine (n=22)**	**Prednisolone (n=21)**	**Cyclosporin (n=17)**	**Algorithm (n=32)**	**Individual (n=31)**
Age in years	43.0 (14.7)	37.0 (14.1)	28.8 (9.9)	30.1 (8.9)	24 (17)	21 (13)
Male sex	10 (50%)	12 (54.5%)	9 (52.4%)	10 (58.8%)	14 (44%)	16 (45%)
Presence of asthma or allergic rhinitis	19 (95%)	18 (82%)	16 (76.2%)	12 (70.6%)	10 (31%)A 19 (59%)R	9 (29%)A 21 (68%)R
Presence of allergen specific IgE	20 (100%)	21 (95%)	-	-	28 (88%)	28 (90%)
Duration of eczema in years	39.8 (16.2)	33.1 (6.8)	-	-	-	-
Disease activity - SCORAD - objective SCORAD - EASI - POEM	57.2 (11.8) 48.9 (13.0) 27.9 (12.3) 19.8 (5.4)	58.4 (10.4) 47.4 (9.2) 30.4 (14.2) 19.6 (4.0)	- 57.6 (10.2) - 22.1 (7.1)	- 54.7 (10.9) - 21.4 (5.4)	- 46.1 (12.1) - -	- 47.1 (19.4) - -

Data are mean (SD) or number (%)
A: asthma, n: number, R: allergic rhinitis, -: not provided

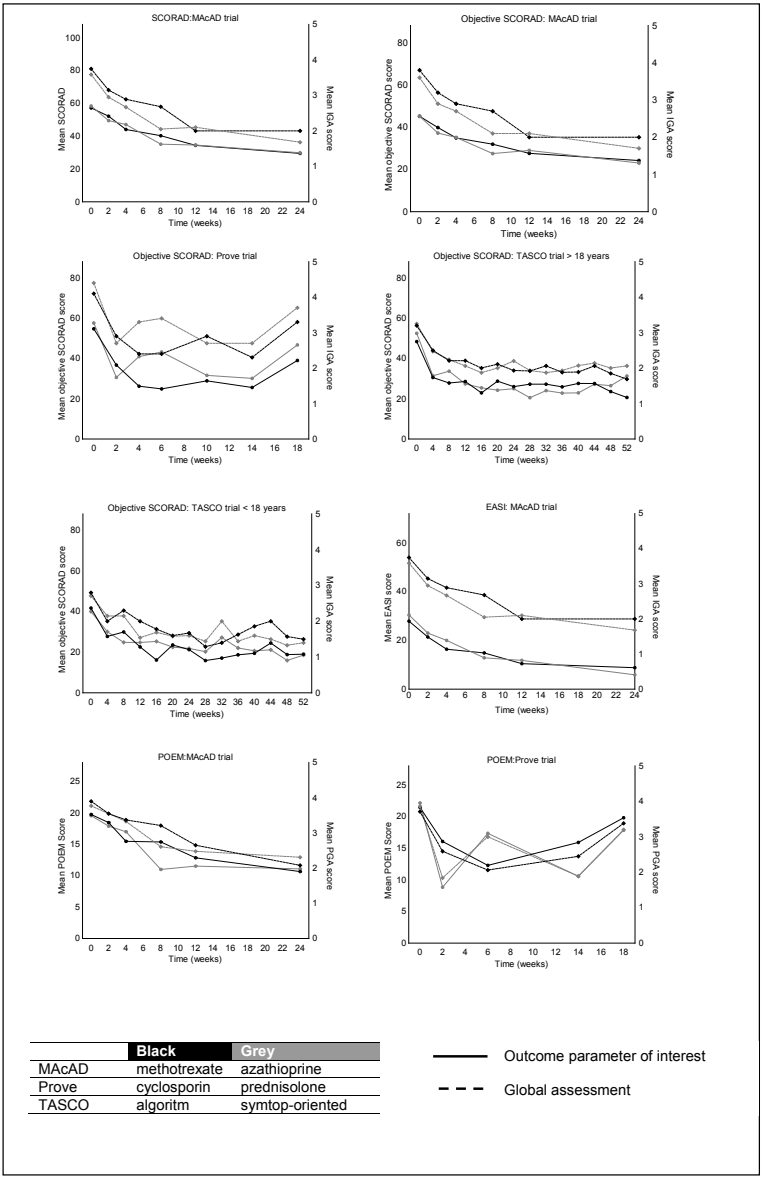


Figure 1. Responsiveness: (objective) SCORAD, EASI and POEM against global assessment.

The SCORAD showed an AUC of 0.70 (95% CI 0.61-0.78) (Table 2). The AUC of objective SCORAD was 0.73 (95% CI 0.70-0.77). The AUC of the EASI is 0.67 (95% CI 0.60-0.76) and the AUC of the POEM is 0.67 (95% CI 0.59-0.75). Although the AUCs of the EASI and POEM seem to be somewhat lower than the AUCs of the SCORADs, the confidence intervals largely overlap. We did separate subgroup analysis for the objective SCORAD, in children and adults. The AUC of the objective SCORAD for adults was 0.73 (95% CI 0.68-0.77) and for children 0.76 (95% CI 0.68-0.84).

Table 2. Responsiveness: area under the curve of ROC

	AUC	95% CI	Cut-off	Sensitivity(95%CI)*	Specificity(95%CI)*
SCORAD	0.70	0.61 - 0.78	4.05	64.7 (51.0 – 76.4)	64.2 (55.8 – 71.8)
Objective SCORAD					
MAcAD & Prove	0.73	0.67 - 0.79	7.00	55.5 (46.5 – 64.1)	80.4 (73.5 – 85.8)
TASCO < 18 years	0.76	0.68 - 0.84	3.45	51.3 (37.7 – 64.1)	82.1 (73.7 – 88.8)
TASCO > 18 years	0.72	0.65 - 0.79	6.55	51.3 (40.4 – 62.1)	82.1 (75.9 – 86.9)
All adults patients	0.73	0.68 – 0.77	6.85	54.8 (47.8 – 61.6)	79.5 (74.6 – 83.2)
All patients	0.73	0.70 – 0.77	6.45	54.8 48.6 – 60.9)	78.8 (74.8 – 82.4)
EASI	0.67	0.60 - 0.76	2.75	73.8 (61.2 - 83.6)	57.4 (46.8 - 67.5)
POEM	0.67	0.59 - 0.75	1.50	62.1 (50.1 – 72.9)	66.1 (57.4 – 73.9)

AUC; area under the curve, CI; confidence interval.
*Sensitivity and specificity reflect the highest correct classification for the cut-off value.

MCID

The anchor-based longitudinal mean difference in change scores for patients with a 1-point improvement in global severity score are shown in Table 3. For the SCORAD, the overall mean difference was 8.7 points (SD 7.8) for an improvement of one point on the IGA. In other words, a mean improvement of 8.7 points on the SCORAD corresponds with a clinically relevant change. The mean difference varied from 6.8 to 12.6 points, when each individual step of improvement on IGA was specified (e.g. from 5 to 4 points on IGA, from 4 to 3, etc). Analysis of the ROC cut-off points for greatest correct classification showed that 4.1 points on the SCORAD was best for identifying a clinically significant improvement with a sensitivity of 64.7% and a specificity of 64.2% (Table 2). The difference in mean score for stable patients on the SCORAD (systematic error/bias) at consecutive visits was 2.2 points (SD 7.8). Limits of agreement ranged from -13.1 to 17.5 points (Figure 2). Six percent of our observations were within this range.

Table 3. Longitudinal mean difference of changes scores.

Outcome measure	RCT	Anchor (IGA/PGA)	*N* of obs.	Mean difference	Min.	Max.	SD
SCORAD	MAcAD	5 -> 4	4	10.8	-1.1	20.9	9.6
		4 -> 3	29	6.8	-9.0	28.3	8.9
		3 -> 2	30	9.8	-3.0	25.6	7.1
		2 -> 1	13	9.5	-0.8	16.3	6.5
		1 ->0	1	12.6	-12.6	12.6	-
		TOTAL	**77**	**8.7**	**-9.0**	**28.3**	**7.8**
Objective SCORAD	MAcAD & PROVE	5 -> 4	8	11.9	2.5	31.0	9.8
		4 -> 3	50	7.7	-11.4	29.0	9.1
		3 -> 2	43	7.9	-7.0	19.5	6.0
		2 -> 1	17	7.5	-3.9	19.5	6.2
		1 ->0	1	10.6	10.6	10.6	.
		TOTAL	**119**	**8.0**	**-11.4**	**31.0**	**7.7**
	TASCO <18 years	5 -> 4	0	-	-	-	-
		4 -> 3	0	-	-	-	-
		3 -> 2	16	12.4	-5.0	37.9	11.1
		2 -> 1	28	7.4	-7.6	25.5	8.5
		1 ->0	7	7.5	1.3	12.2	4.5
		TOTAL	**51**	**9.0**	**-7.6**	**37.9**	**9.2**
	TASCO >18 years	5 -> 4	0	-	-	-	-
		4 -> 3	11	9.9	-6.0	31.6	12.2
		3 -> 2	27	10.4	-22.6	33.0	11.3
		2 -> 1	35	5.8	-9.0	28.8	7.6
		1 ->0	5	6.0	-4.4	17.9	9.2
		TOTAL	**78**	**8.0**	**-22.6**	**33.0**	**9.9**
	All adult patients	5 -> 4	8	11.9	2.5	31.0	9.8
		4 -> 3	61	8.1	-11.4	31.6	9.7
		3 -> 2	70	8.8	-22.6	33.0	8.4
		2 -> 1	52	6.4	-9.0	28.8	7.2
		1 ->0	6	6.8	-4.4	17.9	8.4
		TOTAL	**197**	**8.0**	**-22.6**	**33.0**	**8.6**
	All patients	5 -> 4	8	11.9	2.5	31.0	9.8
		4 -> 3	61	8.1	-11.4	31.6	9.7
		3 -> 2	86	9.5	-22.6	37.9	9.0
		2 -> 1	80	6.7	-9.0	28.8	7.7
		1 ->0	13	7.2	-4.4	17.9	6.3
		TOTAL	**248**	**8.2**	**-22.6**	**37.9**	**8.7**

Table 3. *Continued*

Outcome measure	RCT	Anchor (IGA/PGA)	*N* of obs.	Mean difference	Min.	Max.	SD
EASI	MAcAD	5 -> 4	3	8.6	2.9	18.7	8.8
		4 -> 3	21	7.8	-3.7	24.3	6.9
		3 -> 2	27	7.3	-1.7	21.7	5.7
		2 -> 1	13	3.3	-0.6	6.6	2.4
		1 ->0	1	1.0	1	1	-
		TOTAL	**65**	**6.6**	**-3.7**	**24.3**	**5.9**
POEM	MAcAD & PROVE*	5 -> 4	7	0.1	-5	3	3.1
		4 -> 3	29	4.5	-1	21	5.1
		3 -> 2	25	3.4	-8	18	5.0
		2 -> 1	5	2.0	-2	7	3.2
		1 ->0	2	4.0	3	5	1.4
		TOTAL	**68**	**3.4**	**-8**	**21**	**4.8**

IGA; investigator global assessment, max; maximal change score, min; minimal change score, N; number, obs.; observations, PGA; patient global assessment, SD; standard deviation.
* Anchor by PGA

When pooling all the available data from the different trials, we found an overall mean difference of the objective SCORAD of 8.2 points (SD 8.7) when an improvement in global severity score (IGA) was observed on consecutive visits (Table 3). Taking into account all the observations in adult patients, the mean difference was 8.0 points (SD 8.6). For observations from children, the mean difference was 9.0 (SD 9.2). This was only based on IGA anchor scores ranging from 3 to 0, due to the milder disease of the included children. Overall, analysis of the ROC cut-off points for greatest correct classification showed that 6.5 points on the objective SCORAD was best for identifying a clinically significant improvement with a sensitivity of 54.8% and a specificity of 78.8%. Cut-off points for the different adult populations were similar to that. The cut-off point for the observations of children was 3.5 points. For all patients, the Bland and Altman analysis shows an overall difference in mean score of 1.2 (SD 8.0) points. The limits of agreement ranged from -14.5 to 16.8 (6.4% of the observations within the range). For all adult patients, the same mean score difference was found. The limits of agreement ranged from -14.4 to 16.9 (6.8% of the observations within the range). For observations in children, the mean score difference was 0.7 points (SD 7.1), with a limit of agreement ranging from -13.1 to 14.5 subsequently (7.1% of the observations within the range).

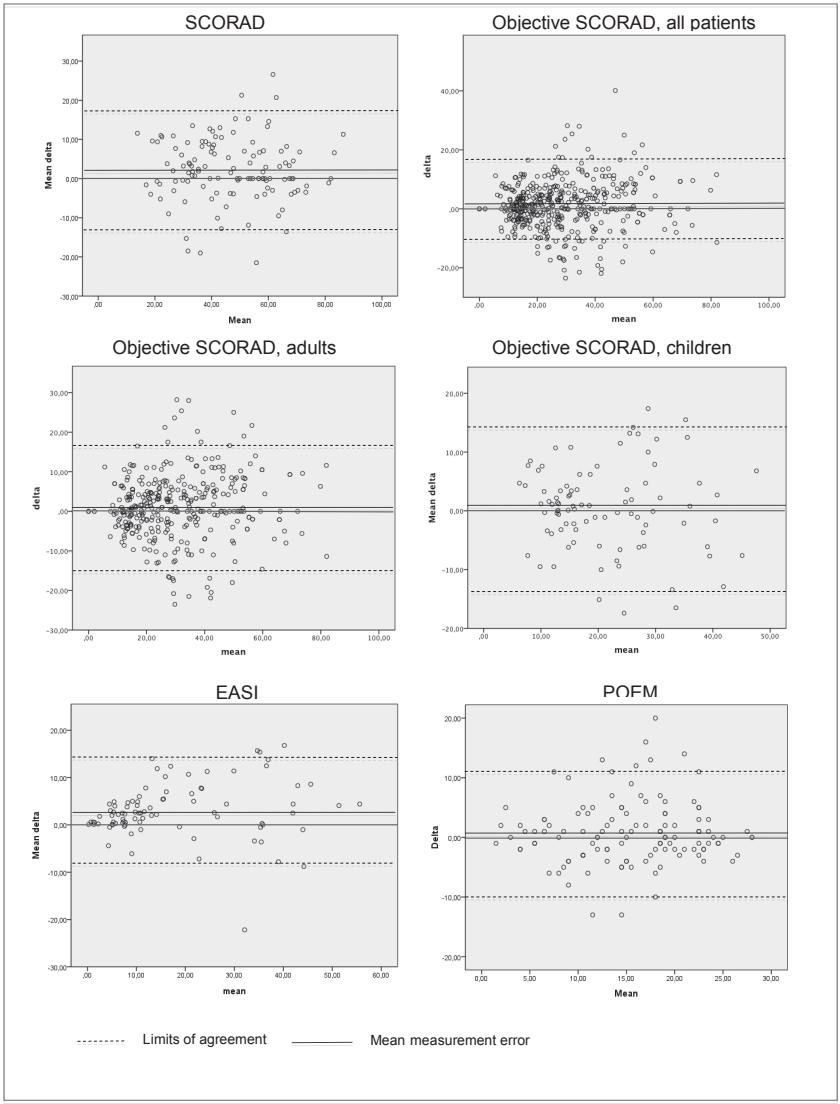


Figure 2. Measurement error by Bland and Altman

For the EASI, the overall mean difference was 6.6 points (SD 5.9) when IGA improved with one point. The MCID varied from 1.0 (anchor from 1 to 0) to 8.6 (anchor from 5 to 4). This range might be due to the small number of observations in both categories. The ROC cut-off point was 2.8 points with 73.8% sensitivity and 57.4% specificity. Measurement error subtracted from the Bland and Altman analysis was 3.0 points (SD 5.8). Limits of agreement ranged from -8.7 to 14.4 points.

The overall mean difference of the POEM was 3.4 points (SD 4.8). The mean difference ranged from 0.1 (anchor from 5 to 4) to 4.5 (anchor from 4 to 3). An improvement of 1.5 points on the POEM was the ROC cut-off point (sensitivity 62.1%, specificity 66.1%). The mean measurement error was 0.8 points (SD 5.3). Therefore, limits of agreement ranged from -9.6 to 11.2 points.

DISCUSSION

Main findings

In this study, we assessed the responsiveness and the MCID of four outcome measures for AD. The results indicate that these measures are all fairly sensitive to change, but that their ability to discriminate between patients who improved and who did not, varies. The responsiveness, or sensitivity to change, of the objective SCORAD and SCORAD was fair, meaning that the area under their ROC curve was larger than 0.70. Although the EASI and the POEM had an AUC below 0.70, the differences in responsiveness were not significant as the confidence intervals overlapped almost entirely.

To calculate the MCID we used two different analytic methods. First, we compared mean differences in the outcome instruments of patients with a 1-point improvement in global assessment score on the consecutive visits. Second, we used the ROC curve to determine the optimal cut-off point for discrimination between those who did and did not have an at least 1- point improvement in global assessment. For the SCORAD we found a mean difference in SCORAD-scores between two subsequent categories of IGA (i.e. one point improvement on the IGA) of 8.7 points. As the SCORAD ranges from 0 to 103, and the IGA ranges from 0 to 5, this means a change about 20% on the IGA only translates to a change of less than 10% on the SCORAD. The optimum cut-off on the ROC curve of the SCORAD was 4.1 points meaning that when considering people with a change of 4.1 SCORAD points or more as being changed, this will result in the highest combination of sensitivity and specificity for detecting change in IGA. For the SCORAD the sensitivity at this optimal

cut-off point was 65% (95% CI 51 to 76%) and the specificity was 64% (95% CI 56 to 72%).

The POEM showed similar sensitivity and specificity at its optimal cut-off point (1.50) and its mean difference was 3.4. Although the objective SCORAD showed higher specificity (79%) at its optimal cut-off (6.5), its sensitivity was lower than that of the other measures (55%). The EASI, on the other hand, showed a higher sensitivity (74%), but a lower specificity (57%) at an optimal cut-off of 2.8. The mean difference for the objective SCORAD was 8.2 points and for the EASI 6.6 points.

Differences in outcomes between the primary and sensitivity analysis could be due to the fact that we measured two slightly different MCIDs. It can be argued that the longitudinal anchor-based mean difference is technically not the most *minimal* difference that is clinically important, but is more a clinically important difference. This explains the trend that the longitudinal mean difference is higher than the optimal cut-off point for best discrimination. We would advice to use the highest approximation of the MCID, since this would be less likely to overestimate clinical meaningfulness of a trial.

When performing power calculations for a parallel-group RCT based on these MCIDs, a group size of 14, 19, 14 or 34 is needed for rejecting the null hypothesis based on the SCORAD, objective SCORAD, EASI or POEM respectively (80% power and 5% significance).

All estimated clinically important differences and optimal cut-offs were within the range of the limits of agreement meaning that individual change scores that are similar to the MCID cannot solely be attributed to a real clinical meaningful change but could also be attributed to measurement error. This is another argument to use the highest approximation of the MCID.

Strengths and weaknesses

This study was performed following the guidelines for performing clinimetric studies by COSMIN (acronym for Consensus-based Standards for the selection of health Measurement INstruments).[23;24] The aim of COSMIN was to establish the definition of the clinimetric properties, standard of design issues and preferred statistical methods and criteria of adequacy.[25;26] The COSMIN does not provide a consensus on how to calculate the MCID. Therefore, we have estimated the MCID using different analytical methods.

Since the COSMIN panel reached consensus that effect sizes and related measures (standardized response mean, relative efficacy statistic, etc.) are inappropriate measures for responsiveness, we have chosen to use the ROC-AUC method. We have only performed ROC's on improvement. Improvement was

defined as improving one point on global assessment, thus improvements of >1 point on global assessment were excluded. This will have resulted in slightly lower outcome on AUC. Since most of the patients improved during the trials, there were too little observations on worsening.

We have chosen to determine the MCID within patients using an anchor-based longitudinal approach. The clear advantage is that change in an outcome parameter is linked to a meaningful external anchor. Disadvantages are that a longitudinal approach can be biased by response shift and that the anchor relies on global ratings.[27] Global ratings used as anchor (gold standard) are broad/not precise and might have a non-linear distribution. For instance, the difference between 'severe' and 'very severe' is smaller than the difference between 'mild' and 'moderate'. In this perspective, some outcomes can be attributed to characteristics of the global assessment rather than to the clinical outcome measure of interest. Also, we used the improvement on the global assessment as anchor for change instead of a transition scale, which is more generally accepted.

Absolute changes to calculate the MCID were used. These changes have the advantage that they are more easily applicable in studies and in practice. Relative changes might be more methodologically sound, but they are not easy to handle on a routine basis.

The limits of agreement found in this study showed a large variation. Although we assumed that the patients had stable disease when global assessment remained stable at consecutive visits, chance variation can never be excluded. Besides, some degree of variation in disease severity could not be excluded due to the long time between visits (4 weeks generally) and because the global assessment measures allow some variation in disease activity within a point on the scale. Furthermore, inter- and intra-rater variability will have played a role, Due to this, we believe that the range of the limits of agreement is an overestimation to some degree.

Generalisability

When using a selected population, it is always questionable whether the study population acts as a representative of the total group of patients, or whether they are mainly a reflection of their current status. We think that the data from the MAcAD and the PROVE trial are representative for all adult patients with chronic severe AD. Data from the TASCO trial also included children. Only the objective SCORAD was calculated with data from children. Those children had generally milder disease than the adult patients that were included. Responsiveness and MCID did not substantially differ between the children

and adults on the objective SCORAD. This might be a positive argument for generalising the outcomes of other outcome measures to children.

It is known that the MCID may vary by population and context and thus a single MCID may be insufficient for all study applications. Thereby, this is the first approximation of the MCID. Confidence in a specific MCID value evolves over time and is confirmed by additional research evidence.

Clinical implications

Currently, the harmonizing outcome measures in eczema (HOME) initiative is undertaken by international specialists on eczema. This is an initiative that resembles the OMERACT for rheumatoid arthritis.[28-30] Its main goal is to define core outcome domains that should be used in every clinical trial concerning AD to enable meta-analysis and thereby improving strength of evidence. In this perspective, it was of great significance to establish the responsiveness of the four selected outcome measures. The quantification of the MCID will offer advantages to analyze and design clinical trials. Based on the MCID, accurate power calculations can be made and it can be decided if statistically significant differences reflect a real change in disease severity to deserve consideration in clinical practise. Demonstrating MCID is also important evidence for achieving successful patient-related outcome claims through regulatory agencies.

REFERENCE LIST

1. Schmitt J, Langan S, Williams HC. What are the best outcome measurements for atopic eczema? A systematic review. *J Allergy Clin Immunol* 2007; **120**: 1389-98.
2. Severity scoring of atopic dermatitis: the SCORAD index. Consensus Report of the European Task Force on Atopic Dermatitis. *Dermatology* 1993; **186**: 23-31.
3. Charman CR, Venn AJ, Williams HC. The patient-oriented eczema measure: development and initial validation of a new tool for measuring atopic eczema severity from the patients' perspective. *Arch Dermatol* 2004; **140**: 1513-9.
4. Tofte SJ, Graeber M, Cherill R *et al*. Eczema area and severity index (EASI): A new tool to evaluate atopic dermatitis. *J Eur Acad Dermatol Venereol* 1998; **11**: S197.
5. Guyatt GH, Deyo RA, Charlson M *et al*. Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol* 1989; **42**: 403-8.
6. Fletcher A. Quality-of-life measurements in the evaluation of treatment: proposed guidelines. *Br J Clin Pharmacol* 1995; **39**: 217-22.
7. Breuer K, Braeutigam M, Kapp A *et al*. Influence of pimecrolimus cream 1% on different morphological signs of eczema in infants with atopic dermatitis. *Dermatology* 2004; **209**: 314-20.

8. Schram ME, Roekevisch E, Leeflang MM *et al*. A randomized trial of methotrexate versus azathioprine for severe atopic eczema. *J Allergy Clin Immunol* 2011.
9. Schmitt J, Schakel K, Folster-Holst R *et al*. Prednisolone vs. ciclosporin for severe adult eczema. An investigator-initiated double-blind placebo-controlled multicentre trial. *Br J Dermatol* 2010; **162**: 661-8.
10. Rajka G, Langeland T. Grading of the severity of atopic dermatitis. *Acta Derm Venereol Suppl (Stockh)* 1989; **144**: 13-4.
11. Schmitt J, Meurer M, Schwanebeck U *et al*. Treatment following an evidence-based algorithm versus individualised symptom-oriented treatment for atopic eczema. A randomised controlled trial. *Dermatology* 2008; **217**: 299-308.
12. 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.
13. Hanifin JM, Thurston M, Omoto M *et al*. The eczema area and severity index (EASI): assessment of reliability in atopic dermatitis. EASI Evaluator Group. *Exp Dermatol* 2001; **10**: 11-8.
14. Fredriksson T, Pettersson U. Severe psoriasis--oral therapy with a new retinoid. *Dermatologica* 1978; **157**: 238-44.
15. Charman CR, Venn AJ, Williams HC. Measurement of body surface area involvement in atopic eczema: an impossible task? *Br J Dermatol* 1999; **140**: 109-11.
16. Barbier N, Paul C, Luger T *et al*. Validation of the Eczema Area and Severity Index for atopic dermatitis in a cohort of 1550 patients from the pimecrolimus cream 1% randomized controlled clinical trials programme. *Br J Dermatol* 2004; **150**: 96-102.
17. Revicki D, Hays RD, Cella D *et al*. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008; **61**: 102-9.
18. Beaton DE, Bombardier C, Katz JN *et al*. A taxonomy for responsiveness. *J Clin Epidemiol* 2001; **54**: 1204-17.
19. Beaton DE, Katz JN, Fossel AH *et al*. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. *J Hand Ther* 2001; **14**: 128-46.
20. Hosmer DW. *Applied logistic regression*. New York: John Wiley and sons, 2000.
21. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; **10**: 407-15.
22. de Winter AF, Heemskerk MA, Terwee CB *et al*. Inter-observer reproducibility of measurements of range of motion in patients with shoulder pain using a digital inclinometer. *BMC Musculoskelet Disord* 2004; **5**: 18.
23. Mokkink LB, Terwee CB, Knol DL *et al*. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010; **10**: 22.
24. Mokkink LB, Terwee CB, Patrick DL *et al*. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010; **63**: 737-45.
25. Mokkink LB, Terwee CB, Knol DL *et al*. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010; **10**: 22.

26. Mokkink LB, Terwee CB, Patrick DL *et al*. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010; **19**: 539-49.
27. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003; **56**: 395-407.
28. Schmitt J, Williams H. Harmonising Outcome Measures for Eczema (HOME). Report from the First International Consensus Meeting (HOME 1), 24 July 2010, Munich, Germany. *Br J Dermatol* 2010; **163**: 1166-8.
29. Schmitt J, Langan S, Stamm T *et al*. Core outcome domains for controlled trials and clinical recordkeeping in eczema: international multiperspective delphi consensus process. *J Invest Dermatol* 2011; **131**: 623-30.
30. Tugwell P, Boers M. OMERACT conference on outcome measures in rheumatoid arthritis clinical trials: introduction. *J Rheumatol* 1993; **20**: 528-30.

ME Schram

E Roekevisch

MMG Leeflang

JD Bos

J Schmitt

PhI Spuls

A randomized trial of methotrexate and azathioprine for severe atopic eczema.
Journal of Allergy and Clnical Immunology 2011, in press

5.1

A RANDOMIZED TRIAL OF METHOTREXATE AND AZATHIOPRINE FOR SEVERE ATOPIC DERMATITIS

SUMMARY

Rationale: Patients with severe atopic dermatitis/eczema (AD) frequently require systemic treatment to control their disease. Methotrexate and azathioprine are proposed as off-label treatment options, but direct comparisons are lacking.

Objective: We sought to compare the efficacy and safety of methotrexate versus azathioprine in adults with severe AD.

Methods: Patients with severe AD were randomly assigned in a 1:1 ratio to receive either methotrexate (dosage, 10-22.5 mg/wk) or azathioprine (dosage, 1.5-2.5 mg/kg/day) for 12 weeks, followed by a 12-week follow-up period. Primary outcome was the mean change in the severity scoring of atopic dermatitis (SCORAD) index after 12 weeks. Efficacy assessors blinded for allocation of treatment were used to perform clinical outcome assessment. Analyses were done on an intention-to-treat basis.

Results: Of the 45 patients screened, 42 were included. At week 12, patients in the methotrexate group had a mean relative reduction in the SCORAD index of 42% (SD 18%) compared with 39% (SD 25%) in the azathioprine group (P= 0.52). Proportions of patients achieving at least mild disease and reductions on impact of quality of life, symptoms, and levels of thymus and activation regulated chemokine (TARC) were similar in both groups at weeks 12 and 24. No statistically significant differences were found in the number and severity of adverse events. Abnormalities in blood count were more common in the azathioprine group. No serious adverse events occurred.

Conclusion: Both treatments achieved clinically relevant improvement and were safe in the short-term. Methotrexate and azathioprine are appropriate options for the treatment of severe AD.

INTRODUCTION

Atopic dermatitis/eczema (AD) is a chronic inflammatory skin disorder that affects approximately 3% to 5% of the adult population in the western world.[1] AD can result in impairment of skin function, poor sleep, and social stigma over a long period of time. Patients often have to bear the burden of considerable psychological comorbidity.[2] Patients with severe AD require prolonged treatment with large amounts of highly potent topical corticosteroids, systemic treatment, or both. Frequently used options for systemic treatment of AD include cyclosporin and systemic corticosteroids. Although proved effective,[3] a proportion of patients have a contraindication for cyclosporin or discontinue treatment because of ineffectiveness or side effects. Moreover, long-term use of cyclosporin raises concerns over (nephro)toxicity.[4] Systemic corticosteroids are used frequently to suppress exacerbations, although high-level evidence is lacking.[5] A recent randomized controlled trial (RCT) comparing short-term cyclosporin versus prednisolone was interrupted because of an unsuspected high proportion of severe rebound in the prednisolone group.[6] Medium- to long-term treatment with prednisolone is relatively contraindicated because of the cumulative effect of the side effects.[5] This illustrates the need for novel medium- to long-term treatment options for patients with severe AD. However, commercial interest for research in eczema is low, and thus investigator-initiated studies are needed.

As health care costs are increasing, dermatologists are looking for cheaper alternatives. Long-existing and relatively cheap disease-modifying antirheumatic drugs seem to be beneficial for AD. Two of those drugs are methotrexate and azathioprine. Azathioprine, a purine synthesis inhibitor that inhibits the proliferation of leukocytes, and methotrexate, a folic acid antagonist that targets several key T-cell activities, are currently used off-label in some (referral) centers. Despite several case series and open-label studies for methotrexate,[7-10] there have been no RCTs supporting a role for methotrexate in the management of AD. The role of azathioprine in AD was established by 2 RCTs in which azathioprine was significantly superior to placebo,with mean improvements of 26% and 37% on clinical outcome scales after three months.[11,12] Numerous uncontrolled studies on azathioprine in adult and juvenile patients showed similar results.[13]

To our knowledge, no comparison of methotrexate with azathioprine in a randomized controlled fashion has been performed. With the present study, we conducted a randomized comparison of methotrexate with azathioprine for the treatment of severe AD evaluating efficacy, safety and effect on quality of life.

METHODS

Design

This study was an investigator-initiated, single-blind, parallel-group (ratio 1:1), RCT evaluating efficacy, safety and quality of life with methotrexate versus azathioprine over a 12-week period. The trial was conducted between July 2009 and December 2010 at the Department of Dermatology of the Academic Medical Center in Amsterdam, The Netherlands. Patients were evaluated every 2 weeks in the first month and monthly thereafter. The follow-up phase consisted of another 12 weeks in which study drugs could be continued, stopped or switched, reflecting normal clinical practice.

The study protocol was reviewed and approved by the local medical ethics committee (institutional review board) and was performed in accordance with the Good Clinical Practice Guidelines of the International Conference of Harmonisation, Declaration of Helsinki. The trial was registered in the Dutch Trial Register (NTR1916). Written informed consent was obtained from all patients before study-related procedures were commenced.

Patients

Patients were recruited from the in- and outpatient clinic of the Academic Medical Center of Amsterdam (referral center for severe AD) or were referred by regional dermatologists. Patients with AD (with and without the presence of allergen-specific IgE) defined according to the Millennium Criteria and the UK Working Party criteria[14] were eligible if they were 18 years or older; the severity grading by the Rajka and Langeland criteria was severe[15]; the patients were unresponsive, contraindicated, or intolerant to cyclosporin treatment; and the patients had not previously been treated with azathioprine or methotrexate.

Excluded were patients who were pregnant, breast-feeding, or planning pregnancy (men and women) until 3 months after discontinuation; those with a history of cancer, alcohol abuse, organ transplantation, chronic or recurrent infectious diseases, or any severe and uncontrolled disease; those with a history of herpes zoster infection within 2 months of baseline or current bacterial skin infection; and those who had received phototherapy, any systemic medication or a potent topical medication within the last 2 weeks.

Because thiopurine methyltransferase (TPMT) is a key enzyme in the purine metabolism and genetic variation in the gene that transcribes TPMT is linked to interpersonal differences in toxicity of azathioprine, patients randomized to the azathioprine group were tested for TPMT activity. When TPMT activity was

low or absent (<21 nmol/g/hour), indicating homozygous TPMT mutations and a subsequent risk for life-threatening myelotoxicity, patients were excluded.

Patients randomized to receive methotrexate were excluded if abnormal laboratory results were discovered after they had taken a test dose of 5 mg of methotrexate.

At every study visit (weeks 0, 2, 4, 8, 12 and 24), laboratory tests were done, including a full blood count and kidney and liver function measurement. Women of childbearing potential underwent a serum pregnancy test at every visit.

Treatment regimens

Treatment with methotrexate was initiated at 10 mg/wk and administered as a single oral dose. Dose escalation with 2.5 to 5 mg per scheduled visit was allowed until 22.5 mg/wk was reached. Because folate supplementation reduces the risk for hepatotoxicity in patients with rheumatoid arthritis, each patient randomized to methotrexate received 5 mg of folate 1 day after methotrexate intake.[16] Azathioprine was initiated at 1.5 mg/kg/day in a single dose, and the dosage could be escalated at each visit with 0.5 mg/kg/day until a maximum of 2.5 mg/kg/day was reached.

Dosage was escalated if patients did not achieve at least a 25% reduction in disease activity at a study visit. The dosage could be decreased according to protocol in case of abnormal findings on physical examination, laboratory markers, and/or adverse events. After the first 12 weeks, dosages in responders were reduced to find the optimum dosage.

Patients were allowed to continue or start with concomitant topical triamcinolone acetonide ointment (body), hydrocortisone ointment (face) and oral antihistamines. In case of an exacerbation or postponed treatment effect in the first 8 weeks of treatment, patients were allowed to receive a maximum of 2 courses of rescue medication: 30 mg/d oral prednisolone for 1 week and a 1 week reduction schedule (20-20-15-15-10-10-5mg).

Outcomes

Efficacy. The primary efficacy outcome parameter was the mean relative and absolute change in the severity of AD at week 12 assessed by means of the SCORAD.[17] The SCORAD score combines both objective items as affected area and intensity of the lesions (erythema, edema/induration, excoriation, oozing/crusting, lichenification, and dryness) and subjective items as extent of pruritus and sleep loss on a visual analog scale (VAS). Scores range from 0 to 108 points.[18]

Secondary outcome parameters included the number of patients with a SCORAD score reduction of 50% or more (SCORAD50) and the number of

patients achieving mild disease (defined as mild, minimal, or no disease activity on investigator global assessment (IGA)), IGA and patient global assessment (PGA), mean change in the eczema area and intensity index (EASI), patient-oriented eczema measurement (POEM), itch and sleeplessness on a VAS, Skindex-17, levels of TARC, amount of concomitant topical corticosteroids and number of courses of rescue medication used.

The IGA and PGA were assessed by using a 6-point Likert scale: 0, clear; 1, almost clear; 2, mild disease; 3, moderate disease; 4, severe disease; and 5, very severe disease. The EASI is based on the extent of the eczematous involvement of the body surface area, as well as the intensity of the lesions (range, 0-72).[19] The POEM includes 7 questions regarding skin symptoms (range, 0-28).[20] Change in quality of life was assessed by the use of the Dutch version of the Skindex-17.[21] Scores range from 0 to 85 points, with higher scores indicating more significantly impaired quality of life. Clinical outcome measures and quality of life were assessed at each visit. Furthermore, at baseline and week 12, serum TARC levels were measured.[22,23]

Safety. The number and severity of adverse events were assessed at each visit by the safety assessor. Adverse events that were transient and easily tolerated by the patient were considered mild. Moderate adverse events were defined as causing discomfort and interrupting the subject's usual activities. Adverse events were severe if the event caused considerable interference with the subject's usual activities and could be incapacitating or life-threatening. Serious adverse events were defined as life-threatening events, death, prolonged or initial hospitalization, disability or permanent damage. The safety assessor defined adverse events as not, possible, probably or definitely related to treatment.

Blinding

Concealment of allocation was achieved by using a computerized program (see the Statistical analysis section). Clinical outcome measurements were assessed by trained efficacy assessors, who were blinded for allocation. Statistical analysis was performed by the third author, who was also blinded for allocation. Patients and safety assessors were not blinded.

Statistical analysis

In the primary analysis the difference in mean SCORAD scores between the treatment groups at week 12 was analyzed by using intention-to-treat analysis. The criterion for including patients in the intention-to-treat analysis was receiving at least 1 dose of study medication.

Randomization was performed in a 1:1 ratio by using a computerized program (TENALEA Clinical Trial Data Management System) with the (nondeterministic) minimization method described by Pocock and Simon.[24] Patient factors (strata) did not influence the allocation scheme. If a patient missed a visit, we used the score from the previous visit for the intention-to-treat analysis.

Without the availability of a formally calculated minimal clinically important difference, we deemed it appropriate to use an 8-point difference in SCORAD scores between groups as the nonequivalence limit.[25,26] Assuming an SD of 10 points in both groups, the power analysis showed that 42 participants were needed for a study with 80% power and 5% significance.

No interim analysis was performed. Two-sided P values of less than 0.05 were considered to indicate statistical significance. Specific statistical tests used are indicated in the legends of the tables. SPSS 18.0 for Windows (SPSS, Inc, Chicago, Ill) was used to perform data analysis.

RESULTS

Patients' characteristics

Between September 2009 and May 2010, 45 patients were screened for enrollment, 43 of whom were randomized (Figure 1). Twenty patients were assigned to the methotrexate group and 23 patients were assigned to the azathioprine group. One patient randomized to the azathioprine group withdrew informed consent before the initiation of study medication and was not included in the analyses. Baseline characteristics of included patients are shown in Table 1. Fifty-two percent of the patients were male, the mean age was 40 years, and the mean duration of eczema was 36 years over the 2 groups. The mean SCORAD score was 58 points at baseline.[18] The mean Skindex-17 score was 51 points.

Mean dosage of methotrexate was 20 mg/wk at week 12 and 17.5 mg/wk at week 24. Mean dosage of azathioprine was 2.2 mg/kg/d at week 12 and 2.1 mg/kg/d at week 24.

During the study, 1 patient in the methotrexate group dropped out after 4 weeks because of nausea and fatigue. Three patients withdrew in the azathioprine group: 1 patient because of lymphocytopenia, increased liver enzyme values, and worsening of preexisting mild anemia at week 4; 1 patient because of nausea and vomiting at week 5; and 1 patient because of failure to adhere to the study protocol at week 2. No patients were lost to follow-up.

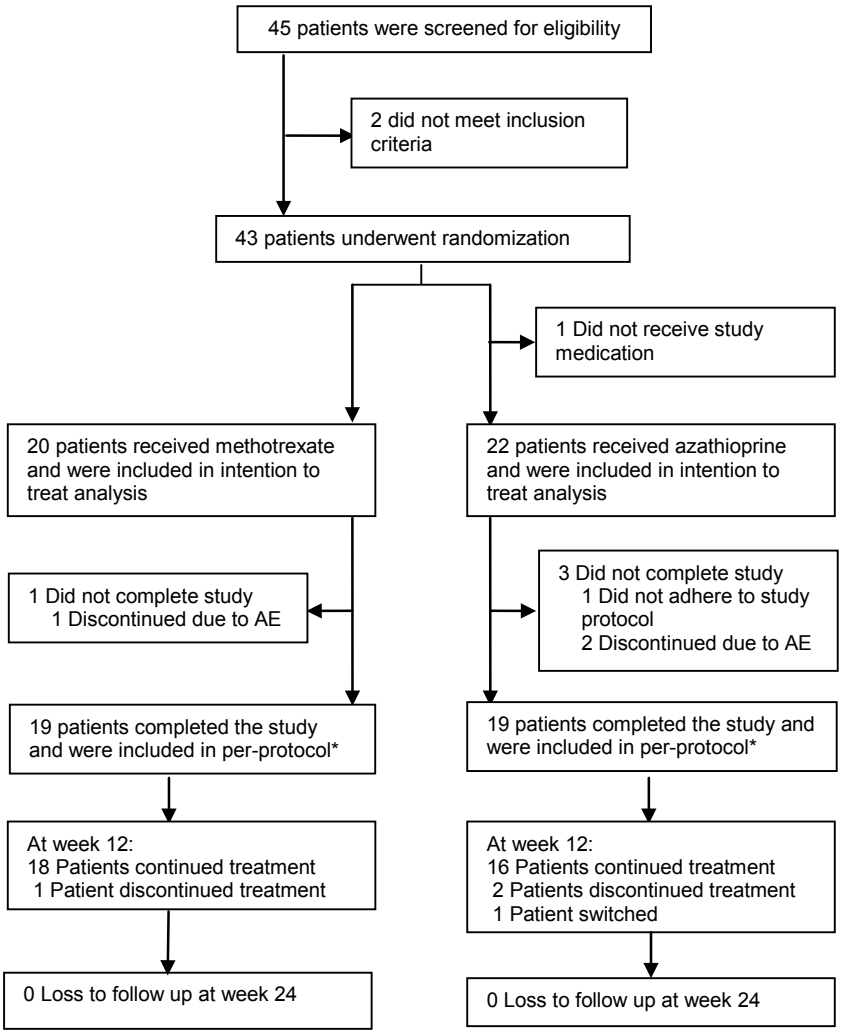


Figure 1. Screening, randomization, treatment and follow-up of study participants. *Data per-protocol analysis not shown. AE; adverse event

Table 1. Demographics and baseline characteristics of the included participants

	Treatment group	
	Methotrexate (n=20)	**Azathioprine (n=22)**
Age (y), mean (SD)	43.0 (14.7)	37.0 (14.1)
Male sex	10 (50%)	12 (55%)
Presence of asthma or allergic rhinitis	19 (95%)	18 (82%)
Presence of allergen-specific IgE	20 (100%)	21 (95%)
Duration of atopic eczema (y), mean (SD)	39.8 (16.2)	33.1 (16.8)
Disease activity		
SCORAD, mean (SD)	57.2 (11.8)	58.4 (10.4)
IGA, mean (SD)	3.8 (0.6)	3.6 (0.6)
PGA, mean (SD)	3.9 (0.6)	3.8 (0.8)
EASI, mean (SD)	27.9 (12.3)	30.4 (14.2)
POEM, mean (SD)	19.8 (5.3)	19.5 (4.0)
Quality of life (Skindex-17), mean (SD)	50.2 (11.7)	51.7 (8.6)

Efficacy

At week 12, mean SCORAD scores in the patients in the methotrexate group changed from 57.2 (SD 11.8) to 34.4 (SD13.0), representing a relative reduction of 42% (P<0.001, Table 2). SCORAD scores in patients randomized to the azathioprine group changed from 58.4 (SD 10.4) to 36.3 (SD 16.9), representing a relative reduction of 39% (P < 0.001). The P value for the absolute difference between the groups is 0.89. Figure 2 shows the mean SCORAD scores during the study period.

Eight (40%) patients in the methotrexate group versus 10 (45%) patients in the azathioprine group achieved a SCORAD50 response (P= 0.76). Fifteen patients in each group (75% in the methotrexate group vs 68% in the azathioprine group) achieved at least mild disease on IGA (P= 0.74). On global assessment, the mean IGA score was reduced to 1.8 (SD 0.7) in the methotrexate group versus 1.4 (SD 0.9) in the azathioprine group (P= 0.20) and the mean PGA score was reduced to 1.3 (SD 0.9) in the methotrexate group versus 1.2 (SD 1.3) in the azathioprine group (P= 0.95). The EASI score was reduced to 17.4 (SD 6.6) points in the methotrexate group compared with 17.2 (SD 14.1) points in the azathioprine group (P= 0.95). Reduction in the mean POEM score was 6.9 (SD 5.7) in the methotrexate group versus 7.9 (SD 7.7) in the azathioprine group (P= 0.65). Clinical improvement was paralleled by a decrease in symptoms. Mean VAS itch scores decreased to 2.5 (SD 2.2) in the methotrexate group versus 2.6

Table 2. Clinical response at week 12 (intention-to-treat analysis)

	Treatment group		
Variable	**Methotrexate (n=20)**	**Azathioprine (n=22)**	**P Value**
Improvement in SCORAD			
Absolute reduction, mean (SD)	22.7 (7.9)	22.2 ± 16.5	0.89*
Relative reduction, mean (SD)	42% (18%)	39% (25%)	0.70*
At least 50% (SCORAD50), no. (%)	8 (40%)	10 (45%)	0.76¥
Improvement IGA			
Reduction, mean (SD)	1.8 (0.7)	1.4 (0.9)	0.20*
Cleared, minimal or mild disease (IGA <2), no.(%)	15 (75%)	15 (68.2%)	0.74¥
Improvement other outcomes			
Reduction in PGA, mean (SD)	1.3 (0.9)	1.2 (1.3)	0.95*
Reduction in EASI, mean (SD)	17.4 (6.6)	17.2 (14.1)	0.95*
Reduction in POEM, mean (SD)	6.9 (5.7)	7.9 (7.7)	0.65*
Reduction in VAS itch score, mean (SD)	2.5 (2.2)	2.6 (2.2)	0.78*
Reduction in VAS sleeplessness score, mean (SD)	2.8 (2.6)	3.8 (2.8)	0.24*
Reduction in Skindex-17, mean (SD)	12.9 (8.8)	10.3 (12.9)	0.46*
Reduction in TARC levels, median (IQR)	1215 (302 to 2496)	885 (122 to 3107)	0.61**
Use of concomitant topical steroids(g), median (IQR)	115.2 (45 to173)	79.1 (22 to 121)	0.16§
No. of rescue medication, no. (%)	2 (10%)	4 (18%)	0.67¥

IQR; interquartile range
* T-test for independent groups;
§ Mann-Whitney U test for independent groups;
¥ Fisher's exact test
** based on ln_TARC.

(SD 2.2) in the azathioprine group (P= 0.78). Mean VAS scores for sleeplessness decreased to 2.8 (SD 2.6) in the methotrexate group versus 3.8 (SD 2.8) in the azathioprine group (P= 0.24).

When comparing quality of life, the mean Skindex-17 score was reduced from 50.2 (SD 11.7) at baseline to 37.8 (SD 9.8) at week 12 in the methotrexate group (P < 0.001). In the azathioprine group the mean Skindex-17 score was reduced from 51.7 (SD 8.6) to 41.5 (SD 13.1; P < 0.001). This equals a reduction of 26% versus 20% (P= 0.65).

Median TARC levels decreased to 1215 ng/mL (interquartile range, 302-2496 ng/mL) in the methotrexate group and 885 ng/mL (interquartile range, 122-3107 ng/mL) in the azathioprine group (P= 0.61).

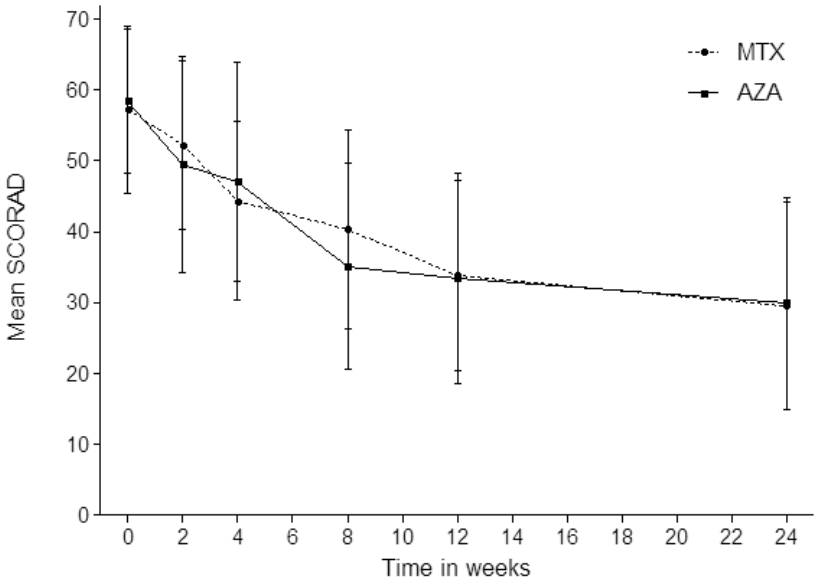


Figure 2. Mean (SD) scores for the SCORAD at baseline, during treatment and follow-up. MTX; methotrexate, AZA; azathiorpine.

Patients in the methotrexate group used a median of 115.2 g of concomitant topical corticosteroids versus 79.1 g in the azathioprine group. Two (10%) patients in the methotrexate group required a course of rescue medication at weeks 1 and 2, respectively, and 4 (18%) patients in the azathioprine group required a course at weeks 1, 2, 4 and 5 respectively (P= 0.67).

Safety

Table 3 shows an overview of the safety results. Abnormalities in blood count (mostly lymphocytopenia) were statistically significantly more frequent in the azathioprine group (P= 0.002). Infections, gastrointestinal adverse events, and increased liver enzyme levels occurred in equal proportion in both groups. Fourteen patients in each group experienced infections, mainly upper airway infections, common colds, and mild skin infections. Skin infections occurred in 5 (25%) patients in the methotrexate group and 7 (32%) patients in the azathioprine group; all were mild. Five infections in the methotrexate group and

Table 3. Adverse events and other key safety data through week 24 Treatment group

	Treatment group		
Variable	**Methotrexate**	**Azathioprine**	**P Value***
Total no. of patients	20	22	
Total no. of patients with adverse events, no. (%)	20 (100%)	22 (100%)	1.00
Serious and severe adverse events[¥]	0	0	-
Adverse events, total events**	113	121	
Adverse events leading to withdrawal	1 (5%)	2 (9%)	1.00
Adverse event requiring dose adjustment, no. of patients (%)	2 (10%)***	2 (9%)	1.00
Treatment-related adverse events[§], no. of patients (%)	14 (70%)	12 (64%)	0.66
Categorized adverse events, no. of patients (%)			
Infections	14 (70%)	14 (64%)	0.19
Gastro-intestinal adverse events	11 (55%)	13 (59%)	0.79
Exacerbation of atopic eczema	3 (15%)	2 (9%)	0.66
Abnormalities in blood count	6 (30%)	17 (77%)	0.002
Increased liver enzymes	7 (35%)	8 (36%)	0.93

* P Values were calculated using Chi-square test or Fisher's exact test.
[¥] Adverse events were classified as serious and severe according to pre-defined definition (see methods section).
** Total number of adverse events: all patients displayed adverse events, the number of events per patient ranging from 1 to 14 (median 5, interquartile range 3.25 to 8). If a patient presented twice with the same adverse effect (e.g. twice migraine), then this was counted as 1 adverse event. This was the case in 17 out of the 42 patients; 2 patients presented a certain symptom 3 times (which was also counted as 1 time).
***One patient required two dose lowerings
[§] Treatment-related adverse events are those classified as possibly, probably or definitely related to the study drug by the safety assessor.

8 in the azathioprine group were considered moderate of intensity. Three (15%) patients in the methotrexate group had an exacerbation of their eczema during the study compared with 2 (9%) patients in the azathioprine group. No severe and serious adverse events occurred.

Follow-up

After 12 weeks, 18 (95%) patients in the methotrexate group continued their treatment, and 1 patient discontinued after induction of remission. In the azathioprine group 16 (84%) patients continued, 1 switched to methotrexate because of lack of efficacy, and 2 discontinued because of induction of remission. At week 24, the mean SCORAD score on intention-to-treat analysis was 30.4 (SD 14.3) in the methotrexate group and 33.7 (SD 16.9) in the azathioprine group

(P= 0.58), representing a relative reduction from baseline of 48% versus 43%. After 24 weeks of treatment, there was no statistically significant difference between the 2 groups in all outcome measures.

DISCUSSION

General conclusion

Disease-modifying antirheumatic drugs, such as methotrexate, azathioprine and mycophenolate, offer a range of off-label therapeutic options for patients with severe AD. In the present RCT methotrexate and azathioprine were compared. Patients in the methotrexate group experienced a statistically significant overall improvement of 42% in mean SCORAD score at week 12. Patients in the azathioprine group showed a statistically significant overall improvement of 39% in mean SCORAD score.

In addition, intensity of symptoms and TARC levels were reduced in a similar fashion in both groups. Effect on quality of life reduced by 26% in the methotrexate group versus 20% in the azathioprine group. Overall, there was no statistical difference between both therapies in any primary or secondary efficacy outcome measures assessed at weeks 12 and 24.

In this study the number and severity of adverse events, including laboratory abnormalities, appeared to be generally similar in short-term treatment, with the exception of mild myelosuppression in the azathioprine group. No severe or serious adverse events occurred.

Because of the relatively slow onset of action of both treatments, it is common in some clinical practices to give patients concomitant oral corticosteroids to support them in the first weeks of treatment. In our study both treatments were given as single systemic therapy with the possibility of adjuvant rescue medication. This was needed in 6 of the 42 patients treated. This might indicate that routine administration of concomitant oral corticosteroids is not necessary.

Our results are in concordance with the previous results of 2 placebo-controlled studies regarding azathioprine, in which improvements of 26% and 37% on the 6-area, 6-sign atopic dermatitis score at 12 weeks were found.[11,12] Similar results were found for the effect of methotrexate in cohort studies, although in some the effect was more pronounced: 52% at week 24 in one study and a greater than 70% reduction in outcome parameters in 65% of the patients at week 12 in another.[7,9] Studies on cyclosporin, which is the first-choice systemic treatment, showed a more marked response. A systematic

review showed that relative improvements were consistently greater than 50% at 6 to 8 weeks.[3]

In conclusion, methotrexate and azathioprine can be considered equally effective for the treatment of severe AD in adults. Overall, this study is limited to conclude about the safety for medium- to long-term use. Nevertheless, because both drugs have been available for more than 50 years, they have a well-known toxicity profile, and dermatologists are familiar with the use of these drugs in the treatment of psoriasis or bullous diseases. Patients treated with azathioprine should be monitored for myelosuppression, and preferably, TPMT levels should be measured before the initiation. Patients receiving methotrexate should be monitored for hepatic and pulmonary toxicity and myelosuppression. Individual treatment decisions should be based on patients' preferences and comorbidity.

Strengths and limitations

Important strengths of our trial include that our study is the first head-to-head comparison of methotrexate versus azathioprine in patients with severe AD. Thereby, the study was investigator-initiated, with the investigators having no conflict of interest; the study included patients naive for methotrexate or azathioprine; and validated outcome measures were used.

Our trial had certain limitations. Power analysis was based on the SCORAD score. Because the minimal clinically important difference of the SCORAD score is unknown, we decided that a difference of at least 8 points would represent a clinical meaningful difference. This was based on other studies and on personal experience with the scale. Nevertheless, it could have been possible that a smaller SCORAD score difference was clinically meaningful. In that case our sample size would not have been sufficient to detect a meaningful difference between the therapies.

In our study the washout period for systemic treatment was relatively short. This might have resulted in a lower baseline SCORAD score and in a smaller difference between baseline and week 12 efficacy outcomes. At baseline, patients were allowed to start or increase the amount of topical steroids. This could have influenced the treatment effect. However, considering the low total amount and low potency of the topical steroids used during the trial, the effect on efficacy outcomes will be minimal. Continuing with a stable dose or implementation of a stabilization period would have been more methodologically sound.

Ideally, patients should be blinded for allocation of treatment to avoid a performance bias. Because each patient received active treatment in this study, this effect will be less pronounced than in a placebo-controlled design. Concerning the non-blinded safety assessors, patient management could have

been biased. However, it should be noted that safety assessors did not have personal preferences or conflicts of interest. A patient-blinded design could not be performed because of a lack of funding and other financial resources.

In the absence of international guidelines for these treatments, our dosing schedule was based on earlier performed studies and experience with the drugs in other indications. Therefore it could be argued that the interventions were not similar.

Currently, there is an international initiative to harmonize outcome assessment in AD: the Harmonizing Outcome Measurement in Eczema initiative. During a Delphi round on core outcome domains, it was shown that symptoms, physician assessed clinical signs, and measurement for long-term control were of particular interest.[27] Symptoms were addressed by means of VASs on itch and sleeplessness and by means of the POEM. Physician-assessed clinical signs were addressed by using the SCORAD and EASI. Both were sufficiently validated according to a recently published systematic review on the validity of clinical outcome measures.[28] Although we have performed a 3-month follow-up period to indicate long-term control, we could not totally comply with the measurement for long-term control because of the design of the study. An observational follow-up study, which is currently being conducted, is directed at that particular outcome domain.

Generalisability

This study was performed in adults (18-75 years of age) with severe AD who were unresponsive or contraindicated for cyclosporin treatment. Sex was equally distributed. Because the included patients were 40 years old and had a mean duration of disease of 36 years, we can conclude it concerned a very chronic population. In some cases included patients were very severely affected and had frequent hospitalizations for disease management in the past. All except 1 of the patients showed the presence of allergen-specific IgE on a Phadiatop test (Pharmacia, Uppsala, Sweden). We believe that our data are applicable to all adults with difficult-to-treat chronic AD, depending on their comorbidity and preferences. Studies in children should be performed to establish the role of these treatments in children.

Clinical research implications

We have demonstrated that both methotrexate and azathioprine are effective and short-term safe treatment options for adults with severe AD. We believe that the results from this study justify treatment with these drugs when regular treatment is insufficient. Furthermore, the results from this study can be used to

update or formulate clinical practice guidelines for the treatment of severe AD. In light of the rules and regulations concerning off-label treatment, it is important to generate evidence-based guidelines for its use. Future studies should be performed to confirm the long-term safety profile of both methotrexate and azathioprine, to confirm their role in children, and to compare both treatments with other therapies, such as cyclosporin and oral corticosteroids.

AKNOWLEDGEMENTS

We thank all the patients who participated in this study, B. Arents of the patient association of atopic dermatitis (VMCE), and all our fellow dermatologists for referring patients (in particular F. Wu, M. Nahuys, M. D. Njoo and B. Nanninga). We also thank G. A. Appel and A. C. Q. de Vries for being the blinded efficacy assessors, M. Bonnerjee for monitoring the study, E. M. Kemper for being the involved pharmacologist, A. B. P. van Kuilenberg for assessing the TPMT levels, and K. Szegedi and P. Res for assessing TARC levels.

REFERENCES

1. Williams H, Robertson C, Stewart A, it-Khaled N, Anabwani G, Anderson R, et al. Worldwide variations in the prevalence of symptoms of atopic eczema in the International Study of Asthma and Allergies in Childhood. J Allergy Clin Immunol 1999; **103**: 125-38.
2. Schmitt J, Romanos M, Pfennig A, Leopold K, Meurer M. Psychiatric comorbidity in adult eczema. Br J Dermatol 2009; **161**:878-83.
3. Schmitt J, Schmitt N, Meurer M. Cyclosporin in the treatment of patients with atopic eczema-a systematic review and meta-analysis. J Eur Acad Dermatol Venereol 2007; **21**:606-19.
4. Behnam SM, Behnam SE, Koo JY. Review of cyclosporin immunosuppressive safety data in dermatology patients after two decades of use. J Drugs Dermatol 2005; **4**:189-94.
5. Schmitt J, Schakel K, Schmitt N, Meurer M. Systemic treatment of severe atopic eczema: a systematic review. Acta Derm Venereol 2007; **87**:100-11.
6. Schmitt J, Schakel K, Folster-Holst R, Bauer A, Oertel R, Augustin M, et al. Prednisolone vs. ciclosporin for severe adult eczema. An investigator-initiated double-blind placebo-controlled multicentre trial. Br J Dermatol 2010; **162**:661-8.
7. Weatherhead SC, Wahie S, Reynolds NJ, Meggitt SJ. An open-label, dose-ranging study of methotrexate for moderate-to-severe adult atopic eczema. Br J Dermatol 2007; **156**: 346-51.
8. Zoller L, Ramon M, Bergman R. Low dose methotrexate therapy is effective in late-onset atopic dermatitis and

idiopathic eczema. Isr Med Assoc J 2008; **10**: 413-4.

9. Goujon C, Berard F, Dahel K, Guillot I, Hennino A, Nosbaum A, et al. Methotrexate for the treatment of adult atopic dermatitis. Eur J Dermatol 2006;**16**: 155-8.
10. Shaffrali FC, Colver GB, Messenger AG, Gawkrodger DJ. Experience with lowdose methotrexate for the treatment of eczema in the elderly. J Am Acad Dermatol 2003; **48**: 417-9.
11. Berth-Jones J, Takwale A, Tan E, Barclay G, Agarwal S, Ahmed I, et al. Azathioprine in severe adult atopic dermatitis: a double-blind, placebo-controlled, crossover trial. Br J Dermatol 2002; **147**: 324-30.
12. Meggitt SJ, Gray JC, Reynolds NJ. Azathioprine dosed by thiopurine methyltransferase activity for moderate-to-severe atopic eczema: a double-blind, randomized controlled trial. Lancet 2006; **367**: 839-46.
13. Schram ME, Borgonjen RJ, Bik CJMJ, van der Schroeff JG, van Everdingen JJE, Spuls PhI. Off-label use of azathioprine in dermatology: a systematic review. Arch Dermatol 2011; **147**: 474-88.
14. Brenninkmeijer EE, Schram ME, Leeflang MM, Bos JD, Spuls PhI. Diagnostic criteria for atopic dermatitis: a systematic review. Br J Dermatol 2008; **158**: 754-65.
15. Rajka G, Langeland T. Grading of the severity of atopic dermatitis. Acta Derm Venereol Suppl (Stockh) 1989; **144**: 13-4.
16. Hoekstra M, van Ede AE, Haagsma CJ, van de Laar MA, Huizinga TW, Kruijsen MW, et al. Factors associated with toxicity, final dose, and efficacy of methotrexate in patients with rheumatoid arthritis. Ann Rheum Dis 2003; **62**: 423-6.
17. Severity scoring of atopic dermatitis: the SCORAD index. Consensus Report of the European Task Force on Atopic Dermatitis. Dermatology 1993; **186**: 23-31.
18. Oranje AP, Glazenburg EJ, Wolkerstorfer A, de Waard-van der Spek FB. Practical issues on interpretation of scoring atopic dermatitis: the SCORAD index, objective SCORAD and the three-item severity score. Br J Dermatol 2007; **157**: 645-8.
19. Tofte SJ, Graeber M, Cherill R, Omoto M, Thurston M, Hanifin JM. Eczema area and severity index (EASI): A new tool to evaluate atopic dermatitis. J Eur Acad Dermatol Venereol 1998; **11**(suppl 2):S197.
20. Charman CR, Venn AJ, Williams HC. The patient-oriented eczema measure: development and initial validation of a new tool for measuring atopic eczema severity from the patients' perspective. Arch Dermatol 2004; **140**: 1513-9.
21. Nijsten TE, Sampogna F, Chren MM, Abeni DD. Testing and reducing skindex-29 using Rasch analysis: Skindex-17. J Invest Dermatol 2006; **126**: 1244-50.
22. Hijnen D, De Bruin-Weller M, Oosting B, Lebre C, De Jong E, Bruijnzeel-Koomen C, et al. Serum thymus and activation-regulated chemokine (TARC) and cutaneous T cell-attracting chemokine (CTACK) levels in allergic diseases: TARC and CTACK are disease-specific markers for atopic dermatitis. J Allergy Clin Immunol 2004; **113**: 334-40.
23. Kakinuma T, Nakamura K, Wakugawa M, Mitsui H, Tada Y, Saeki H, et al. Thymus and activation-regulated chemokine in atopic dermatitis: Serum thymus and activation-regulated chemokine level is closely related with disease activity. J Allergy Clin Immunol 2001; **107**: 535-41.
24. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled

clinical trial. Biometrics 1975; **31**: 103-15.

25. Paul C, Lahfa M, Bachelez H, Chevret S, Dubertret L. A randomized controlled evaluator-blinded trial of intravenous immunoglobulin in adults with severe atopic dermatitis. Br J Dermatol 2002; **147**: 518-22.
26. Schmitt J, Meurer M, Schwanebeck U, Grahlert X, Schakel K. Treatment following an evidence-based algorithm versus individualised symptom-oriented treatment for atopic eczema. A randomised controlled trial. Dermatology 2008; **217**: 299-308.
27. Schmitt J, Williams H. Harmonising Outcome Measures for Eczema (HOME). Report from the First International Consensus Meeting (HOME 1), 24 July 2010, Munich, Germany. Br J Dermatol 2010; **163**: 1166-8.
28. Schmitt J, Langan S, Williams HC. What are the best outcome measurements for atopic eczema? A systematic review. J Allergy Clin Immunol 2007; **120**: 1389-98.

ME Schram

PhI Spuls

JD Bos

Off-label use of efalizumab in dermatology.
Expert Review of Dermatology 2010;5:535-47

5.2 OFF-LABEL USE OF EFALIZUMAB IN DERMATOLOGY : A SYSTEMATIC REVIEW

SUMMARY

In reaction to the marketing suspension recommended by the European Medicines Agency, Genentech announced the voluntary withdrawal of efalizumab. Since June 8, 2009, efalizumab has been unavailable. Despite the fact that the use of efalizumab is now ceased, results of studies with efalizumab will remain relevant, especially about its use in an off-label setting for indications other then chronic plaque psoriasis. New biologics as well as chemicals are being developed targeting similar T-cell-mediated pathways. Indications for these new compounds other than psoriasis could be selected based on this systematic review. With this article we aimed to summarize the evidence of effectiveness, efficacy and safety of efalizumab in off-label dermatological use and to illustrate the time scale in which drugs are offered off-label to dermatological patients. We found some evidence supporting the effectiveness of efalizumab in palmoplantar pustulosis, atopic dermatitis, lichen planus and cutaneous lupus erythematosus. In total, 12 serious adverse events occurred, one due to an infection. Efalizumab was given off-label within half a year after registration of the drug for plaque psoriasis.

INTRODUCTION

In reaction to the marketing suspension recommended by the European Medicines Agency (EMA), Genentech announced the voluntary withdrawal of efalizumab. Since June 8, 2009, efalizumab is no longer available. The reason for this withdrawal is that the benefit of the drug in psoriasis is no longer considered to outweigh its risks, after four cases of progressive multifocal leukoencephalopathy (PML) in a patient population of approximately 48,000 patient-years. Although this is not an unknown side effect in different immunosuppressive treatments, the additive value of efalizumab for psoriasis on the comprehensive biological market at this point is questioned.[1]

Efalizumab is a recombinant, humanized, monoclonal IgG-1 antibody that binds to CD11a, the α-subunit of leukocyte function-associated antigen-1. As such, it interferes with T-cell trafficking and T-cell activation.[2] A series of randomized controlled trials (RCTs) have demonstrated the efficacy and safety of efalizumab in the treatment of chronic moderate to severe plaque psoriasis.[3-8] Thus far, the longest evaluation of the efficacy and safety of systemic biological therapy in psoriasis has been a 3-year, phase III, open-label trial of continuous efalizumab therapy.[9-11]

Short-term adverse events such as flu-like symptoms may be seen following the first two doses, with an incidence comparable to placebo after the third dose.[12] There was no evidence of end-organ toxicity.[11,12]

Efalizumab was registered for the treatment of plaque psoriasis, but was also extensively explored in an off-label setting for other, often difficult to treat, dermatological diseases. Efalizumab was suspended due to an unfavourable risk–benefit ratio in psoriasis, therefore it will probably be unavailable for other dermatological diseases also.

However, in this article, we summarize the evidence of safety, efficacy and effectiveness of treatment with efalizumab in patients with dermatological diseases other than psoriasis and show the time scale in which efalizumab was used off-label in dermatology.

METHODS

Inclusion and exclusion criteria

RCTs, case reports and pilot studies in which patients with dermatological diseases other than plaque psoriasis were treated with efalizumab were assessed for eligibility.

Studies that reported efficacy, effectiveness and/or safety were included. Reviews, unpublished articles or poster presentations were excluded. No restrictions were imposed regarding age, gender, skin type and number of subjects in a study. No language restrictions were applied. Double publications were excluded.

Literature search

Between December 2008 and January 2009, a literature search in MEDLINE, EMBASE and CENTRAL was performed (Table 1). As the main search strategy, 'efalizumab' and synonyms were used. An additional search was carried out combining the search term 'monoclonal antibodies' with the different dermatological diseases found in the first search supplemented with Sjögren's disease and pyoderma gangrenosum; diseases that may be T-cell-mediated, but were not found in the initial search. There was no limit to the date of the publication. References of all relevant articles found were checked in order to find additional articles.

Study selection & data extraction

All articles with a title and abstract considering efalizumab treatment in dermatological patients other than psoriasis were selected by the first author

Table 1. Search strategy for MEDLINE and EMBASE

MEDLINE
1. ("efalizumab"[Substance Name] OR "efalizumab"[All Fields]) OR ("efalizumab"[Substance Name] OR "efalizumab"[All Fields] OR "raptiva"[All Fields]) OR hu1124[All Fields] OR CD11a[All Fields] OR xanelim[All Fields]
2. "antibodies, monoclonal"[MeSH Terms] OR "monoclonal antibodies"[All Fields]
3. "dermatitis, atopic"[MeSH Terms] OR ("dermatitis"[All Fields] AND "atopic"[All Fields]) OR "atopic dermatitis"[All Fields] OR ("atopic"[All Fields] AND "dermatitis"[All Fields])
4. "palmoplantar pustulosis"[All Fields] OR "palmoplantar pustular psoriasis"[All Fields] OR "foot psoriasis"[All Fields]
5. "Pityriasis Rubra Pilaris"[Mesh] OR "Pityriasis Rubra Pilaris"[All Fields]
6. "Lichen Planus"[Mesh] AND "Lichen Planus"[All Fields]
7. "Alopecia Areata"[Mesh] AND "Alopecia Areata"[All Fields]
8. "Stomatitis, Aphthous"[Mesh] AND "Stomatitis, Aphthous"[All Fields]
9. "Pyoderma Gangrenosum"[Mesh] AND "Pyoderma Gangrenosum"[All Fields]
10. "Granuloma Annulare"[Mesh] AND "Granuloma Annulare"[All Fields]
11. "Dermatomyositis"[Mesh] AND "Dermatomyositis"[All Fields]
12. "Vitiligo"[Mesh] AND "Vitiligo"[All Fiels]
13. "Sjogren's Syndrome"[Mesh] AND "Sjogren's Syndrome"[All fields]
14. "Lupus Erythematosus, Discoid"[Mesh] OR "lupus erythematosus"[All Fields]
15. 2 AND (3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14)
16. 1 AND 15

Continued on next page

Table 1. *Continued*

EMBASE
1. Efalizumab/
2. hu1124.mp. or monoclonal antibody CD11a/ or xanelim.mp.
3. 1 or 2
4. exp Monoclonal Antibody/
5. Atopic Dermatitis/
6. ((atopic$ or intrins$ or allergic$) adj3 (dermatit$ or eczem$)).tw.
7. 5 or 6
8. exp Pustular Psoriasis/ or exp Pustulosis Palmoplantaris/or exp Hand and foot psoriasis/
9. Pityriasis Rubra Pilaris.mp. or exp Pityriasis Rubra Pilaris/
10. hidradenitis suppurativa.mp. or exp Suppurative Hidradenitis/
11. lichen planus.mp. or exp Lichen Planus/
12. aphthous stomatitis.mp. or exp Aphthous Stomatitis/
13. pyoderma gangrenosum.mp. or exp Pyoderma Gangrenosum/
14. granuloma annulare.mp. or exp Granuloma Annulare/
15. vitiligo.mp. or exp vitiligo/
16. cutaneous lupus erythematosus.mp. or exp Skin Lupus Erythematosus/
17. sjogren.mp. or exp Sjoegren Syndrome/
18. 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17
19. 4 AND 18
20. 3 AND 19

(M.S.) for relevance. To determine eligibility, the full text of the selected articles was screened by two reviewers (M.S. & drs. J. La Verge).

Data on study characteristics, efficacy, effectveness and safety were extracted by two reviewers independently (M.S. & drs. J. La Verge) using a data extraction form. Study characteristics included: dermatological disease concerned, design of the study, number of patients included, dose of efalizumab and duration of study. Disagreements about study selection and data extraction were solved by discussion.

RCTs were assessed following the criterion grading system described in the *Cochrane Handbook For Systematic Reviews Of Interventions 5.0.1* (updated February 2008).[13] To help assess the risk of bias within included RCTs, the following parameters of methodological quality were assessed; sequence generation, allocation of concealment, blinding (of participants, researchers and outcome assessment), handling of withdrawals and losses and other potential threats to validity.

Each item is scored with 'adequate', 'inadequate' or 'unclear' in order to assess the risk of bias. Each reviewer graded the selected studies for each disease according to the grading of recommendations assessment, development and evaluation system (GRADE). This system uses the following definitions in grading the quality of evidence: high – further research is very unlikely to change our

confidence in the estimate of effect; moderate – further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate; low – further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate; very low – any estimate of effect is very uncertain.[14]

RESULTS

Search results

An initial search retrieved 2429 articles. After screening the title and abstract for eligibility, 46 articles were selected. Then, after screening the full texts of the articles, 28 articles were identified to be relevant. Figure 1 summarizes the selection process. Reasons for exclusion were lack of relevance in six articles[15-20], review design without additional new evidence in 11 articles[21-31] and one double publication.[32] One RCT, seven pilot studies and 20 case reports were identified concerning 11 dermatological diseases other than psoriasis.

Study characteristics of included studies are shown in Table 1, whereas study results are summarized in Table 2. The presentation of the results in Table 2 resembles the original wording as used in the articles. Studies reporting efficacy of efalizumab in patients with two different dermatological diseases were discussed separately for each disease. Although palmoplantar pustulosis (PPP) and hand and foot psoriasis (HFP) are seen as distinct entities, these diseases are often seen as one in many studies. Therefore, we have chosen to combine these data.

Palmoplantar pustulosis and hand & foot psoriasis

Six studies were found that deal with efalizumab treatment in PPP and/or HFP; three pilot studies and three case reports.[33-38] In a pilot study concerning ten patients with HFP, six patients showed a mean physician global assessment improvement of 1.2 points after 3 months of treatment, while the other four patients had inadequate responses and dropped out.[37] However, in a pilot study with 17 patients with either HFP or PPP, all patients showed a clinical benefit from treatment.[38] A total of 12 of them had more than 90% improvement of their lesions. Overall, efalizumab was effective in the treatment of PPP and HFP. In two of the six articles, lesions on the hands improved quicker and better than lesions on the feet.[33,38] Clinically significant side effects reported in these studies included one with hemolytic anemia, one case of thrombocytopenia, one case of ulcerative colitis and three severe headaches.

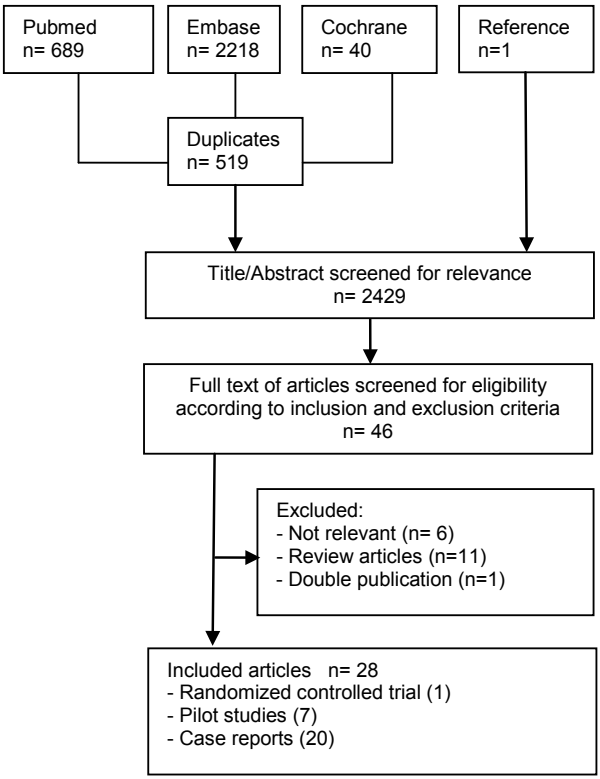


Figure 1. Summary of selection process for studies concerning efalizumab treatment in dermatological diseases other than psoriasis.

According to the GRADE system, the quality of this evidence is considered very low, indicated by the fact that no RCTs were conducted and publication bias is likely.

Pityriasis rubra pilaris

There have been two reports of the use of efalizumab in pityriasis rubra pilaris (PRP), a condition in which the pathogenesis is not known.[39,40] The first involved a 10-year-old child with unresponsive PRP and comorbidity with psoriasis, who achieved 50% improvement of his PRP after the first dose and remission after 9 months of treatment with efalizumab.[39] The second report involved a 60 year

Table 2. Characteristics of included articles

Author (year)	Study design	Treatment (months)	FU (months)	Disease of subjects	Number of subjects	Age of subjects (years)	Dose (mg/kg/week)
Palmoplantar pustulosis							
Fretzin *et al.* (2006)	Pilot	2- 18	1-9	PPP & HFP	17	29 - 83	1
Cohen *et al.* (2007)	Pilot	5		HFP	7	36 - 74	1
Colsman *et al.* (2008)	CR	4 & 6	4	PPP	2	45 & 55	1
Stinco *et al.* (2008)	CR	7		PPP	1	74	1
Varma *et al.* (2008)	Pilot	3	3	HFP	10	52.2 (mean)	1
Wozel *et al.* (2008)	CR	17		PPP	1	55	*
Pityriasis Rubra Pilaris							
Klein *et al.* (2007)	CR	2.2		PRP	1	60	1
Gomez *et al.* (2007)	CR	9		PRP & PS	1	10	1
Hidradenitis suppurrativa							
Strober *et al.* (2007)	Pilot	3	1	HS	5	18 - 64	1
Atopic Dermatitis							
Farshidi *et al.* (2006)	CR	> 8		AD	1	30	1
Kaelin *et al.* (2006)	CR	7		AD & AA	1	19	1
Weinberg *et al.* (2006)	CR	8 &19		AD & AA	2	8 & 48	1
Hassan *et al.* (2007)	CR	11		AD	1	19	1
Siegfried *et al.* (2007)	CR	24	6	AD & AA	1	<18	*
Takiguchi *et al.* (2007)	Pilot	3	2	AD	10	≥ 18	1
Lichen Planus							
Cheng *et al.* (2006)	CR	2.5		OLP	1	54	1
Böhm *et al.* (2007)	CR	3		LP	1	29	1
Heffernan *et al.* (2007)	Pilot	3	2	OLP	4	52 - 71	1

Alopecia areata							
Kaelin *et al.* (2006)	CR	7		AA & AD	1	19	1
Weinberg *et al.* (2006)	CR	19		AA & AD	1	8	1
Siegfried *et al.* (2007)	CR	24	6	AA & AD	1	<18	*
Kolde *et al.* (2008)	CR	8	6	AA	1	44	*
Price *et al.* (2008)	RCT/OLE	3 RCT 3 OLE	3	AA	62	18 – 59 (mean 35.7)	1
Aphthous stomatitis							
Zribi (2007)	CR	3	2	AS & PS	1	50	1
Granuloma annulare							
Goffe *et al.* (2004)	CR	9		GA & PS	1	52	*
Dermatomyositis							
Huber *et al.* (2006)	CR	12		DM	1	82	1 to 1.8
Vitiligo							
Fernandez *et al.* (2008)	CR	13		VI & PS	1	43	1
Wakkee *et al.* (2008)	CR	3	2	VI & PS	1	52	1
Lupus Erythematosus							
Clayton *et al.* (2006)	CR	5		CLE	1	47	1 to 1.25
Hamprecht *et al.* (2007)	CR	1.5		CLE	1	42	0.7
Usmani *et al.* (2007)	Pilot	1.5- 21		DLE	13	32-66	1 to 1.25

AA; alopecia areata, AD; Atopic dermatitis, AS; Aphthous stomatitis, CLE; cutaneous lupus erythematosus, CR; case report, DLE; discoid lupus erythematosus, DM; dermatomyositis, FU; follow up, GA; granuloma annulare, HFP; hand & foot psoriasis, HS; hidradenitis suppurativa, LP; Lichen Planus, OLE; open label extention, OLP; Oral Lichen Planus, PPP; palmoplantar pustulosis, PRP; pityriasis rubra pilaris, PS; psoriasis, RCT; randomized controlled trial, VI; vitiligo.
*; not reported.

old woman, who was misdiagnosed with psoriasis and started treatment with efalizumab.[40] During treatment she experienced a dramatic aggravation of her skin lesions and therapy was discontinued. A skin biopsy confirmed clinical diagnosis of PRP. No adverse events other than the flare of PRP were reported.

The quality of the evidence according to the GRADE system is very low. The small amount of available evidence is conflicting.

Hidradenitis suppurativa

Hidradenitis suppurativa (HS) is a chronic, inflammatory condition in which TNF-blocking agents have been shown to be effective, but T cells are not generally believed to play an initial role in this condition.[41]

Despite these considerations, the effect of efalizumab on HS was studied in five women with severe, refractory disease, in a single-center, prospective, open-label clinical trial.[42] Three patients prematurely discontinued from the study due to worsening of the condition, migraine headaches and loss to follow-up. None of the patients derived clinical benefit from therapy. Notable adverse events during treatment were one case of anemia, a ductal breast carcinoma and severe headaches.

According to the GRADE system the quality of evidence is very low. However, these findings might indicate that efalizumab is not useful in HS. Also, this might suggest that this disease is not dependent on T-cell activation or migration to sites of inflammation.

Atopic dermatitis

Atopic dermatitis (AD) is a chronic, relapsing, inflammatory skin disease. Whether atopy is viewed as an immune regulation disorder or as the result of an impaired skin barrier, there is no doubt that in AD, T cells have a role in the formation and inflammation of eczema lesions.[43]

In a pilot study involving ten patients, the investigators reported a more than 50% improvement in the Eczema Area and Severity Index in six patients after 12 weeks of treatment.[44] Three clinically significant adverse events occurred: thrombocytopenia, viral gastroenteritis and a rebound. Thereby, five of the ten patients experienced secondary bacterial infection of the skin and three subjects were diagnosed with eczema herpeticum. All skin infections were mild in nature. Furthermore, five case reports regarding six patients in total were found.[45-49] All patients showed improvement to some extent. One patient was primarily treated for his alopecia areata (AA) when his eczema improved, and he became independent of topical steroids.[47] Unfortunately, outcome assessment

was very heterogeneous. Long-term efficacy of efalizumab was reported in a case study describing a child with AD.[48] Efalizumab was well tolerated in these reports.

Although six studies were conducted showing good efficacy of efalizumab in AD, the quality of evidence is very low according to the GRADE system.

Lichen planus

Lichen planus is a papulosquamous inflammatory disorder that may affect the skin, nails, hair and mucous membranes.[50] T cells are assumed to play a major role in its immunopathogenesis, perhaps activated by human herpes virus type 7.[51]

Three articles were found concerning efalizumab in lichen planus.[52-54] In a pilot study involving four patients with oral, erosive lichen planus, a clinically significant reduction in oral mucosal surface area involvement was reported after 12 weeks of treatment in all patients, as well as improvements in visual analogue scale scores for pain.[54] However, clinically significant adverse events were also reported, including one case of drug-induced subacute cutaneous lupus erythematosus, hospitalization for urticaria and a staphylococcal abscess of an artificial hip joint. One case report of a patient with oral, erosive lichen planus described 75% improvement with the use of efalizumab.[53] In another case report, clearance of the skin lesions was described after 8 weeks of treatment in a patient with widespread, relapsing, generalized cutaneous disease.[52] Both case reports showed minimal adverse events. Findings from case reports and a small pilot study suggest that efalizumab could be effective in lichen planus. The quality of evidence is considered very low according to the GRADE system.

Alopecia areata

Alopecia areata is thought to be an autoimmune disease targeted at the hair follicles, in which T cells play an important pathogenic role.

Five articles are available on the effectiveness and safety of efalizumab in this condition; four case reports and one RCT. In two case reports, AA was the primary target for treatment, whereas in the other two AD was the primary target.[47-49,55] All four reports showed marked improvement of AA. However, a recent RCT investigating the efficacy of efalizumab in 62 patients with moderate-to-severe AA failed to reach statistical significance in the efficacy end points after 6 months of treatment.[56] Details on the quality assessment for this study are given in Table 4. Besides frequent mild adverse events, such as headaches, nausea, myalgia and rash, one case of aseptic meningitis and one case of myocardial infarction

were reported. There has been one reported case of AA that developed during treatment with efalizumab; relatedness is not certain.[57]

According to the GRADE system, the quality of evidence is moderate. It may be concluded that there is evidence for a lack of efficacy of efalizumab in AA.

Aphthous stomatitis

Aphthous stomatitis is characterized by minor and major mouth ulcers that are painful and recurrent. Pathogenesis of the disease is unknown. A case report has described the prevention of developing new aphthous ulcers in a patient who was treated with efalizumab for plaque psoriasis.[58]

The quality of evidence according to the GRADE system is very low.

Granuloma annulare

Granuloma annulare (GA) is a benign granulomatous disease with unknown pathogenesis.

One case report was found describing that efalizumab was effective in GA.[59] A 52-year-old man with a history of moderate to severe plaque psoriasis and subsequent GA showed marked improvement of his GA within 4 weeks of initiation of treatment with efalizumab, and at 4 months of continuous efalizumab therapy it was completely cleared. However, it should be noted that the patient was given subsequent therapy with clofazimine, which may also have some effect in the treatment of GA. Withdrawal of clofazimine after clearing of the GA did not result in a recurrence. Adverse events were not reported.

The quality of evidence according to the GRADE system is very low.

Recalcitrant dermatomyositis

Recalcitrant dermatomyositis is a multisystem idiopathic inflammatory disorder, most commonly affecting the muscles and skin. The effectiveness of efalizumab plus azathioprine and prednisolone in recalcitrant dermatomyositis has been documented in one case report in which treatment resulted in a prominent reduction of cutaneous symptoms.[60] Adverse events were not reported. Also for this study, a very low quality of evidence according to GRADE applies.

Vitiligo

Vitiligo is a chronic and often progressive skin disorder characterized by depigmentation due to melanocyte loss. It is believed to have a T-cell-mediated, autoimmune etiology.[61]

Table 3. Results of data extraction

	Efficacy	Concomitted medication (number of patients)	Reported side effects (number of events)
Palmoplantar pustulosis			
Fretzin *et al.* (2006)	>90% improvement in 12 of 17 patients, no non-responders	Acitretin 10mg/day (1)	Mild flu-like symptoms (2), Folliculitis (1), Haemolytic anaemia (1), Mild myalgia (1)
Cohen *et al.* (2007)	Clearance of affected hands and feet within 5 months	Acitretin 25mg/day (1) Topical urea 40% (1) Cyclosporin 5mg/kg/day (1)	Injection site reaction (1), Mild flu-like symptoms (1), Severe headache (1)
Colsman *et al.* (2008)	Evident improvement in one patient, clearance in the other	Topical mometasone furoate, calcipotriol and betamethasone (1)	Thrombocytopenia (1)
Stinco *et al.* (2008)	Initial improvement, loss of efficacy after 4 months	Acitretin 0.5 mg/kg/day (1)	None
Varma *et al.* (2008)	6 of 10 patients had a beneficial response. mean PGA improvement = 1.2 4 of 10 patients withdrew due to inadequate response		Myalgia (8), Chills (4), Infection (4), Aggravation of psoriasis (3), Fatigue (2), Headache (2), Back pain (1), Fever (1), Foot numbness (1), Herpes zoster (1), Malaise (1), Pruritis (1), Rhinorrhea (1), Swollen ankle (1), Tinnitus (1), Ulcerative colitis (1)
Wozel *et al.* (2008)	Hands clear, feet almost clear after 2.5 months.	Unsure whether MTX 7.5 mg/week was continued (1)	Not reported
Pityriasis Rubra Pilaris			
Klein *et al.* (2007)	Severe aggravation of skin lesions		Aggravation of PRP (1)
Gomez *et al.* (2007)	50% improvement after first admission, remission after 9 months		Not reported
Strober *et al.* (2007)	No clinical benefit in any of the patients		Aggravation of HS (2), Upper respiratory infection (2), Anaemia (1), Chills (1), Ductal breast cancer (1), Fever (1), Headaches (1), Weight loss (1)
Atopic dermatitis			
Farshidi *et al.* (2006)	Decreased prurigo papules, less lichenification and hyperpigmentation	Antihistamines, topical steroids (1)	Mild flu-like symptoms (2)
Hidradenitis suppurrativa			
Kaelin *et al.* (2006)	Marked improvement after 6 months, independence from topical steroids	Topical steroids (1)	None

Table 3. *Continued*

	Efficacy	Concomitted medication (number of patients)	Reported side effects (number of events)
Weinberg *et al.* (2006)	Sustained and substantial improvement	Interferon gamma in first month (1), Topical tacrolimus 0,1%, pimecorlimus 1% and steroids (1)	Minimal adverse events
Hassan *et al* .(2007)	SCORAD from 51 to 9 in 9 months	Topical steroids (1)	None
Siegfried *et al.* (2007)	Clearing of eczema after 18 months		Acute autoimmune thrombocytopenia (1)
Takiguchi *et al.* (2007)	EASI score 37.1± 13.5 to 17.6 ± 14.5 at week 12		Superficial bacterial skin infection (5), Herpes simplex (3), Flu-like symptoms (2), Change in eczema morphology (2), Loose stools (1), Thrombocytopenia (1)
Lichen Planus			
Cheng *et al.* (2006)	75% improvement at week 10	Topical tacrolimus (1)	Minimal adverse events
Böhm *et al.* (2007)	Clearance of skin lesions at 8 weeks		Abdominal discomfort (1), Fatigue (1), Post inflammatory hyperpigmented macules (1), Vertigo (1)
Heffernan *et al.* (2007)	At week 12 improvement of mean CLS of 2, VAS score of 82%, BSA of 71.1%	Topical clobetasol 0.05% (1) Topical tacrolimus 0.1% (1) Prednisolone syrup (1)	SCLE (1), Urticaria (1) Streptococcal abcess left hip (1)
Alopecia areata			
Kaelin *et al.* (2006)	90% regrowth of scalp hair after 6 months of treatment	Topical steroids (1)	None
Weinberg *et al.* (2006)	Partial regrowth of scalp hair after 19 months of treatment	Interferon gamma in first month (1) and steroids (1)	Minimal adverse events
Siegfried *et al.* (2007)	Regrowth of scalp hair after 18 months		Acute autoimmune thrombocytopenia (1)
Kolde *et al.* (2008)	80-100% regrowth of scalp hair after 8 months, no relapse after 6 months		None
Price *et al.* (2008)	12 weeks: 1,7% hair growth in placebo vs -3.3% hair growth in efalizumab group 24 weeks: 8.8% hair growth in placebo/efalizumab vs -6.0% hair growth in efalizumab/efalizumab group	All patients received at least one concomitant medication	Headache (19), Nausea (9), Rash (7), Myalgia (6), Pharyngitis (5), Fever (4), Infection (4), Myocardial infarction (1), Aseptic meningitis (1), Nail disorder (1), Pustular rash (1), Toothache (1)

Aphthous stomatitis			
Zribi (2007)	After 12 weeks no new ulcers developed		Paradoxical papular eruption (1), Aggravation of psoriasis (1)
Granuloma annulare			
Goffe *et al.* (2004)	Marked improvement after 4 weeks, clearance within 4 months	Clofazimine (1)	Not reported
Dermatomyositis			
Huber *et al.* (2006)	Prominent reduction of facial and thoracic oedema	Prednisolone 40 mg/day 1st months, 10mg/day 2nd month, azathioprine 50mg/day(1)	Not reported
Vitiligo			
Fernandez *et al.* (2008)	Clinical improvement of vitiligo with evident repigmentation, psoriasis flared	Methyl prednisolone 32 mg/day for 5 days, MTX 5-7.5mg/week (1)	Massive erythrodermic reaction (1)
Wakkee *et al.* (2008)	Repigmentation of vitiligo universalis, psoriasis minimally improved		Cosmetically undesirable pattern of regimentation (1)
Lupus erythematosus			
Clayton *et al.* (2006)	Dramatical improvement of skin lesions within 6 weeks		Slight flare (1)
Hamprecht *et al.* (2007)	Substantial improvement of malar rash after 4 weeks	Prednisolon 5 mg/day (1)	Elevated anti-dsDNA antibodies (1), Lupus nephritis (1)
Usmani *et al.* (2007)	11 of 13 patients very good to excellent overall response, mean time of treatment response is 5,5 weeks	Oral prednisolone (5) and methyl prednisolone injections (1), dapsone (2), hydroxychloroquine (1), mepacrine (1)	Generalized rash (3), Flare of joints (2), Joint pain (2), Mild flare of skin lesions (2), Headache (1), Loss of appetite (1), Skin infection (2), Vomiting and diarrhea (1)

CLS; clinical lesion score, HS; hidradenitis suppurativa, PGA; physician global assessment, PRP; pityriasis rubra pilaris, SCLE ; subacute cutane lupus erythematosis, VAS; visual analogue scale.

Table 4. Risk of bias of included RCT

	Randomisation	Allocation	Blinding	Attrition	Overall
Price *et al.* (2008)	UNCLEAR	ADEQUATE	Participants ADEQUATE Researchers ADEQUATE Outcome assessment UNCLEAR	Reporting drop-outs ADEQUATE Missing data INADEQUATE	Unclear risk of bias

Two cases have been reported of patients with vitiligo and comorbid psoriasis who have been treated with efalizumab.[62,63] Although the psoriasis flared in one report, evident repigmentation of the vitiligo affected skin occurred. In the other case report, a patient discontinued from treatment with efalizumab due to irregular and thus cosmetically undesirable repigmentation of his vitiligo.

According to the GRADE system, the quality of evidence is very low.

Cutaneous lupus erythematosus

Chronic discoid lupus erythematosus, which is the most common form of cutaneous lupus erythematosus (CLE), and systemic lupus erythematosus are chronic autoimmune diseases.[64,65] Cutaneous lesions are thought to be initiated by ultraviolet damage, opening up autoantigens for binding by autoantibodies, followed by an inflammatory reaction in which T cells are involved. One pilot study and two case reports demonstrated beneficial treatment response with efalizumab.[66-68] In a small retrospective analysis of 13 patients with chronic discoid lupus erythematosus, the general assessment of overall clinical responses to treatment was rated as 'good' to 'excellent' in 12 patients.[68] The mean time to treatment response was 5.5 weeks. One patient dropped out due to severe headaches. A case report of a 47-year old woman with subacute CLE, in whom conventional medication had yielded no improvement, showed that efalizumab was effective within 6 weeks of initiation, with minimal cutaneous disease remaining after 5 months of weekly treatment.[66] In another case report, a 42-year old woman with recalcitrant malar rash due to CLE was treated with efalizumab in combination with oral prednisolone. After initial improvement of her malar rash, she developed a lupus nephritis within 6 weeks of therapy.[67] Also, a case of subacute CLE developing in a patient undergoing efalizumab therapy has been reported.[69]

The quality of evidence is considered very low.

DISCUSSION

As with other new therapies, off-label use in dermatological diseases other than the registered indication was explored. Efalizumab was registered for use in plaque psoriasis in October 2003 by the US FDA, and in September 2004 by the EMA. The first reported off-label administration of efalizumab in patients without concomitant plaque psoriasis was in May 2004, and was published in 2006.[49] Since then many studies followed, of which seven were published in 2006, 12 in 2007 and eight in 2008.

In this review, a total of 151 patients with dermatological disorders other than psoriasis were found, treated with efalizumab for approximately 80 patient-years. Duration of individual treatment ranged from 1 week to 24 months. The dose of efalizumab was generally 1.0 mg/kg/week, but varied from 0.7 to 1.8 mg/kg/week. The mean age of the participants was approximately 39 years. In general, the evidence is very low in quality according to the GRADE system, since mainly case reports and pilot studies were conducted.

The main difficulties in assessing the effectiveness, efficacy and safety of treatment with efalizumab was generally the poor quality of study designs, the heterogeneity of the data, lack in comprehensive reporting and the use of different outcome measures. Primarily case reports and pilot studies were conducted. Case reports and pilot studies are prone to give biased results, such as selection and publication bias. While performing pilot studies, the question remains as to when the point is reached that new pilot studies no longer add to the strength of evidence. In general, performing numerous pilot studies is not of scientific value, as health-related questions are answered based on higher levels of evidence. Administration of off-label drugs in difficult-to-treat diseases is meaningful in order to find new treatment options. The evidence generated from daily practice and pilot studies should be used as a proof of concept and thus could be an indication for further research.

Although efalizumab is suspended from the market, the available evidence indicates that treatment with efalizumab might be efficacious in severe and recalcitrant PPP, AD, lichen planus and CLE. However, it should be kept in mind that these results are solely based on case reports and pilot studies. Only one case report regarding one patient was found for treatment with efalizumab in dermatomyositis, GA and aphthous stomatitis; all showed clinical benefit. Treatment of efalizumab in vitiligo resulted in repigmentation of affected skin in two patients, but unfortunately this was not always cosmetically desirable. Conflicting results were found in studies concerning efalizumab in patients with PRP. Lack of efficacy was found in treatment for HS and AA. For the latter, lack

of efficacy was based on a RCT concerning 64 patients and with a moderate quality rating, while a beneficial response was found in four case reports concerning four patients.

In the light of the developments in psoriasis, we also summarized the demonstrated severe adverse events for all off-label indications in our review. First of all, no cases of PML were reported. Serious adverse events reported were two cases of anemia and thrombocytopenia and one case of staphylococcal abscess of an artificial hip joint, hospitalization for urticaria, myocardial infarction, aseptic meningitis, breast cancer, lupus nephritis, ulcerative colitis and one case of drug-induced subacute CLE. All these patients dropped out of study. Frequently reported mild adverse events were headache in 23 cases, myalgia in 15 cases and mild infections in 29 cases, of which 12 were skin infections. Exacerbations occurred in ten patients. Special consideration should be taken into account regarding skin infections in patients without an intact skin barrier and formation of autoantibodies and flares in patients with CLE.

Thrombocytopenia has been observed in 0.3% of patients receiving efalizumab during clinical trials, whereas haemolytic anemia has been occasionally reported.[8] These adverse effects are not easy to explain knowing that efalizumab is directed at T cells. As with all immunomodulating therapies, there is a risk of malignancies, especially lymphomas and non-melanoma skin cancer. Studies thus far have not indicated an increased incidence of malignancies.[3-7,9-12] Unfortunately, the quality of the evidence on off-label safety data in this review is too low to compare the results with psoriasis safety data.

CONCLUSION

In this article, it has been shown that some patients with difficult-to-treat dermatological diseases and without alternative treatment options available to them, responded well to efalizumab treatment. Thereby, the demonstrated effectiveness of efalizumab in PPP and GA might indicate that T-cell-mediated pathways play a role in their pathogenesis. The information summarized in this review will also be useful when a new drug is developed that targets similar T-cell-mediated pathways as efalizumab. Efficacy of the new drug might therefore be predicted.

The EMA found the additional value of efalizumab in plaque psoriasis to be limited in view of the risk of development of PML. It is not certain whether or not this also applies to other dermatological indications.

With this article, we would like to create awareness among healthcare professionals about the rapid introduction of efalizumab in the off-label setting

and the safety issues concerned. The reason for the withdrawal of efalizumab was four cases of PML. In September 2008, the first case of PML was identified in a patient using efalizumab for 4 years. This unfortunate circumstance emphasizes the need for long-term follow-up and long-term registries for registered indications, as with the Psonet initiative, as well as for unregistered indications.[70]

Many drugs prescribed in dermatological daily practice are off-label, approaching 50% of prescriptions. Currently, national and international regulatory agencies are making new rules and regulations concerning off-label drug use. This has led to the fact that in some countries prescription of off-label medication in the absence of guidelines is prohibited unless pharmacists are consulted and patients sign informed consents. The withdrawal of efalizumab makes clear that the use of new and potentially harmful drugs in an off-label setting is not without risk.[71] With the administration of drugs in unregistered indications, there is an unknown balance between the dose, efficacy and safety pattern. Thereby, prescribing physicians are legally responsible for any consequences.

Five-year view

Thanks to the growing awareness around off-label prescription, there is much opportunity in the next five years to find suitable treatment options for patients with often difficult-to-treat and refractory dermatological conditions or with rare conditions and subsequent scarce therapeutic options. That is, if more reviews on off-label drugs become available to provide evidence for or lack of evidence for efficacy and safety of off-label prescription in the coming years. Subsequently, these reviews will reveal new research areas. Thereby, obligatory registries for off-label treatment should be created to generate long-term effectiveness and safety data. If alarming side effects occur, this could be rapidly identified. Coordination between registries and pharmacovigilance authorities should take place.

Combining the data from these reviews and registries, will allow us to make evidence-based recommendations and to formulate guidelines concerning a specific drug in off-label use. In case of promising results of a new drug in an off-label setting, case reports should be followed by well-designed RCTs using standardized and validated diagnostic and outcome tools to reduce heterogeneity and enable meta-analysis when multiple studies are performed. For rare diseases, trial networking on national and international level will be necessary.

Altogether, this will lead to more evidence-based treatment options for difficult-to-treat and rare diseases and to less exposure of patients to treatments that are not effective and potentially harmful.

REFERENCE LIST

1. Calabrese LH, Molloy ES. Progressive multifocal leucoencephalopathy in the rheumatic diseases: assessing the risks of biological immunosuppressive therapies. *Ann Rheum Dis* 2008; **67 Suppl 3**: iii64-iii65.
2. Schon MP, Zollner TM, Boehncke WH. The molecular basis of lymphocyte recruitment to the skin: clues for pathogenesis and selective therapies of inflammatory disorders. *J Invest Dermatol* 2003; **121**: 951-62.
3. Dubertret L, Sterry W, Bos JD *et al*. CLinical experience acquired with the efalizumab (Raptiva) (CLEAR) trial in patients with moderate-to-severe plaque psoriasis: results from a phase III international randomized, placebo-controlled trial. *Br J Dermatol* 2006; **155**: 170-81.
4. Gordon KB, Papp KA, Hamilton TK *et al*. Efalizumab for patients with moderate to severe plaque psoriasis: a randomized controlled trial. *JAMA* 2003; **290**: 3073-80.
5. Lebwohl M, Tyring SK, Hamilton TK *et al*. A novel targeted T-cell modulator, efalizumab, for plaque psoriasis. *N Engl J Med* 2003; **349**: 2004-13.
6. Leonardi CL, Papp KA, Gordon KB *et al*. Extended efalizumab therapy improves chronic plaque psoriasis: results from a randomized phase III trial. *J Am Acad Dermatol* 2005; **52**: 425-33.
7. Papp KA, Bressinck R, Fretzin S *et al*. Safety of efalizumab in adults with chronic moderate to severe plaque psoriasis: a phase IIIb, randomized, controlled trial. *Int J Dermatol* 2006; **45**: 605-14.
8. Schmitt J, Zhang Z, Wozel G *et al*. Efficacy and tolerability of biologic and nonbiologic systemic treatments for moderate-to-severe psoriasis: meta-analysis of randomized controlled trials. *Br J Dermatol* 2008; **159**: 513-526
9. Gottlieb AB, Gordon KB, Lebwohl MG *et al*. Extended efalizumab therapy sustains efficacy without increasing toxicity in patients with moderate to severe chronic plaque psoriasis. *J Drugs Dermatol* 2004; **3**: 614-24.
10. Gottlieb AB, Hamilton T, Caro I *et al*. Long-term continuous efalizumab therapy in patients with moderate to severe chronic plaque psoriasis: updated results from an ongoing trial. *J Am Acad Dermatol* 2006; **54**: S154-S163.
11. Leonardi C, Menter A, Hamilton T *et al*. Efalizumab: results of a 3-year continuous dosing study for the long-term control of psoriasis. *Br J Dermatol* 2008; **158**: 1107-16.
12. Papp KA, Camisa C, Stone SP *et al*. Safety of efalizumab in patients with moderate to severe chronic plaque psoriasis: review of clinical data. part II. *J Cutan Med Surg* 2005; **9**: 313-23.
13. Higgings JPT, Altman DG. Assessing risk of bias in included studies. In: *Cochrane Handbook for Systematic Reviews of interventions Version 5.0.1*. Higgings JPT, Green S (Eds). The Cochrane Collaboration, Oxford, UK (2008).
14. Guyatt GH, Oxman AD, Kunz R *et al*. What is "quality of evidence" and why is it important to clinicians? *BMJ* 2008; **336**: 995-8.
15. Dedrick RL, Walicke P, Garovoy M. Anti-adhesion antibodies efalizumab, a humanized anti-CD11a monoclonal antibody. *Transpl Immunol* 2002; **9**: 181-6.
16. Gauvreau GM, Becker AB, Boulet LP *et al*. The effects of an anti-CD11a mAb, efalizumab, on allergen-induced airway responses and airway inflammation in subjects with atopic

asthma. *J Allergy Clin Immunol* 2003; **112**: 331-8.

17. Goncalves DP, Laurindo I, Scheinberg MA. The appearance of pustular psoriasis during antitumor necrosis factor therapy. *J Clin Rheumatol* 2006; **12**: 262.
18. Harper EG, Simpson EL, Takiguchi RH *et al*. Efalizumab therapy for atopic dermatitis causes marked increases in circulating effector memory CD4+ T cells that express cutaneous lymphocyte antigen. *J Invest Dermatol* 2008; **128**: 1173-81.
19. Jacquemin MG, Machiels JJ, Lebrun PM *et al*. Successful treatment of atopic dermatitis with complexes of allergen and specific antibodies. *Lancet* 1990; **335**: 1468-9.
20. Su C, Teitelman M. Anti-TNF therapy: Is this our answer to pyoderma gangrenosum. *Gastroenterology* 2006; **131**: 1347-9.
21. Callen JP. Immunomodulatory treatment for dermatomyositis. *Curr Allergy Asthma Rep* 2008; **8**: 348-53.
22. Chen SL, Yan J, Lui Y *et al*. Efalizumab for psoriasis: A systematic review. *Chinese Journal of evidence-based Medicine* 2006; **6**: 267-72.
23. Gottlieb AB. Therapeutic options in the treatment of psoriasis and atopic dermatitis. *J Am Acad Dermatol* 2005; **53**: S3-16.
24. Grassi W, Montecucco C. [Approved indications and other possible utilizations of anti-TNFalpha biologic drugs.]. *Reumatismo* 2005; **57 Suppl 1**: 30-3.
25. Graves JE, Nunley K, Heffernan MP. Off-label uses of biologics in dermatology: rituximab, omalizumab, infliximab, etanercept, adalimumab, efalizumab, and alefacept (part 2 of 2). *J Am Acad Dermatol* 2007; **56**: e55-e79.
26. Guhl G, az-Ley B, Fernandez-Herrera J. [Off-label use of biologic agents in the treatment of dermatosis, part 2: etanercept, efalizumab, alefacept, rituximab, daclizumab, basiliximab, omalizumab, and cetuximab]. *Actas Dermosifiliogr* 2008; **99**: 5-33.
27. Heymann WR. Antipsoriatic biologic agents for the treatment of atopic dermatitis. *J Am Acad Dermatol* 2007; **56**: 854-5.
28. Kerns MJ, Graves JE, Smith DI *et al*. Off-label uses of biologic agents in dermatology: a 2006 update. *Semin Cutan Med Surg* 2006; **25**: 226-40.
29. O'Neill ID. Off-label use of biologicals in the management of inflammatory oral mucosal disease. *J Oral Pathol Med* 2008.
30. Prens EP. Off-label use and side-effects of biologics. *Ned Tijdschr voor Dermatologie en Venereologie* 2008; **18**: 118-9.
31. Wenzel J, Bieber T, Uerlich M *et al*. [Systemic treatment of cutaneous lupus erythematosus]. *J Dtsch Dermatol Ges* 2003; **1**: 694-704.
32. Wakkee M, Assen YJ, Thio HB *et al*. Repigmentation of vitiligo during efalizumab. *J Am Acad Dermatol* 2008; **59**: S57-S58.
33. Cohen DJ, Scherschun L. Case reports: practical experience with efalizumab in hand and foot psoriasis. *J Drugs Dermatol* 2007; **6**: 1224-30.
34. Colsman A, Carrascosa JM, Ferrandiz C *et al*. Successful treatment of recalcitrant palmoplantar psoriasis with efalizumab. *J Eur Acad Dermatol Venereol* 2008; **22**: 1131-4
35. Fretzin S, Crowley J, Jones L *et al*. Successful treatment of hand and foot psoriasis with efalizumab therapy. *J Drugs Dermatol* 2006; **5**: 838-46.
36. Stinco G, Piccirillo F, Patrone P. Treatment of palmoplantar pustular psoriasis with efalizumab: a quick response with early recurrence. *Eur J Dermatol* 2008; **18**: 91-2.

37. Varma R, Cafardi JA, Cantrell W *et al*. Safety and efficacy of subcutaneously administered efalizumab in adults with moderate-to-severe hand and foot psoriasis: an open-label study. *Am J Clin Dermatol* 2008; **9**: 105-9.

38. Wozel G, Vitez L. Palmoplantar pustular psoriasis: successful therapy with efalizumab after non-response to infliximab. *Acta Derm Venereol* 2008; **88**: 169-70.

39. Gomez M, Ruelas ME, Welsh O *et al*. Clinical improvement of pityriasis rubra pilaris with efalizumab in a pediatric patient. *J Drugs Dermatol* 2007; **6**: 337-9.

40. Klein A, Szeimies RM, Landthaler M *et al*. Exacerbation of pityriasis rubra pilaris under efalizumab therapy. *Dermatology* 2007; **215**: 72-5.

41. Mekkes JR, Bos JD. Long-term efficacy of a single course of infliximab in hidradenitis suppurativa. *Br J Dermatol* 2008; **158**: 370-4.

42. Strober BE, Kim C, Siu K. Efalizumab for the treatment of refractory hidradenitis suppurativa. *J Am Acad Dermatol* 2007; **57**: 1090-1.

43. Bos JD. Proliferating prevalence of concepts and controversies in atopic eczema. *J Cosmet Dermatol* 2004; **3**: 243-4.

44. Takiguchi R, Tofte S, Simpson B *et al*. Efalizumab for severe atopic dermatitis: a pilot study in adults. *J Am Acad Dermatol* 2007; **56**: 222-7.

45. Farshidi A, Sadeghi P. Successful treatment of severe refractory atopic dermatitis with efalizumab. *J Drugs Dermatol* 2006; **5**: 994-8.

46. Hassan AS, Kaelin U, Braathen LR *et al*. Clinical and immunopathologic findings during treatment of recalcitrant atopic eczema with efalizumab. *J Am Acad Dermatol* 2007; **56**: 217-21.

47. Kaelin U, Hassan AS, Braathen LR *et al*. Treatment of alopecia areata partim universalis with efalizumab. *J Am Acad Dermatol* 2006; **55**: 529-32.

48. Siegfried EC. Long-term follow-up of a child treated with efalizumab for atopic dermatitis. *Arch Dermatol* 2007; **143**: 1077-8.

49. Weinberg JM, Siegfried EC. Successful treatment of severe atopic dermatitis in a child and an adult with the T-cell modulator efalizumab. *Arch Dermatol* 2006; **142**: 555-8.

50. Sugerman PB, Savage NW, Walsh LJ *et al*. The pathogenesis of oral lichen planus. *Crit Rev Oral Biol Med* 2002; **13**: 350-65.

51. De Vries HJ, van MJ, Teunissen MB *et al*. Lichen planus is associated with human herpesvirus type 7 replication and infiltration of plasmacytoid dendritic cells. *Br J Dermatol* 2006; **154**: 361-4.

52. Bohm M, Luger TA. Lichen planus responding to efalizumab. *J Am Acad Dermatol* 2007; **56**: S92-S93.

53. Cheng A, Mann C. Oral erosive lichen planus treated with efalizumab. *Arch Dermatol* 2006; **142**: 680-2.

54. Heffernan MP, Smith DI, Bentley D *et al*. A single-center, open-label, prospective pilot study of subcutaneous efalizumab for oral erosive lichen planus. *J Drugs Dermatol* 2007; **6**: 310-4.

55. Kolde G, Meffert H, Rowe E. Successful treatment of alopecia areata with efalizumab. *J Eur Acad Dermatol Venereol* 2008; **22**: 1519-20

56. Price VH, Hordinsky MK, Olsen EA *et al*. Subcutaneous efalizumab is not effective in the treatment of alopecia areata. *J Am Acad Dermatol* 2008; **58**: 395-402.

57. Tosti A, Pazzaglia M, Starace M *et al*. Alopecia areata during treatment with biologic agents. *Arch Dermatol* 2006; **142**: 1653-4.

58. Zribi H, Crickx B, Descamps V. Prevention of recurrent aphthous stomatitis by efalizumab (Raptiva). *J Eur Acad Dermatol Venereol* 2007; **21**: 1286-7.
59. Goffe BS. Disseminated granuloma annulare resolved with the T-cell modulator efalizumab. *Arch Dermatol* 2004; **140**: 1287-8.
60. Huber A, Gaffal E, Bieber T *et al*. Treatment of recalcitrant dermatomyositis with efalizumab. *Acta Derm Venereol* 2006; **86**: 254-5.
61. Chow S, Rizzo C, Ravitskiy L *et al*. The role of T cells in cutaneous autoimmune disease. *Autoimmunity* 2005; **38**: 303-17.
62. Fernandez-Obregon AC. Clinical management with efalizumab of a patient with psoriasis and comorbid vitiligo. *J Drugs Dermatol* 2008; **7**: 679-81.
63. Wakkee M, Thio HB, Neumann FJ. Is efalizumab effective in the treatment of vitiligo? *Ned Tijdschr voor Dermatologie en Venereologie* 2008; **18**: 131-2.
64. Wenzel J, Uerlich M, Worrenkamper E *et al*. Scarring skin lesions of discoid lupus erythematosus are characterized by high numbers of skin-homing cytotoxic lymphocytes associated with strong expression of the type I interferon-induced protein MxA. *Br J Dermatol* 2005; **153**: 1011-5.
65. Wenzel J, Zahn S, Mikus S *et al*. The expression pattern of interferon-inducible proteins reflects the characteristic histological distribution of infiltrating immune cells in different cutaneous lupus erythematosus subsets. *Br J Dermatol* 2007; **157**: 752-7.
66. Clayton TH, Ogden S, Goodfield MD. Treatment of refractory subacute cutaneous lupus erythematosus with efalizumab. *J Am Acad Dermatol* 2006; **54**: 892-5.
67. Hamprecht A, Tuting T, Bieber T *et al*. Successful treatment of recalcitrant malar rash in a patient with cutaneous lupus erythematosus with efalizumab. *Clin Exp Dermatol* 2008; **33**: 347-8.
68. Usmani N, Goodfield M. Efalizumab in the treatment of discoid lupus erythematosus. *Arch Dermatol* 2007; **143**: 873-7.
69. Bentley DD, Graves JE, Smith DI *et al*. Efalizumab-induced subacute cutaneous lupus erythematosus. *J Am Acad Dermatol* 2006; **54**: S242-S243.
70. Lecluse LL, Naldi L, Stern RS *et al*. National Registries of Systemic Treatment for Psoriasis and the European 'Psonet' Initiative. *Dermatology* 2009; **218**: 347-56
71. Nijsten T, Spuls PI, Naldi L *et al*. The misperception that clinical trial data reflect long-term drug safety: lessons learned from Efalizumab's withdrawal. *Arch Dermatol* 2009; **145**: 1037-9.

ME Schram
RJ Borgonjen
CMJM Bik
JG van der Schroeff
JJE van Everdingen
PhI Spuls

Off-label use of azathioprine in Dermatology: a systematic review.
Archives of Dermatology 2011;147:474-488

5.3 OFF-LABEL USE OF AZATHIOPRINE IN DERMATOLOGY: A SYSTEMATIC REVIEW

SUMMARY

Objective: To summarize evidence regarding the effectiveness, efficacy and safety of off-label azathioprine (AZA) use in dermatology.

Data Sources: We searched the MEDLINE (1950- 2009), EMBASE (1980-2009), and CENTRAL (1996- 2009) databases on October 9, 2009. The main search terms were *azathioprine* and its synonyms. No restrictions were imposed regarding publication date. Only articles in English, French, German or Dutch were included.

Study Selection: Randomized controlled trials, cohorts and case series concerning the use of AZA in an off-label dermatologic setting were independently assessed for eligibility by 2 co-authors. The search retrieved 3870 articles and 148 articles were selected for detailed review.

Data Extraction: Forty-three articles matching the inclusion and exclusion criteria were reviewed for methodologic quality by 2 reviewers independently, including an evaluation of components associated with biased estimates of treatment effect.

Data Synthesis: High-quality evidence (level A) was found for a moderate therapeutic effect in severe atopic dermatitis (AD). Evidence of moderate quality (level B) was found for efficacy in parthenium dermatitis (PD), bullous pemphigoid (BP), chronic actinic dermatitis (CAD), and leprosy type 1 reaction (Lep T1R). Furthermore, favorable therapeutic effects existed for erythema multiforme, lichen planus, and pityriasis rubra pilaris, although the quality of evidence was low (level C).

Conclusions: A strong clinical recommendation was given for AZA in AD. Conclusions regarding safety in an off-label setting could not be reached because of scarce and incomplete data (level C evidence). Long-term registries and prospective studies could add to the existing evidence and provide legal support for off-label drug use in dermatology.

INTRODUCTION

Azathioprine (AZA) is an immune-modulating drug widely used in medicine today. Originally, AZA was developed and used to prevent graft rejection in transplant surgery, but soon after its introduction, application expanded to a range of autoimmune and immune-mediated diseases. The therapeutic effect is thought to derive from its purine antimetabolism, and AZA is often seen as a corticosteroidsparing agent. In the United States, AZA use in dermatology is off-label. Although in most European countries the use of AZA is licensed for the treatment of pemphigus vulgaris and dermatomyositis (http://www.eudrapharm.eu), it is extensively used off-label.

In an off-label setting, patients are treated with drugs that are not registered for that specific indication.[1] Only pharmaceutical companies can apply for registration of drugs for specific indications. However, the interests of pharmaceutical companies and those of physicians and patients often differ regarding the registration of drugs in difficult-to-treat and refractory dermatologic conditions or in rare conditions with subsequently scarce therapeutic options. As a result, in daily practice, many drugs are prescribed off-label, approaching 50% of prescriptions in dermatology.[1]

International and national regulatory agencies are applying new rules and regulations concerning off-label medication use (http://www.fda.gov, http://www.mhra.gov.uk, and http://www.rivm.nl). Moreover, dermatologists can be held accountable for any unfavorable effects that result from off-label prescription.

To ensure evidence-based health care decision making, guidelines should be developed for drugs that are often prescribed off-label. Because AZA is one such drug, we aimed to summarize the available evidence in the literature regarding the efficacy, effectiveness and safety of off-label treatment with AZA in dermatologic patients. On the basis of this evidence and taking clinical experience into account, we formulated clinical recommendations for the application of AZA in specific conditions.

METHODS

Inclusion and exclusion criteria

Included were randomized controlled trials (RCTs), cohort studies and case series with more than 5 patients in which dermatologic patients treated with off-label AZA were assessed to determine the efficacy, effectiveness and/or safety of the drug with or without concomitant therapy. Excluded were reviews,

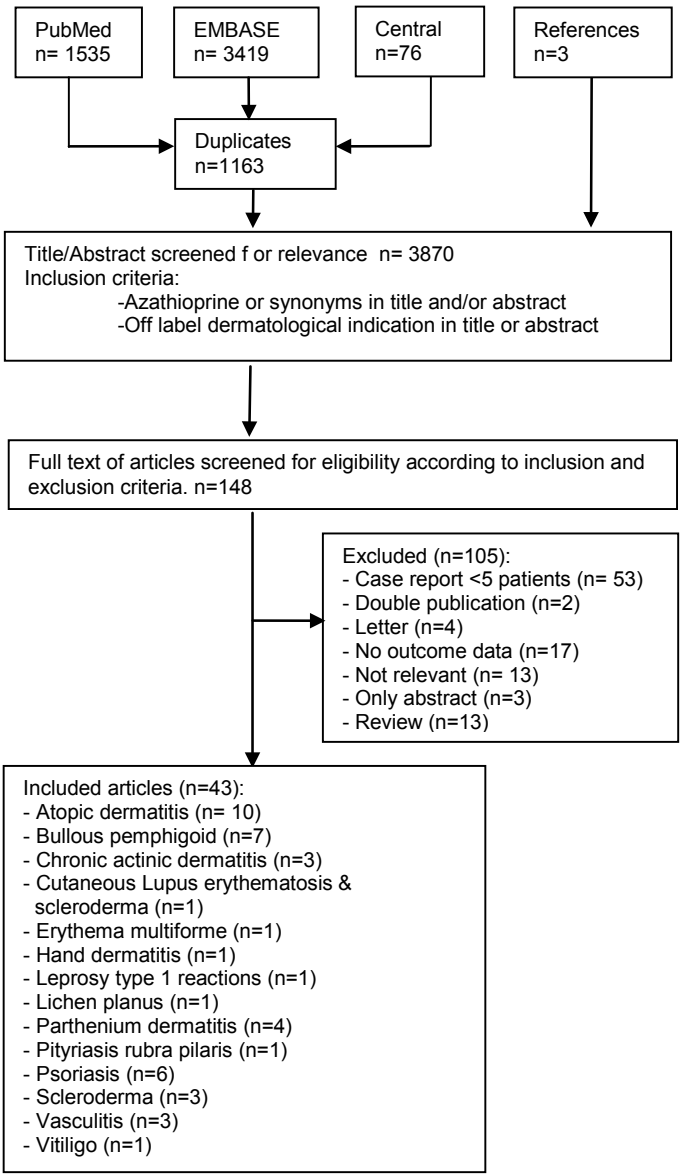


Figure. Flowchart summarizing the selection process for studies concerning off-label azathioprine treatment in dermatological diseases.

case series with fewer than 5 patients, animal and in vitro studies, abstracts, unpublished data, articles concerning topical treatment, articles concerning diseases primarily treated by other specialists, and duplicate publications. No restrictions were imposed regarding age, sex, skin type, or date of publication. Only articles in English, French, German, or Dutch were included.

Literature Search

On October 9, 2009 a literature search of the MEDLINE (1950-2009), EMBASE (1980-2009) and CENTRAL databases (1996-2009) was performed. The keyword *azathioprine* and its synonyms were used in combination with the names of skin diseases (Table 1). Reference lists of included articles and reviews were hand-searched to find additional eligible articles.

Table 1. Search strategy for MEDLINE

Search no.	Term(s)
1	derm*.jn.
2	Azathioprine/
3	(azathioprine* or imuran* or immuran* or imurel*).ab.
4	(azathioprine* or imuran* or immuran* or imurel*).ti.
5	(azathioprine* or imuran* or imurel* or immuran*).kw.
6	4 or 2 or 5 or 3
7	6 and 1
8	exp Skin Diseases/
9	6 and 8
10	7 or 9
11	limit 10 to (humans and (dutch or english or french or german))

Study Selection and Data Extraction

Of the articles located, the title and/or abstract was screened for off-label AZA treatment in dermatologic patients and selected if considered potentially relevant. To determine eligibility, the full text of the selected articles was assessed according to predefined inclusion and exclusion criteria. Data concerning methodologic quality, study characteristics, efficacy and safety were extracted using a data extraction form. Study characteristics included study population (number, age range, sex of included patients, dermatologic disease concerned, previous treatments, and inclusion criteria for participation), design

of the study (study type, duration of treatment, dosage of AZA, and duration of follow-up period), publication date, safety and severity outcome measurements used, and subsequent outcomes, including onset of action and duration of remission. Preferably, changes in mean objective outcome measurements from baseline to end of active treatment were reported. If this was not possible, percentages of patients with a specific outcome were presented.

All stages of study selection and data extraction were performed by 2 reviewers (M.E.S. and C.M.J.M.B.) independently. Disagreements regarding study selection and data extraction were solved by discussion.

Data analysis

We assessed RCTs according to the criterion grading system described in the *Cochrane Handbook for Systematic Reviews of Interventions 5.0.2* (updated in February 2008).[2] To assess the risk of bias, the following measurements were assessed: sequence generation, concealment of allocation, blinding (of participants, researchers, and outcome assessment), handling of withdrawals and losses, and other potential threats to validity. Each item was scored as adequate, inadequate, or unclear. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist for observational studies was used to identify methodologic errors in cohort studies.[3] Case reports were not assessed for methodologic quality.

Strenght of evidence

Each reviewer graded the total body of evidence found for each disease according to the grading system proposed by Robinson *et al.*[4] Because it is inappropriate to formulate safety conclusions based on information derived from studies with relatively small case numbers and short follow-up periods, we compared the found adverse events (AEs) and serious adverse events (SAEs) with the safety data from the Summary of Product Characteristics dated May 27 2009 http://www.medicines.org.uk/emc/medicine/2881/SPC/Imuran_Tablets_25mg/). Particular attention was given to the identification of AEs that seemed to be specific for an indication.

Clinical recommendations

The clinical recommendations are based on the available evidence derived from the literature in combination with considerations such as known AEs, patient preferences, availability of other treatment options, organizational aspects, social consequences and costs. Two reviewers (M.E.S. and C.M.J.M.B.)

independently assessed the methodologic quality, strength of evidence, and magnitude of effect and generated clinical recommendations balancing the benefits, burdens, risks, and costs. All details were discussed with the members of the working group.

RESULTS

Study search

The Figure summarizes the selection process. An initial search retrieved 3870 articles. After screening the title and/or abstract for eligibility, 148 articles were selected. The main reason for exclusion was lack of relevance (eg, use of AZA in registered indications). After screening the full texts of the articles, 43 articles matched our inclusion and exclusion criteria, of which 11 were RCTs, 2 were cohort studies, and 30 were case series.[5-47] In total, 1128 patients with 12 different dermatologic diseases were described. Study demographics are listed in Table 2.

Data extraction

Concomitant treatment was allowed in most studies, not reported in 9 studies, and not allowed in 3. Because AZA is often regarded and used as a corticosteroid-supporting and sparing agent, oral corticosteroids were most frequently used as concomitant medication (n=20). Oral corticosteroids were also used to support patients between initiation and the onset of action of AZA, after which the dosage was tapered. Topical corticosteroids were allowed in 9 studies. Diagnostic criteria and disease definitions were often not specified. Also, the definition of the onset of action was mostly absent.

The severity outcome measures used were heterogeneous. Objective and validated severity outcome measurements were used in only 11 (25.6%) of the studies and included the Six Area, Six Sign Atopic Dermatitis (SASSAD) severity score; Severity Scoring of Atopic Dermatitis index; Psoriasis Area and Severity Index; modified Rodnan skin score; scores on visual analogue scales; and percentage of body surface area affected by disease.[48-50] Non-validated outcome measures included the modified Psoriasis Area and Severity Index and the Clinical Severity Score. In most of the studies, only descriptive means were given to indicate changes in disease severity.

In 4 studies, thiopurine methyltransferase (TPMT) activity was measured before the initiation of AZA therapy. In 2 of these studies, the dosage of AZA was adjusted to the degree of TPMT activity, and efficacy was shown not to be altered by this.[32,33] Onset of action varied from 1 to 28 weeks and occurred on

Table 2. Characteristics of included articles

Disease	Reference (year of publicatiuon)	Study design	Treatment arms	Mean (SD) Treatment duration, mo	FU Duration, mo
AD	8. (2002)	RCT	AZA/PCB PCB/AZA	3	3 -
AD	33. (2006)	RCT	AZA PCB	3	6.5
AD	22. (2009)	CS Pros		6	-
AD	32. (2001)	CS Pros		1-4 (2)	6
AD	9. (1998)	CS Pros		12-84	8
AD	23. (2008)	CS		4-35	0-216
AD	30. (2005)	CS		0.3-108 (24)	120
AD	35. (2002)	CS		(20.5)	-
AD	26.(2002)	CS		24	unk
AD	28. (1996)	CS		1-21	12
BP	7. (2007)	RCT	Pred plus AZA Pred plus MM	24[1]	10
BP	19. (1993)	RCT	Pred plus AZA Pred plus PE Pred	6	-
BP	11. (1978)	RCT	Pred plus AZA Pred	36	-
BP	5. (1977)	Cohort	Pred plus AZA Pred	unk[1]	8-45
BP	10. (1974)[2]	CS			48
BP	41. (1973)	CS		2-30	-
BP	18. (1971)	CS		unk[1]	7
CAD	34. (1989)	RCT	AZA PCB	1.5-12 (8.4)	unk
CAD	47. (2003)	CS		Unk[1]	unk
CAD	29. (1984)	CS		1.5-33 (11.5)	unk
CV	20. (1991)	RCT	Pred plus AZA Conventional	3	29
CLV[3]	13. (1991)	CS Pros		unk[1]	unk
CLV	12. (1987)	CS		unk	unk
SCLE CCLE	13. (1991)	CS Pros		unk[1]	unk
EM	16. (1995)	CS		15-60[1]	0-12
HD	38. (2009)	CS		unk	3-48
Lep T1R	31. (2004)	RCT	Pred plus AZA Pred	3	3

No. of patients ((M/F)	Age of patients, Mean (SD) [range], y	Daily dose of AZA
37 (25/12)	38 (17-73)	2.5mg /kg
		2.5mg /kg
41 (19/22)	30 (11)	0.5-2.5 mg/kg
20 (16/4)	36 (12)	-
17 (9/8)	16.1 [9-22]	1.2-3.5 mg/kg
12	[18-53]	0.5-2.5 mg/kg
10 (5/5)	39.7 [28-56]	0.7-2.5mg /kg
37 (17/20)	43 [19-83]	0.7-2.5 mg/kg
24	29 [13-48]	1.5-3.0 mg/kg
48 (28/20)	6.9 [6-16]	2.5-3.5 mg/kg
38 (21/17)	42.4 [23-87]	25-200 mg
35 (24/11)	31 [5-70]	100 mg
38 (12/26)	75.3 (12.8)	2.0 mg/kg
35 (15/20)	74.8 (13.4)	-
36 (19/17)	77.2 (8.3)	100-150mg
31 (14/17)	74.8 (10)	-
31 (17/14)	75.4 (11)	-
12 (6/6)	75.6 (unk)	2.5 mg/kg
13 (3/10)	74.1 (unk)	-
15	unk [43-92]	1.5 mg/kg
14		
10 (3/7)	68.2 [53-87]	2.5 mg/kg
5 (3/2)	81.4 [70-91]	75-250 mg
12	Unk [47-79]	2.5 mg/kg
8 (6/2)	64 [48-75]	50 mg
10 (10/0)	66 [48-84]	-
12 (unk)	62.7 [26-85][5]	1.0-2.5 mg/kg
14 (11/3)	69.1 [55-80]	100-200 mg
8 (5/3)	66 [51-86]	2 mg/kg
9 (2/7)	65 [37-81]	-
6 (2/4)	32.2 [20-50]	150-250 mg
8	Unk	150 mg
6 (2/4)	42.5 [31-62]	100-150mg
5 (2/3)	29.6 [16-39]	100-150mg
6 (5/1)	49.0 [34-60]	unk
21 (18/3)	46 [22-65]	3 mg/kg
19 (17/2)	41 [20-65]	-

Table 2. *Continued*

Disease	Reference (year of publicatiuon)	Study design	Treatment arms	Mean (SD) Treatment duration, mo	FU Duration, mo
LP	43. (2001)	CS		3-7 (5)	6-9
PD	44. (2008)	RCT	AZA Betamethason	6	6
PD	42. (2000)	Cohort	AZA AZA AZA	6-36	36
PD	45. (2006)	CS		6	-
PD	40. (1998)	CS		6	9-12
PRP	24. (1972)	CS		2-48	18-36
PsA	27. (1990)	CS		32.5 (34)[4]	0-93
Ps	15. (1974)	CS		2-31[4] (21)	unk
PsA	6. (1973)	CS		unk	unk
Ps	46. (1972)	CS		2.5- 7.5	24
Ps	17. (1970)	CS		1,5	-
Ps & PsA	21. (1970)	CS		0.7 – 23 (2)	unk
SSc	36. (2006)	RCT	AZA CYC	30	-
SSc	37. (2007)	CS		12	-
ScD/SSc	14. (1973)	CS		0.5- 1.7	-
ScD	25. (1968)	CS		5-23	-
Vi	39. (2006)	RCT	PUVA plus AZA PUVA	4	-

For all abreviations, see list at end of article
CYC; cyclophosphamide, FU; follow-up, MM; mycofenolaat mofetyl, PE; plasma exchange, Pred; prednisolone, unk; unknown, y;years
- = not applicable
[1] Depended highly on individual clinical response
[2] Follow-up study. Same patients used as in reference 18.
[3] Severe and recalcitrant disease with mild systemic involvement
[4] Treatment until remission was achieved
[5] age of total group

average 4 to 6 weeks after the initial dosage. Duration of remission varied from 0 to 93 months.

Study results are summarized in Table 3. Studies reporting AZA efficacy in patients with 2 different dermatologic diseases were discussed separately for each disease.

No. of patients ((M/F)	Age of patients, Mean (SD) [range], y	Daily dose of AZA
9(4/5)	32.2 [5-54]	100 mg
20 (13/7)	48.5 [25-60]	100 mg
21 (16/5)	50.7 [26-73]	
22 (2/20)	31-75	100mg/day
11 (5/6)		50 mg/day+300mg/28d
10 (10/0)		100 mg/day+300mg/28d
12 (10/2)	53.5 [39-65]	300 mg/ week
20 (19/1)	54 [40-72)	100-150 mg
5(4/1)	55.2 [47-67]	50-200 mg
11(6/5)	46.3 (11.2)	Mean 2.14mg/kg
29	51 [23-79]	100-300 mg
11 (9/2)	47.2 [27-68]	25-150 mg
66 (42/24)	[20-60]	50-100 mg
10 (2/8)	54.3 [31-73]	2.5 mg/kg
20 (11/9)	50.6 [23-8]	50-450 mg
30 (4/26)	36 [19-63]	2.0-2.5 mg/kg
30 (3/27)	38 [20-65]	2.0 mg/kg
13 (1/12)	38 [23-53]	50-100 mg
5 (2/3)	34.2 [9-71]	50-150 mg
20 (14/6)	43.5 [12-65]	50-250 mg
30(unk)	unk	0.6-0.75 mg
30(unk)		-

Methodologic quality and recommendations

As seen in the risk of bias table, the methodologic quality of the included RCTs varied substantially (Table 4). Adequate randomization was performed in 7 (63.6%) of the studies and was unclear or not performed in 4 (36.4%). Concealment of allocation was judged unclear in 3 studies (27.3%) and inadequate in 2 (18.2%). Blinding was judged adequate in 3 studies (27.3%) and was inadequate or unclear in 8 (72.7%). Summary of the quality of evidence and the evidence-based recommendations for clinical practice are given in Table 5.

Table 3. Results

Disease	Reference (year)	Efficacy/effectiveness	Mean onset of action, wk
AD	8.(2002)	Reduction disease severity SASSAD: AZA: 10.2 points (26%) versus PCB: 1.0 points (3%) Between groups: $p<0.01$; ITT	unk
AD	33. (2006)	Reduction disease severity SASSAD: AZA 12.0 points (37%) versus PCB 6.6 points (20%) Difference: 5.4 (95%CI 1.4-9.3), and 17% (95%CI 4.3-29%); ITT	unk
AD	22. (2009)	Mean SCORAD reduction 18.0 points (27%) after 6 months	unk
AD	32. (2001)	Mean SASSAD improvement 12.3 units (95% CI 1.0-23.7)	4
AD	9. (1998)	5 (50%) complete remission 3 (30%) refractory after initial response 1 (10%) good reaction, but persistent sites 1 (10%) poor reaction	unk
AD	23. (2008)	15 (40.5%) complete remission 9 (24%) good response, no remission 7 (19%) under treatment, no remission 5 (13.5%) stopped due to SAE 1 (3%) no response	8
AD	30. (2005)	20 (83.3%) patients 94% improvement of BSA 2 (8.3%) lack of response 2 (8.3%) stopped due to SAE	8 [range 2-28]
AD	35. (2002)	28 (58.3%) excellent improvement 13 (27.1%) good improvement 7 (18.4%) inadequate improvement	4 [range 2-6]
AD	26. (2002)	30 (79%) good response 4 (10.5%) stopped due to SAE 4 (10.5%) lack of response	3.7 [range 1-8]
AD	28. (1996)	18 (69.2%) good response 3 (8.6%) little effect 3 (8.6%) stopped due to SAE	[range 4- 24]
BP	7. (2007)	100% complete remission after a median duration of 23.8 (SD 18.9) days in the Pred & AZA group vs 100% complete remission after a median duration of 42.0 (SD 55.3) days in Pred & MM group. (p=0.09); ITT	unk

	Duration of remission, mo	Concomitant medication (no. of patients)	AE's (no. of events)	SAE's (n of events)
	-	TS (all)	Liver enzymes ↑(8), URTI(5), Fatigue (3), Yellow urine (3), Migraine (2), Malaise (2), Folliculitis (2), Lymphopenia (2), Arthalgia (1), Asthma(1), Bruising (1), Depression (1), Hay fever (1), Headache (1), Impetigo (1), Neutropenia (1), Painful foot (1), Shaking (1), Sore tongue (1), Yawning (1)	None
	3	TS (all)	>1 lymphopenia (28), Nausea (17), ALAT increase (6), Severe nausea (4), Headaches (5), Abdominal pain (4), Lightheadedness (3), Folliculitis (3), Drug hypersensitivity (2), LRTI (2), URTI (2), >1 neutropenia (2), Malaise(1)	None
	-	TS (all)	Liver enzymes ↑(3)	None
	3-6	unk	Liver enzymes ↑((2)	None
	>36	unk	Transient lymphopenia (3), Transient gamma-glutamyl transferase increasement (1), Lethargia (1), Recurrent herpes simplex labialis (1)	Non Hodgekin Lymfoma (1)*
	1-35	TS (all)	Lymphopenia (19), Neutropenia (5), Nausea (4) Persistant lymphopenia (2), Flu-like reaction (1), Pancytopenia (1), abdominal cramps (1), myalgia (1), Severe nausea and vomiting (1)	None
	9	unk	Pancytopenie (1), Abdominal distress (1), Acute bronchitis (1), Severe herpes labialis (1), Liver enzymes ↑((1)	Pancreatitis (1) ruptured aneurysm (1)†*
	-	OS (23)#	Lymphopenia (15), Thrombocytopenia (1), Liver enzymes ↑((5), Microcytosis (3), Eczema herpeticum (1), Nausea, vomiting, diarroea (1), Hypersensitivity(1)	None
	-	TS (all)	Nausea (5), Fatigue (5), Leukopenia (3) Myalgia (2), Drug eruption (2), Megaloblastic anemia (1)	Severe pancytopenia (1)
	12 (61.5% patients)	TS (all)	Severe nausea and epigastric pain (3) Mild nausea (3)	None
	5.8 (SD 4.8)	OS 0.5 mg/kg (all)#	Liver enzymes ↑((106), Dizziness (2), Hyperglycemia (1), Myalgia (1)§	Death(2) Infection (1) Abnormal liver function (2)

Table 3. *Continued*

Disease	Reference (year)	Efficacy/effectiveness	Mean onset of action, wk
BP	19. (1993)	Percentage controlled disease AZA & Pred: week 4: 80.5% (95% CI 68-93%) 6 months: 39.0% (95% CI 23-55%) Pred: week 4: 71% (95% CI 55-87%) 6 mo: 42% (95% CI 25-59) Pred & PE: week 4: 71% (95% CI 55-87) 6 mon: 29% (95% CI 13-45%) No statistically significant difference between groups at 4 weeks and 6 mo. PP	unk
BP	11. (1978)	7 (58%) remission in Pred plus AZA group versus 3 (23%) in Pred group. Not significant. Significant reduction in prednisone dose (45%) in Pred plus AZA group.	unk
BP	5. (1977)	50% reduction in maintenance dose of prednisolone in AZA group 30% reduction in length of medicinal therapy in AZA group	unk
BP	10. (1974)	4 (40%) well controlled/remission 1 (10%) moderate response 1 (10%) Intolerant 4 (40%) Deceased	unk
BP	41. (1973)	2 (40%) very good, no new blisters 2 (40%) good, occasional new blister 1 (20%) withdrew after 2 weeks	4-6
BP	18. (1971)	9 (82%) withdrawal of OS without relapse 2 (18%) considerable reduction of pred dose	unk
CAD	34. (1989)	*Mean reduction in rash VAS*: AZA; 4.9 , PCB; -1.1, $p<0.05$ Difference: 6.0 (95%CI 1.7-10.2) Mean reduction in BSA of rash: AZA; 34.4%, PCB; -27.8% $p<0.05$ Difference: 62.2 (95% 11.8-112.5); ITT	< 4
CAD	47. (2003)	11 (92%) Partial/good response 1 (8%) Withdrew	4-6
CAD	29. (1984)	9 (57%) Cleared/improved 2 (14%) Cleared, but refractory 2 (14%) No response 2 (14%) Withdrew	6-20
CV	20. (1991)	7 (88%) remission AZA ($p<0.05$). 6 (55%) remission conventional ($p<0.05$). No sign. diff. between groups	unk
CLV	13. (1991)	2 (33%) complete response 3 (50%) Partial response 1 (17%) No response	4-6
CLV	12. (1987)	8 (100%) responsive to treatment	4-8

	Duration of remission, mo	Concomitant medication (no. of patients)	AE's (no. of events)	SAE's (n of events)
	-	OS 1mg/kg, (all)[#]	Leukopenia (4)[$]	Death (6) Hepatitis (3) Severe cytopenia (4)
	-	OS 30-80mg/day (all)	Transient leucopenia (2)	Heart failure (1)† and CVA (2)†
	unk	OS 10-240 mg/ day (all)	None	None
	4-36	OS	Herpes zoster (1) Mild urticaria (1) Diarrhoea & vomiting (1)	adenoCA(1)†, mammaCA (1)†, MI (1)†, CVA(1)†
	-	None	None	None
	0.3- 7	OS	Transient leucopenia (4) Nausea, vomiting, diarrhea (2)	None
	unk	unk	Gastro-intestinal intolerance (1)	None
	unk	unk	Liver enzymes ↑((1)	None
	10	None	Vomiting (2), Diarrhea (2), Abdominal discomfort (2)	CVA (1)† Airway disease(1)† heart disease(1)†
	18	OS 10-60 mg (all)	Gastrointestinal complaints (1)	Septic arthritis (1) Epidural abcess (1)
	unk	OS (3) Col (1)	Verrucae (3), Liver enzymes ↑((1) Herpes zoster (1)	None
	unk	OS (all)	Mild hepatitis (1)	None

Table 3. *Continued*

Disease	Reference (year)	Efficacy/effectiveness	Mean onset of action, wk
SCLE/CCLE	13. (1991)	1 (17%) complete response 3 (50%) partial response 2 (33%) no response	<6
EM	16. (1995)	3 (60%) complete remission 2 (40%) well-controlled disease	unk
HD	38. (2009)	2 (33%) >75% improvement 1 (17%) >50 % improvement 3 (50%) withdrew	unk
Lep T1R	31. (2004)	*Skin sympt:* AZA: 52% improvement (p<0.01) Pred: 65% improvement (p<0.01) *Rescue medication*: 48% subjects in AZA versus 37% in pred group. *Nerve function*: no effect in both groups. No sign. diff. between groups; ITT	1
LP	43. (2001)	7 (77.8%) excellent response 1 (11.1%) good response 1 (11.1%) poor response	4-6
PD	44. (2008)	AZA: CSS decreased from 64.5 (SD16.4) to 4.3 (SD 5.6) p <0.01 Beta: CSS decreased from 67.1 (SD 17.4) to 0.6 (SD 2.2) p <0.01; PP	4
PD	42. (2000)	38 (88%) complete remission. 3 (7%) incomplete remission. 2 (5%) withdrew No sign. diff. between groups	unk
PD	45. (2006)	CSS decreased from mean 40.4 (SD 8.0) to 10.9 (SD8.4) p<0.01	4-6
PD	40. (1998)	10(66%): 90% reduction in PASI 3 (20%): 50% reduction in PASI 1 (2%): <50% reduction in PASI 1 (2%): No improvement	8-28
PRP	24. (1972)	4 (80%) cleared completely 1 (20%) withdrew	1
PsA	27. (1990)	8 (80%) complete remission 1 (10%) incomplete remission 1 (10%) relapse	6.1 (SD 6.7)
Ps	15. (1974)	3 (10.3%) 50-75% improvement 16 (55%) 75-100% 7 (24%) no improvement 3 (10.3%) stopped due to SAE	unk
PsA	6. (1973)	10 (91%) complete remission 1 (9%) marked remission	0.5-4.5
Ps	46. (1972)	40 (60%) complete remission 20 (30%) partial reduction 6 (10%) no significant reduction	<10

Duration of remission, mo	Concomitant medication (no. of patients)	AE's (no. of events)	SAE's (n of events)
unk	OS 20-30mg (all)	Drug-induced fever (2), Nausea (1) Leucopenia (1)	Pancreatitis (1)
8	OS (1) Dap (1)	unk	unk
unk	AB, TS (all)	unk	unk
>3	OS 5-40mg (all)[#]	Mild nausea (2), Herpes zoster (1) Transient leucopenia (1)	None
>6-9	unk	Gingivitis (1) [$]	None
1-6	TS (all)	Infections (15), Loss of appetite (7), Dyspepsia (5), Fever (5), Nausea/vomiting (3), Weight gain (3), Backache (3), Hirsutism (1)	None
unk	OS (n=22)	Infection (8), Liver enzymes ↑((2), Fever (1), Hepatitis (1), Malaise (1), Nausea (1), Palpitations (1), Vomiting (1)	None
-	TS (all)	Vomiting (2), Weight loss 2-3 kg (2), Paresthesia of hands (1)	None
unk	OS (all)[#]	None	None
3- appr. 36	OS (1)	Stomach ache (2), Gastric hiatus hernia (1) Nausea (1), Solar keratosis (1), Skin cancers (1)	None
0-93 (Mean 18)	OS (5) SZ (2)	Herpes zoster infection (1), Transient leucopenia(1) Transient leucothrombopenia (1)	Pulmonary embolus(1)† Resp. insuf. (1)†
6-9	None	Nausea, diarrhea, abdominal pain (12), Leucopenia (10), Reversable mild portal fibrosis (8), Abnormalities of taste (2), Cholestasis (2) Thrombocytopenia (1)	None
1-10	unk	Nausea, vomiting, diarrhea (5), Leukopenia (2) Anemia (1), Liver enzymes ↑((1)	None
0-12	unk	Fever (Unk), Nausea (Unk)	None

Table 3. *Continued*

Disease	Reference (year)	Efficacy/effectiveness	Mean onset of action, wk
Ps	17. (1970)	2 (20%) 75-100% BSA reduction 1 (10%) 50-75% BSA reduction 2 (20%) 25-50% BSA reduction 5 (50%) 0-25% BSA reduction	2-6
Ps(A)	21. (1970)	4 (20%) complete remission 6 (30%)almost complete remission 9 (45%) incomplete remission 1 (5%) no response	1-8
SSc	36. (2006)	AZA: mRss 14.3 (SD 1.0) to 14.5 (SD 1.2) after 18 mo. $p>0.05$ CYC: mRss 14.7 (SD $\pm$ 1.1) to 5.2 (SD $\pm$ 0.5) after 18 mo. $p<0.01$ Difference between groups $p<0.01$; ITT	unk
SSc	37. (2007)	mRss 8.23 (SD 2.9) to 6.38 (SD 3.4) ($p<0.01$).	unk
ScD/ SSc	14. (1973)	2 (40%) sign. improvement 2 (40%) moderate improvement 1 (20%) no response	unk
ScD	25. (1968)	8 (40%) improved 7 (35%) no change 1 (5%) worsening 1 (5%) loss to FU 3 (15%) withdrew	unk
Vi	39. (2006)	Mean BSA repigmentation after 4 months: 58.4% in PUVA plus AZA versus 24.8% in PUVA group ($p<0.01$). ITT	3-5

For all abreviations, see list at end of article.
Beta; bethamethason, BSA; body surface area, CA; carcinoma, CI; confidence interval, Col; colchicine, CSS; Clinical Severity Score, CVA; cerebro vascular accident, CYC; cyclophosphamide, d; days, Dap; dapsone, Diff; difference, FU; follow-up, LRTI; lower respiratory tract infection, MI.; myocardinal infarction, MM; mycofenolaat mofetyl, OS; oral steroids, PASI; Psoriasis Area and severity index, PE; Plasma exchange, PP; per protocol analysis, Pred; prednisolone, Resp. insuf.; respiratory insuffiency, sign; significant, SZ; salazopyrine, TS; topical steroids, unk; unknown, URTI; upper respiratory tract infection, y; years.
† subject deceased
- not applicable
was tapered during treatment
*post-treatment
§ Incomplete reporting of adverse events

	Duration of remission, mo	Concomitant medication (no. of patients)	AE's (no. of events)	SAE's (n of events)
	-	None	Transient leucopenia (2), Transient aneamia (1) Liver enzymes ↑((1)	None
	1-8	unk	Transient leucopenia (14), Transient aneamia, (5) Transient thrombocytose (3), Gastro-intestinal complaints (3) Transient thrombocytopenia (1)	Aneamia (1) Myocard infarct (1)
	-	OS 15 mg/day (all)	Nausea (3), Dyspepsia (2), Leukocytopenia (2), Otitis media (1)	None
	-	OS 10 mg (all)	unk	unk
	-	OS (4) AB (4)	unk	unk
	-	-	Granulocytopenie (4), Anemie (3), Gastro-intestinal complaints (3), Serum sickness (3), Thrombocytopenie (2), Joint and muscle pain (1)	None
	-	-	Gastric complaints (2)	None

Table 4. Risk of bias of included RCT's

Disease	Adequate randomization?	Adequate concealment of allocation?	Adequate blinding?
AD	YES	YES	Participants YES Researchers YES Outcome assessment YES
AD	YES	YES	Participants YES Researchers YES Outcome assessment YES
BP	YES	UNCLEAR	Participants NO Researchers NO Outcome assessment NO
BP	YES	YES	Participants NO Researchers NO Outcome assessment NO
BP	YES	YES	Participants NO Researchers NO Outcome assessment NO
CAD	UNCLEAR	UNCLEAR	Participants UNCLEAR Researchers UNCLEAR Outcome assessment UNCLEAR
CV	UNCLEAR	UNCLEAR	Participants NO Researchers NO Outcome assessment NO
Lep T1R	NO	NO	Participants NO Researchers NO Outcome assessment NO
PD	YES	YES	Participants YES Researchers YES Outcome assessment UNCLEAR
SSc	NO	NO	Participants NO Researchers NO Outcome assessment NO
Vitiligo	YES	YES	Participants NO Researchers NO Outcome assessment NO

[a] Any form of bias other than the aforementioned (eg, analytical, attrition or ascertainment)

Incomplete data reported?	Free of selected reporting?	Free of other bias?[a]	Ref (year)
UNCLEAR	YES	YES	8. (2002)
YES	YES	YES	33. (2006)
YES	YES	YES	7. (2007)
UNCLEAR	UNCLEAR	NO	19. (1993)
UNCLEAR	UNCLEAR	UNCLEAR	11. (1978)
NO	UNCLEAR	UNCLEAR	34. (1989)
NO	UNCLEAR	UNCLEAR	20. (1991)
UNCLEAR	YES	NO	31. (2004)
NO	UNCLEAR	UNCLEAR	44. (2008)
NO	YES	UNCLEAR	36. (2006)
NO	NO	NO	39. (2006)

Table 5. Summary of the evidence and Quality of Evidence-based recommendations for clinical practice. [b] A consensus-based assignment of the working group based on the raw data per disease. Available categories for magnitude of treatment are: good, moderate, low or worsening.

Diseases	Grade of recommen-dation[a]	Quality of evidence[a]	Magnitude of therapeutic effect[b]
AD	1	A	Moderate
BP	2B	C	Worsening to good
BP (combination therapy with oral corticosteroids)	2A	B	Moderate (steroid sparing)
CAD	2A	B	Moderate
CV (combination therapy with oral corticosteroids)	2A/2B	B/C	Moderate
Cutaneous lupus erythematosus (combination therapy with oral corticosteroids)	2B	C	Low
EM	2B	C	Good
HD	2B	C	Moderate

Table 5. *Continued*

	Recommendation	Sources
	AZA can be used for the treatment of AD if registered treatment options fail or are contra-indicated (very certain estimate for a very certain moderate effect). Attention should be given to safety aspects when prescribing AZA (uncertain off-label safety)	Berth-Jones et al,[8] Meggitt et al,[33] Hon et al,[22] Meggitt et al,[32] Buckley et al,[9] Hudges et al,[23] Malthieu et al,[30] Murphy et al,[35] Kuanprasert et al,[26] Lear et al.[28]
	A weak recommendation is given for treating BP with AZA alone if treatment with oral corticostroids is contraindicated or has failed (very uncertain estimate for an effect of very uncertain magnitude).	Van Dijk and van Velde[41]
	Although there is no significant benefit of AZA with oral corticosteroids compared to mono therapy with oral steroids, this combination therapy may be considered when there is a need for a corticosteroid-sparing effect and a reduction of the cumulative steroid dose (uncertain estimate for a moderate steroid sparing effect). Attention should be given to safety aspects when prescribing AZA (uncertain off-label safety).	Beissert et al,[7] Guillaume et al,[19] Burton et al,[10] Burton et al,[11] Ahmed et al,[5] Greaves et al.[18]
	If other effective options have failed or are contra-indicated, there is a weak recommendation for the use of AZA (uncertain estimate for an uncertain moderate effect). Attention should be given to safety aspects when prescribing AZA (uncertain off-label safety).	Murphy et al,[34] Yap et al,[47] Leigh and Hawk[29]
	A weak recommendation is given for treating cutaneous vasculitis with AZA in combination with oral corticosteroids if conventional treatment options are contra-indicated or have failed. It is very uncertain if the efficacy outweighs the safety aspects (very uncertain estimate for a very uncertain moderate effect). Patients with cutaneous vasculitis caused by an identifiable agent that can be eliminated should not be initiated on this treatment. Elimination of the identifiable agent should be the first treatment step.	Heurkens et al,[20] Callen et al,[13] Callen and af Ekenstam.[12]
	A weak recommendation is given for treating cutaneous lupus erythematosus with AZA in combination with oral corticosteroids if conventional treatment options are contra-indicated or have failed. It is very uncertain if the beneficial effects outweigh the safety aspects (very uncertain estimate for a very uncertain low effect).	Callen et al.[13]
	A weak recommendation is given for treating recurrent erythema multiforme with AZA if conventional treatment options are contra-indicated or have failed (very uncertain estimate for a very uncertain good effect). Attention should be given to safety aspects when prescribing AZA (uncertain off-label safety).	Farthing et al.[16]
	There is a weak recommendation for the use of AZA to reduce swelling in patients with chronic hand dermatitis if conventional treatment options are contra-indicated or have failed (very uncertain estimate for a very uncertain moderate effect). Extra attention should be given to safety aspects when prescribing AZA (uncertain off-label safety).	Pearce and Mortimer.[38]

Diseases	Grade of recommen-dation[a]	Quality of evidence[a]	Magnitude of therapeutic effect[b]
Lep T1R (combination therapy with oral corticosteroids)	2A	B	Moderate (corticosteroid sparing)
LP	2B	C	Good
PD (combination therapy with oral or topical steroids)	2A	B	Good
PRP	2B	C	Good
Ps	2B	C	Moderate
ScD (combination therapy with oral steroids)	2B	B/C	Worsening to low
Vi (combination therapy with psoralen photochemotherapy)	2A	B	Moderate

[a]According to the *Archives of Dermatology* criteria for assessing the quality of the evidence to support recommendations.[4] Criteria for grade of recommendation: (1) strong recommendation: high-quality, patient-oriented evidence; (2A) weak recommendation: limited-quality, patient-oriented evidence; and (2B) weak recommendation: low-quality evidence. Criteria for assessing the quality of the evidence: (A) systematic review/meta-analysis, randomized controlled trials with consistent findings, or all-or-none observational study; (B) systematic review/meta-analysis of lower-quality clinical trials or studies with limitations and inconsistent findings, lower-quality clinical trial, cohort study, or case-control study; and (C) consensus guidelines, usual practice, expert opinion, or case series.

Recommendation	Sources
A weak recommendation is given for the use of AZA with concomitant oral corticosteroids for the treatment of a severe leprosy type 1 reaction. Although no significant beneficial effect of AZA was shown compared with conventional (corticosteroid) treatment, a possible steroid-sparing effect was suggested. Duration of treatment was considered too short by the working group to evaluate the treatment effect. Attention should be given to safety aspects when prescribing AZA (uncertain off-label safety).	Marlowe et al.[31]
A weak recommendation is given for treating lichen planus with AZA, if conventional treatment options are contra-indicated or have failed (very uncertain estimate for an uncertain good effect). Attention should be given to safety aspects when prescribing AZA (uncertain off-label safety).	Verma et al.[43]
A weak recommendation is given for treating PD with AZA (certain estimate for a certain good effect). Attention should be given to safety aspects when prescribing AZA (uncertain off-label safety).	Verma et al,[44] Verma et al,[42] Verma et al,[45] Sharma et al.[40]
A weak recommendation is given for treating pityriasis rubra pilaris with AZA if conventional treatment options are contra-indicated or have failed (very uncertain estimate for a uncertain good effect). Attention should be given to safety aspects when prescribing AZA (uncertain off-label safety).	Hunter et al.[24]
A weak recommendation is given for treating moderate to severe psoriasis, erythrodermic psoriasis and/or arthritis psoriatica with AZA (uncertain estimate for a very uncertain moderate effect). There are numerous other registered treatment options available with a stronger clinical recommendation. Therefore, it is only indicated when registered/conventional treatment fails or is contra-indicated. Attention should be given to safety aspects when prescribing AZA (uncertain off-label safety).	Le Quintrec et al,[27] Du Vivier et al,[15] Baum et al,[6] Weitgasser et al,[46] Greaves and Dawber,[17] Hewitt et al.[21]
A weak recommendation is given against the use of AZA in the treatment of scleroderma or cutaneous involvement of systemic sclerosis, because there seems to be no significant benificial effect of AZA (uncertain estimate for a very uncertain low effect or worsening).	Nadashkevich et al,[36] Paone et al,[37] Dethlefs and Tronnier,[14] Jansen et al.[25]
A weak recommendation is given against treating vitiligo with a combination of AZA and PUVA. Although there is low evidence for a very uncertain moderate repigmentation, there are additional risks associated with the combination of oral immunosuppressive therapy and photo therapy (non-melanoma skin cancer) that outweigh the potential benefits.	Radmanesh and Saedi[39]

RESULTS PER DISEASE

Atopic Dermatitis

All 10 studies concerning AZA in patients (n=319) with moderate to severe AD, whose conditions were often refractory to conventional treatment, showed an overall decrease in disease severity after active treatment with AZA.[8, 9, 22, 23, 26, 28, 30, 32, 33, 35] Two included RCTs showed that AZA treatment was significantly more favorable than placebo treatment.[8,33] A mean SASSAD improvement of 26% when receiving AZA treatment was found after 3 months of treatment with 2.5 mg/kg of AZA in the study by Berth-Jones *et al*.[8] Noticeably, 12 of the 37 patients withdrew during AZA treatment, 4 of whom did so because of AEs, compared with 4 of the 37 patients who withdrew during placebo treatment. Because an intention-to-treat (ITT) analysis was performed, the risk of overestimation of the effect is considered minimal. Meggitt *et al*[33] showed a mean SASSAD improvement of 37% in the AZA group vs 20% in the placebo group on ITT analysis ($P<0.01$). Six patients in the AZA group withdrew, compared with 1 in the placebo group. A meta-analysis of the RCTs could not be performed because of lack of data.

Eight case series, of which 3 were prospective, were included concerning 221 patients with AD. One study[35] was performed solely with children and showed a good to excellent response in 85% of the patients. In the study by Hon *et al,*[22] a mean Severity Scoring of Atopic Dermatitis improvement of 27% was seen during a 6-month period. The cohort study by Meggitt and Reynolds[32] found a mean SASSAD improvement of 26%, which was statistically significant. In all the other studies, descriptive means were used to report changes in disease severity. In those studies, at least 60% of the patients had a favorable response to AZA, although no clear definition was given. In 3 studies, complete remissions were reported in 40.5% to 58.3% of the patients. Onset of action of AZA ranged from 1 week to 7 months.

Four SAEs were described when using AZA to treat AD: severe pancytopenia (TPMT activity not measured), pancreatitis, non-Hodgkin lymphoma and a ruptured cerebral aneurysm. The last 2 SAEs occurred in the follow-up period.

Bullous Pemphigoid

Seven published studies[5,7,10,11,18,19,41] were found in the literature in which patients with BP were treated with AZA: 3 RCTs, 1 cohort study, and 3 case series. In total, 252 patients were enrolled, with a mean age of approximately 76 years. In all patients, the diagnosis of BP was established by histologic analysis of skin biopsy specimens. Concerning the RCTs, Beissert *et al*.[7] compared methylprednisolone

plus AZA with methylprednisolone plus mycophenolate mofetil. Results show that in both groups, 100% remission was achieved. Time to remission was more prompt in the AZA group, although this was not statistically significant.

Burton *et al.*[11] found a significant reduction (45%) of the cumulative prednisone dosage during a 3-year period when adding AZA. Although the percentage of remissions and number of deaths were lower in the prednisone plus AZA group, it was not significant. In the study by Guillaume *et al*,[19] prednisolone monotherapy was compared with prednisolone plus AZA and prednisolone plus plasma exchange. This trial was interrupted after the interim analysis showed no appreciable benefit resulting from the addition of AZA or from adding plasma exchange to prednisolone. A meta-analysis could not be performed because of clinical and methodologic heterogeneity.

In the cohort study by Ahmed *et al*,[5] 2 groups were compared: prednisone and prednisone plus AZA. A 50% reduction in the maintenance dosage of prednisolone was found in the AZA group and a 30% reduction was found in the duration of drug therapy. Regarding the case series, Greaves *et al*.[18] found that the dosage of prednisolone at which the patients relapsed was lower when patients were using concomitant AZA.

They found that by giving AZA, it was possible to withdraw prednisolone maintenance treatment in 9 of 11 patients without relapses. In the follow-up study by Burton and Greaves,[10] 44% of these patients remained in remission during a 4-year period. One study[41] used AZA as monotherapy and showed a good to very good response in 4 of the 5 patients.

Because of the mean age of the study participants, the relatively long follow-up period, and the severity of the disease, SAEs were not uncommon. Fourteen deaths occurred, of which 7 were unspecified, 3 were due to cerebral vascular accidents, 2 were due to heart diseases, 1 was due to an adenocarcinoma, and 1 was due to a preexistent mamma carcinoma.

Chronic Actinic Dermatitis

In total, 3 studies[29,34,47] were included in which evidence for the efficacy of AZA in patients with chronic actinic dermatitis was found. Murphy *et al.*[34] performed a double-blinded, placebo-controlled RCT. At treatment months 1, 3, and 6, the mean reduction in rash score and extent of rash and itch on a visual analogue scale was statistically different between the AZA and placebo group. The trial was prematurely terminated because the clinical status of actively treated patients showed marked improvement compared with placebo. The study had serious limitations concerning randomization, concealment of allocation and

blinding. Thereby, the 2 treatment groups were not comparable concerning severity of the disease at baseline and UV-exposure during the trial.

In the 2 case series, 14 patients with chronic actinic dermatitis were included. In total, 57% to 92% of the patients cleared or improved markedly. Three patients died, of a cerebral vascular accident, airway disease, or heart disease, respectively, after 12 to 15 months of AZA treatment.

Cutaneous vasculitis

One non-blinded RCT and 2 case series were eligible for this study.[12,13,20] Regarding the RCT, prednisolone with AZA (2.0 mg/kg/d) was compared with different conventional antirheumatic treatments (including nonsteroidal anti-inflammatory drugs, hydroxychloroquine, aurothioglucose, penicillamine, and sulfasala-zine) in 19 patients with cutaneous vasculitis characterized by palpable purpura, ulcers, nail fold infarcts, peripheral gangrene, and histologic features of small vessel infiltration.[20] In the AZA group, a significant reduction of skin vasculitis occurred after 18 months; 7 patients (88%) achieved complete remission. This finding was not statistically significant compared with the conventional treatment group. One patient in the AZA group withdrew because of gastrointestinal AEs.

Concerning the methodologic quality, details regarding randomization and concealment of allocation were not given and no blinding was performed. Furthermore, the causes of vasculitis were not specified, and it was unclear how the severity of vasculitis was measured and by whom.

Of the 19 patients included in the case series, 83% to 100% were responsive to AZA treatment and were able to lower or discontinue their dosage of prednisone.[12,13] One patient experienced pancreatitis during the trial. Two other SAEs of infectious origin occurred: septic arthritis and epidural abscess. One patient with severe and rapid-onset vasculitis died of renal failure during treatment with AZA, which could be attributed to the natural course of the underlying disease.

Leprosy Type 1 Reaction

One non-blinded RCT in which patients with leprosy type 1 reaction were treated with AZA (3.0 mg/kg/d) and concomitant prednisolone (40 mg/d to 5 mg/d) or prednisolone monotherapy was eligible.[31] In total, 40 patients were enrolled. At week 12, skin symptoms improved in 52% of the patients in the AZA plus prednisolone group vs 65% of the patients in the prednisolone-only group. At 12 and 24 weeks, the nerve function (sensory and voluntary muscle testing) showed no improvement in either groups. Nerve pain and tenderness improved in both groups. No statistically significant difference was observed between the

groups concerning all these outcomes. The RCT has serious limitations because the study was not blinded, it was inadequate in randomization and concealment of allocation and the outcome measure was not validated.

Parthenium Dermatitis

Four studies[40,42,44,45] concerning AZA treatment in PD (an Airborne Plant Allergen Contact Dermatitis) were found: 1 single-blinded, parallel-group RCT; 1 cohort comparing dosage regimens; and 2 case series. Concerning the RCT, 73% in the AZA group (100 mg/d) vs 72% in the betamethasone group (2 mg/d) had an excellent response to treatment.[44] During follow-up, 45% had a relapse in the AZA group compared with 67% in the betamethasone group ($P > 0.05$). It was unclear who measured the severity of PD and whether this was performed in a blinded fashion. Also, a substantial number of patients were lost to follow-up without specifications given.

In the cohort study, 3 dosage regimens were compared: 100 mg daily, 50 mg daily plus 300 mg every 28 days, and 100 mg/d plus 300 mg every 28 days.[42] Complete remission was achieved by 11 patients (50%) in group 1, as well as 2 patients (18%) in group 2 and 3 patients (30%) in group 3. These results were not statistically significant. Patients who were using AZA for less than 6 months were excluded from analysis. Furthermore, the groups were not clearly described and the outcome measures were not defined. Of the 32 patients with PD included in the case series, approximately 60% had a greater than 80% reduction in severity scores.

Psoriasis

In total, 6 eligible case series[6,15,17,21,27,46] were found in the literature, in which 147 patients with moderate to severe psoriasis, erythrodermic psoriasis, and/or arthritis psoriatica, whose conditions were often refractory to conventional therapy, were treated. Overall, the case studies gave similar results for effectiveness, with approximately 80% of the patients experiencing marked improvement. Three SAEs were described during treatment: a pulmonary embolus, severe anemia and myocardial infarction.

Scleroderma

Four studies[14,25,36,37] were eligible in which patients with scleroderma or cutaneous involvement of systemic sclerosis were treated: a non-blinded RCT and 3 case series. In the RCT, cyclophosphamide (2 mg/kg/d) combined with prednisolone was more effective than AZA (2.5 mg/kg/d) combined with prednisolone after 30 months.[36] The RCT had serious limitations in randomization (according to date

of birth), concealment of allocation and blinding. The case series demonstrated some beneficial effects of AZA in combination with prednisone.

Vitiligo

A non-blinded RCT was performed comparing oral psoralen-UVA (PUVA) photochemotherapy twice per week and AZA (0.6-0.75 mg/d) with PUVA monotherapy (twice/wk) in 92 patients with vitiligo.[39] Thirty-two patients were lost to follow-up during the study (reasons not specified). After 4 months, the mean total repigmentation rate was 58.4% in the PUVA plus AZA group vs 24.8% in the PUVA-only group ($P<0.001$).

Combining phototherapy with systemic immunomodulating drugs is not advised by the working group because of the risk of developing non-melanoma skin cancer.

Cutaneous Lupus Erythematosus, Erythema Multiforme, Hand Dermatitis, Lichen Planus, and Pityriasis Rubra Pilaris

For these diseases, the evidence was limited to 1 case series with fewer than 10 patients per disease.[13,16,24,38,43] Therefore, the quality of evidence is considered low (level C) for each disease and the estimate of effect uncertain. Efficacy and safety details are given in Table 3.

DISCUSSION

Main findings

A strong clinical recommendation was given for the use of AZA as an alternative treatment in AD. High quality of evidence was found for a moderate therapeutic response. Although the efficacy is not as pronounced as for cyclosporin (SASSAD improvement of 39% to 57%),[51] AZA can be considered an alternative treatment option for severe AD.

A level 2A recommendation was given for the use of AZA in BP in combination with oral corticosteroids. Moderate strength of evidence (level B) was found for a moderate therapeutic response of AZA with concomitant corticosteroids, although no significant benefit was found for combined treatment with AZA and oral corticosteroids compared with monotherapy with oral corticosteroids. AZA is considered to be a corticosteroid-sparing agent that can reduce corticosteroid-related adverse events. A level 2A recommendation was given for AZA in combination with topical or oral corticosteroids as an alternative treatment option for PD. Whether these results can be extrapolated

to all forms of airborne contact dermatitis remains uncertain. Also, a level 2A clinical recommendation was given for chronic actinic dermatitis, leprosy type 1 reactions (in combination with oral corticosteroids), and cutaneous vasculitis (in combination with oral corticosteroids). Evidence of a moderate therapeutic effect existed, although the strength of evidence was moderate (level B or B/C). AZA should only be considered if other treatments (registered or known alternatives) have failed. A moderate to good treatment response was found for erythema multiforme, hand dermatitis with lymphedema, lichen planus, pityriasis rubra pilaris, and psoriasis. Because this was mostly based on only 1 case series per disease, the strength of evidence was considered very low (level C), and a weak recommendation (level 2B) was given. Evidence generated from these studies should be used as a proof of concept and thus should be followed by well-designed further research with a higher level of evidence. Because no beneficial effect of AZA on scleroderma seemed to exist, a level 2B recommendation against the use in scleroderma (in combination with oral corticosteroids) was given. Although a moderate treatment response for AZA in combination with PUVA was found for vitiligo, a level 2A recommendation against its use in this context was given because of an unfavorable risk-benefit ratio.

Overall, the strength of evidence concerning safety issues was low to very low (level C). When comparing the AEs stated in the Summary of Product Characteristics with the reported (S)AEs, we conclude that the reported (S)AEs corresponded with the known adverse effects or that they might be explained by the age of the patients and/or the severity of their disease (eg, BP). Gastrointestinal concerns were common (n=120), as were increased liver enzyme levels (n=131) and mild infections (n=36). Severe infections were reported in 3 events. AZA-induced pancreatitis was observed twice. Malignant neoplasms occurred 3 times: an adenocarcinoma; a mamma carcinoma, which was preexistent; and a non-Hodgkin lymphoma 8 months after discontinuation of AZA therapy. Concerns exist that prolonged treatment with AZA predisposes patients to tumors, predominantly squamous cell carcinomas and non-Hodgkin lymphomas, because carcinogenicity is well recognized in organ transplant recipients treated with AZA and other immunosuppressive drugs.[52,53] Squamous cell carcinomas did not occur in our study, which could be explained by the relatively short duration of treatment and follow-up. Patients treated for short to medium periods, as in inflammatory bowel disease, appear not to be at increased risk for tumors.[54] The myelosuppressive effect of AZA was commonly reported: leukocytopenia (n=127), mainly lymphocytopenia (n=71), thrombocytopenia (n=9), anemia (n=9), and pancytopenia (n=2). Most of these cases were mild and transient. However, severe myelotoxic reactions were seen.

Because the risk for severe myelotoxic effects of AZA is increased in individuals with a TPMT deficiency, measuring TPMT activity before the initiation of AZA therapy is advised and might prevent lifethreatening myelosupression.[55,56]

The studies in this review were mostly of a short-term nature, with limited populations and often incomplete or missing reports of AEs, probably because of the emphasis on efficacy in the study design. Therefore, conclusions regarding (long-term) safety could not be drawn. Nonetheless, the reported (S)AEs did not reveal any particular new safety concern.

Methodologic quality

The main difficulty in assessing the off-label efficacy and safety of AZA was the generally poor quality of study design (inadequate randomization, concealment of allocation and blinding procedures) of RCTs and other studies in general, the heterogeneity of the data, the high drop-out rates and a lack in comprehensive reporting. A major source of heterogeneity was the use of different and un-validated diagnostic and outcome measures and the many studies in which only descriptive means were used. Therefore, the numerous case series were prone to selection and publication bias. The retrospective design raised concerns regarding recruitment bias and confounding.

Limitations

In this review, we excluded case series with fewer than 5 patients because we considered the additional value of including approximately 300 case series with fewer than 5 patients to be minimal and because doing so would lead to an unclear overview. To overcome the fact that potentially important safety issues could be missed, we selected a random sample of 20 case reports and screened those articles for relevant safety issues. Because we did not encounter any, we concluded it was appropriate to exclude those case reports. In 2 studies that compared their results against an undefined control group, the control groups were excluded from analysis and the studies were considered case series. In studies concerning systemic conditions (eg, systemic sclerosus or vasculitis), patients without dermatologic involvement were excluded.

Clinical research implications

This review clearly reveals the need to standardize the manner in which studies regarding off-label treatment are being conducted. It is vital to perform methodologically sound studies using standardized and validated diagnostic and outcome measures to reduce heterogeneity and thus to enable

meta-analysis.[57,58] As shown, the available evidence for off-label treatment is scarce in some indications, and promising case series are often not followed by well-designed RCTs. If we do not move forward and start generating a high level of evidence for (rare) diseases, for which we do not have (many) effective or safe treatments, improvement of treatment options will stagnate. Initiatives to generate high levels of evidence, such as the United Kingdom Dermatology Clinical Trials Network, are necessary. Therefore, long-term registries, such as the Psonet initiative for psoriasis, should be created to generate long-term safety data for off-label drug use within dermatology.[59]

IMPLICATIONS FOR CLINICAL PRACTICE

Results from this review can be used to update or formulate clinical practice guidelines for off-label AZA prescription and thus to ensure evidence-based health care decision making. Future studies should be performed to confirm the role of AZA in diseases that currently have a weak recommendation because of low quality of evidence but in which AZA has shown promising results.

ACKNOWLEDGEMENTS

We are indebted to ME Wessels of the Dutch Order of Medical Specialists. We thank CAFM Bruijnzeel, SAPM Lavrijsen, EP van Puijenbroek, L Wigersma, HB Thio, K.AC van Grootheest and EMGJ de Jong of the Off-Label Project Group of the Dutch Society of Dermatology and Venereology. We thank P van der Valk, WR Faber, AY Goedkoop A Horikx, RIF van der Waal, A Floor-Schreudering, W Goldschmidt and NEL Swart of the Off-Label Working Group of the Dutch Society of Dermatology and Venereology.

LITERATURE

1. Stafford RS. Regulating off-label drug use—rethinking the role of the FDA. *N Engl J Med*. 2008; **358**:1427-1429.
2. Higgins JPT, Altman DG, eds. Assessing risk of bias in included studies. Section 8.3. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.0.2 Website. http://www.cochrane-handbook.org/. Accessed February 17, 2011.
3. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting

observational studies. *Lancet*. 2007; **370**: 1453-1457.

4. Robinson JK, Dellavalle RP, Bigby M, Callen JP. Systematic reviews: grading recommendations and evidence quality. *Arch Dermatol*. 2008; **144**: 97-99.
5. Ahmed AR, Maize JC, Provost TT. Bullous pemphigoid: clinical and immunologic follow-up after successful therapy. *Arch Dermatol*. 1977; **113**: 1043-1046.
6. Baum J, Hurd E, Lewis D, Ferguson JL, Ziff M. Treatment of psoriatic arthritis with 6-mercaptopurine. *Arthritis Rheum*. 1973; **16**: 139-147.
7. Beissert S, Werfel T, Frieling U, et al. A comparison of oral methylprednisolone plus azathioprine or mycophenolate mofetil for the treatment of bullous pemphigoid. *Arch Dermatol*. 2007; **143**: 1536-1542.
8. Berth-Jones J, Takwale A, Tan E, et al. Azathioprine in severe adult atopic dermatitis: a double-blind, placebo-controlled, crossover trial. *Br J Dermatol*. 2002; **147**: 324-330.
9. Buckley DA, Baldwin P, Rogers S. The use of azathioprine in severe adult atopic eczema. *J Eur Acad Dermatol Venereol*. 1998; **11**: 137-140.
10. Burton JL, Greaves MW. Azathioprine for pemphigus and pemphigoid- a 4 year follow-up. *Br J Dermatol*. 1974; **91**: 103-109.
11. Burton JL, Harman RM, Peachey RD, Warin RP. A controlled trial of azathioprine in the treatment of pemphigoid [proceedings]. *Br J Dermatol*. 1978; **99** (suppl 16):14.
12. Callen JP, af Ekenstam E. Cutaneous leukocytoclastic vasculitis: clinical experience in 44 patients. *South Med J*. 1987; **80**: 848-851.
13. Callen JP, Spencer LV, Burruss JB, Holtman J. Azathioprine: an effective, corticosteroid- sparing therapy for patients with recalcitrant cutaneous lupus erythematosus or with recalcitrant cutaneous leukocytoclastic vasculitis. *Arch Dermatol*. 1991; **127**: 515-522.
14. Dethlefs B, Tronnier H. Experiences with antimetabolic therapy in dermatology. *Dermatol Monatsschr*. 1973; **159**: 34-44.
15. Du Vivier A, Munro DD, Verbov J. Treatment of psoriasis with azathioprine. *Br Med J*. 1974; **1**: 49-51.
16. Farthing PM, Maragou P, Coates M, Tatnall F, Leigh IM, Williams DM. Characteristics of the oral lesions in patients with cutaneous recurrent erythema multiforme. *J Oral Pathol Med*. 1995; **24**: 9-13.
17. Greaves MW, Dawber R. Azathioprine in psoriasis. *Br Med J*. 1970; **2**: 237- 238.
18. Greaves MW, Burton JL, Marks J, Dawber RPR. Azathioprine in treatment of bullous pemphigoid. *Br Med J*. 1971; **1**:144-145.
19. Guillaume J-C, Vaillant L, Bernard P, et al. Controlled trial of azathioprine and plasma exchange in addition to prednisolone in the treatment of bullous pemphigoid. *Arch Dermatol*. 1993; **129**: 49-53.
20. Heurkens AHM, Westedt ML, Breedveld FC. Prednisone plus azathioprine treatment in patients with rheumatoid arthritis complicated by vasculitis. *Arch Intern Med*. 1991; **151**: 2249-2254.
21. Hewitt J, Escande JP, Leibowitch M, Franceschini P. Trial therapy of psoriasis with azathioprine. *Bull Soc Fr Dermatol Syphiligr*. 1970; **77**: 392- 396.
22. Hon K-LE, Ching GKW, Leung T-F, Chow C-M, Lee KKC, Ng P-C. Efficacy and tolerability at 3 and 6months following use of azathioprine for recalcitrant AD in children and young adults. *J Dermatolog Treat*. 2009; **20**: 141-145.

23. Hughes RC, Collins P, Rogers S. Further experience of using azathioprine in the treatment of severe AD. *Clin Exp Dermatol*. 2008; **33**: 710- 711.

24. Hunter GA, Forbes IJ. Treatment of pityriasis rubra pilaris with azathioprine. *Br J Dermatol*. 1972; **87**: 42-45.

25. Jansen GT, Barraza DF, Ballard JL, Honeycutt WM, Dillaha CJ. Generalized scleroderma: treatment with an immunosuppressive agent. *Arch Dermatol*. 1968; **97**: 690-698.

26. Kuanprasert N, Herbert O, Barnetson RS. Clinical improvement and significant reduction of total serum IgE in patients suffering from severe atopic dermatitis treated with oral azathioprine. *Australas J Dermatol*. 2002; **43**: 125-127.

27. Le Quintrec JL, Menkès CJ, Amor B. Severe psoriatic rheumatism: treatment with azathioprine: report of 11 cases. *Rev Rhum Mal Osteoartic*. 1990; **57**: 815-819.

28. Lear JT, English JSC, Jones P, Smith AG. Retrospective review of the use of azathioprine in severe atopic dermatitis. *J Am Acad Dermatol*. 1996; **35**: 642-643.

29. Leigh IM, Hawk JL. Treatment of chronic actinic dermatitis with azathioprine. *Br J Dermatol*. 1984; **110**: 691-695.

30. Malthieu FG, Guillet G, Larregue M. Azatropin in severe atopic dermatitis: 24 cases. *Ann Dermatol Venereol*. 2005; **132**: 168-170.

31. Marlowe SNS, Hawksworth RA, Butlin CR, Nicholls PG, Lockwood DNJ. Clinical outcomes in a randomized controlled study comparing azathioprine and prednisolone versus prednisolone alone in the treatment of severe leprosy type 1 reactions in Nepal. *Trans R Soc Trop Med Hyg*. 2004; **98**: 602-609.

32. Meggitt SJ, Reynolds NJ. Azathioprine for atopic dermatitis. *Clin Exp Dermatol*. 2001; **26**: 369-375.

33. Meggitt SJ, Gray JC, Reynolds NJ. Azathioprine dosed by thiopurine methyltransferase activity for moderate-to-severe atopic eczema: a double-blind, randomised controlled trial. *Lancet*. 2006; **367**: 839-846.

34. Murphy GM, Maurice PD, Norris PG, Morris RW, Hawk JL. Azathioprine treatment in chronic actinic dermatitis: a double-blind controlled trial with monitoring of exposure to ultraviolet radiation. *Br J Dermatol*. 1989; **121**: 639- 646.

35. Murphy L-A, Atherton D. A retrospective evaluation of azathioprine in sever childhood atopic eczema, using thiopurine methyltransferase levels to exclude patients at high risk of myelosuppression. *Br J Dermatol*. 2002; **147**: 308-315.

36. Nadashkevich OD, Davis P, Fritzler M, Kovalenko W. A randomized unblended trial of cyclophosphamide versus azathioprine in the treatment of systemic sclerosis. *Clin Rheumatol*. 2006; **25**: 205-212.

37. Paone CC, Chiarolanza I, Cuomo G, et al. Twelve-month azathioprine as maintenance therapy in early diffuse systemic sclerosis patients treated for 1-year with low dose cyclophosphamide pulse therapy. *Clin Exp Rheumatol*. 2007; **25**: 613-616.

38. Pearce VJ, Mortimer PS. Hand dermatitis and lymphoedema. *Br J Dermatol*. 2009; **161**: 177-180.

39. Radmanesh MS, Saedi K. The efficacy of combined PUVA and low-dose azathioprine for early and enhanced repigmentation in vitiligo patients. *J Dermatolog Treat*. 2006; **17**: 151-153.

40. Sharma VK, Chakrabarti A, Mahajan V. Azathioprine in the treatment of

Parthenium dermatitis. *Int J Dermatol*. 1998; **37**: 299-302.

41. van Dijk TJ, van Velde JL. Treatment of pemphigus and pemphigoid with azathioprine. *Dermatologica*. 1973; **147**: 179-185.
42. Verma KK, Manchanda Y, Pasricha JS. Azathioprine as a corticosteroid sparing agent for the treatment of dermatitis caused by the weed Parthenium. *Acta Derm Venereol*. 2000; **80**: 31-32.
43. Verma KK, Mittal R, Manchanda Y. Azathioprine for the treatment of severe erosive oral and generalized lichen planus. *Acta Derm Venereol*. 2001; **81**: 378-379.
44. Verma KK, Mahesh R, Srivastava P, Ramam M, Mukhopadhyaya AK. Azathioprine versus betamethasone for the treatment of parthenium dermatitis: a randomized controlled study. *Indian J Dermatol Venereol Leprol*. 2008; **74**: 453-457.
45. Verma KK, Bansal A, Sethuraman G. Parthenium dermatitis treated with azathioprine weekly pulse doses. *Indian J Dermatol Venereol Leprol*. 2006; **72**:24-27.
46. Weitgasser H. Experiences with an ambulatory immunosuppressive therapy of generalized psoriasis. *Hautarzt*. 1972; **23**: 316-318.
47. Yap LM, Foley P, Crouch R, Baker C. Chronic actinic dermatitis: a retrospective analysis of 44 cases referred to an Australian photobiology clinic. *Australas J Dermatol*. 2003; **44**: 256-262.
48. Schmitt J, Langan S, Williams HC; European Dermato-Epidemiology Network. What are the best outcome measurements for atopic eczema? a systematic review. *J Allergy Clin Immunol*. 2007; **120**: 1389-1398.
49. Spuls PI, Lecluse LLA, Poulsen M-LNF, Bos JD, Stern RS, Nijsten T. How good are clinical severity and outcome measures for psoriasis? quantitative evaluation in a systematic review. *J Invest Dermatol*. 2010; **130**: 933-943.
50. Merkel PA, Clements PJ, Reveille JD, Suarez-Almazor ME, Valentini G, Furst DE; OMERACT 6. Current status of outcome measure development for clinical trials in systemic sclerosis: report from OMERACT 6. *J Rheumatol*. 2003; **30**:1630- 1647.
51. Schmitt J, Schäkel K, Schmitt N, Meurer M. Systemic treatment of severe atopic eczema: a systematic review. *Acta Derm Venereol*. 2007; **87**: 100-111.
52. Kinlen LJ. Incidence of cancer in rheumatoid arthritis and other disorders after immunosuppressive treatment. *Am J Med*. 1985; **78**: 44-49.
53. Taylor AE, Shuster S. Skin cancer after renal transplantation: the causal role of azathioprine. *Acta Derm Venereol*. 1992; **72**: 115-119.
54. Connell WR, Kamm MA, Ritchie JK, Lennard-Jones JE, Dickson M, Balkwill AM. Long-term neoplasia risk after azathioprine treatment in inflammatory bowel disease. *Lancet*. 1994; **343**: 1249-1252.
55. Sanderson J, Ansari A, Marinaki T, Duley J. Thiopurine methyltransferase: should it be measured before commencing thiopurine drug therapy? *Ann Clin Biochem*. 2004; **41**: 294-302.
56. Winter JW, Gaffney D, Shapiro D, et al. Assessment of thiopurine methyltransferase enzyme activity is superior to genotype in predicting myelosuppression following azathioprine therapy in patients with inflammatory bowel disease. *Aliment Pharmacol Ther*. 2007; **25**: 1069-1077.
57. Nijsten T, Spuls PI, Naldi L, Stern RS. The misperception that clinical trial data reflect long-term drug safety: lessons learned from Efalizumab's withdrawal. *Arch Dermatol*. 2009; **145**: 1037-1039.

58. Schram ME, Spuls PI, Bos JD. Off-label use of efalizumab in dermatology. *Exp Rev Dermatol*. 2010; **5**: 535-547.

59. Lecluse LLA, Naldi L, Stern RS, Spuls PI. National registries of systemic treatment for psoriasis and the European 'Psonet' initiative. *Dermatology*. 2009; **218**: 347-356.

List of Abbreviations

AB	Antibiotics
AD	Atopic dermatitis
AE	Adverse event
AZA	Azathioprine
BP	Bullous pemphigoid
CAD	Chronic actinic dermatitis
CCLE	Chronic cutaneous lupus erythematosus
CLV	Cutaneous leukocytoclastic vasculitis
CV	Cutaneous vasculitis
EM	Erythema multiforme
HD	Hand dermatitis
ITT	Intention-to-treat
LP	Lichen planus
Lep T1R	Leprosy type 1 reactions
LV	Leukocytoclastic vasculitis,
PCB	Placebo
PD	Parthenium dermatitis
PRP	Pityriasis rubra pilaris
Ps	Psoriais
PsA	Arthritis psoriatica
RCT	Randomized controlled trial
SAE	Serious adverse event
SASSAD	Six Area, Six Sign Atopic Dermatitis
ScD	Scleroderma
SCLE	Subacute cutaneous lupus erythematosis
SD	Standard deviation
SSc	Systemic sclerosis with cutaneous involvement
TPMT	Thiopurine methyltransferase
UV	Urticarial vasculitis
Vi	Vitiligo

6 SUMMARY AND CONCLUSIONS/ SAMENVATTING EN CONCLUSIES

SUMMARY AND CONCLUSIONS

The general introduction in ***Chapter 1*** provides an overview on the many facets of atopic dermatitis/eczema (AD). It highlights what is known about the pathogenesis, clinical features and patterns, nomenclature, epidemiology and treatment of AD and its challenges.

Chapter 2 adds to our current knowledge in the understanding of the pathogenesis of AD by summarizing the available evidence on prevalences of AD in urban versus rural area's and answering the question whether AD is more prevalent in urban over rural societies.

Numerous factors have been hypothesised to contribute to the fact that the overall prevalence of AD has been increasing in western societies. A plausible explanation for that is the role of environmental factors as part of the 'hygiene hypothesis'. This view postulates that hygienic environments in modern/urban society results in insufficient microbial stimulation in the newborns' immune system. This leads to lack of signals diverting type 2 responses to regulatory and type 1 T cell responses and thereby induces eczema.

An extensive literature search was performed in which all primary studies comparing the prevalence rate of eczema between urban and rural populations were included. Twenty-six suitable articles were found. Nineteen showed a higher risk of having eczema in an urbanised area, of which 11 were significant. Six studies showed a lower risk of having eczema in an urbanised area of which 1 was statistically significant. One study had a relative risk of 1.00. Results showed to be more homogeneous among studies of good methodological quality. Due to heterogeneity, it was not possible to pool the results.

We concluded that there was some evidence for a higher risk of having eczema in urban over rural areas. This might suggest that place of residency has a role in the pathogenesis of eczema. However, which elements in place of residency are responsible for this effect and at which time point in the development of an individual (antenatal versus postnatal) should be subject of further research.

Chapter 3 is centred on the diagnostic criteria for AD. Generally it is assumed that every dermatologist can recognise AD at first sight. However, diagnostic criteria are needed to provide standardized ways to diagnose AD and thereby ensure homogenous study populations in clinical research (hospital population) and accurate estimates of prevalences of AD in population based surveys.

Firstly, we summarized the available evidence on the validity of the ten available diagnostic criteria. For that purpose, we systematically reviewed the literature for eligible validity studies. Twenty articles, including 27 validation

studies, met our inclusion criteria. All studies were assessed for methodological quality by the Quality Assessment of Diagnostic Accuracy (QUADAS) tool.

The Hanifin and Rajka diagnostic criteria, which are the most extensively used criteria in clinical trials, were only validated twice: specificity ranged from 87.9% to 96.0% and sensitivity from 77.6% to 93.8%. In contrast with that, 19 validation studies of the U.K. diagnostic criteria showed sensitivity and specificity ranging from 10% to 100% and 89.3% to 99.1%. Other diagnostic criteria were only validated once or twice with variable outcomes. Methodological quality of the studies varied considerably. Of the 10 proposed diagnostic criteria, it was remarkable that 4 were not validated at all.

With this systematic review, we found a startling discrepancy between most commonly used and most commonly validated diagnostic criteria. Furthermore it can be concluded, that although the U.K. diagnostic criteria lacked sensitivity in some studies, they were the most validated criteria and seem suitable for estimation of prevalences of AD in large populations.

The second study was aimed to validate and further refine the millennium criteria. The millennium criteria, proposed by professor J.D. Bos, are a unique set of criteria that are not only able to differentiate between AD and other entities, but is also between AD and atopiform dermatitis by the use of the criterium 'presence of allergen specific IgE'. Unlike the U.K. criteria, the millennium criteria are aimed for hospital-based populations and use in clinical trials.

In order to validate and refine the millennium criteria, all the individual criteria of the millennium criteria, U.K. criteria and the Hanifin and Rajka criteria were assessed in 210 consecutive patients who were suspected for having AD or atopiform dermatitis. After logistic regression of the data of the individual criteria of the diagnostic lists, a set of 5 criteria were identified as best discriminators: typical morphology, early age of onset, Dennie-Morgan fold, historical and actual flexural involvement. The refined millennium criteria were constituted from these criteria. When comparing the different diagnostic criteria against clinical diagnosis for relative value, the refined millennium criteria showed a relative value of 0.81, the U.K. criteria 0.71 and the Hanifin & Rajka criteria 0.51.

We have shown that the refined millennium criteria are a valid tool to diagnose AD, which are validated in and aimed for hospital-based populations. Furthermore, the millennium criteria are able to distinguish AD from atopiform dermatitis, thereby enabling subgroup analysis.

With uniformity in the use of well-validated and applicable diagnostic criteria for clinical research and population-based studies, future studies will be comparable and a reduction of bias is established.

Chapter 4 presents the results of a study that was conducted in order to improve clinical outcome measurement of trials and other clinical studies.

The study was performed to further explore the clinimetric qualities, responsiveness and minimally clinically important difference (MCID), of four frequently used, valid and reliable clinical outcome measures: (objective) Severity scoring of atopic dermatitis (SCORAD), Eczema area and severity index (EASI) and Patient-oriented eczema measure (POEM). Responsiveness indicates the property of an outcome measurement to capture real changes in disease activity. The MCID is the smallest change in an outcome measurement that indicates a real clinically relevant change in disease severity.

The clinical data of three recently performed investigator-initiated randomized controlled trials were combined to calculate each property. Responsiveness was calculated by using area under the curve (AUC) of the receiver operating characteristic (ROC). MCID was established by analysing the absolute changes in the outcome measurements observed within individuals during treatment using the global assessment as anchors for clinical change.

The SCORAD showed an AUC of 0.70 (95% confidence interval (CI) 0.61-0.78).The AUC of objective SCORAD was 0.73 (95% CI 0.70-0.77). The AUC of the EASI is 0.67 (95% CI 0.60-0.76) and the AUC of the POEM is 0.67 (95% CI 0.59-0.75). Although the AUCs of the EASI and POEM seem to be somewhat lower than the AUCs of the SCORADs, the confidence intervals largely overlap. Scores above 0.70 represent a fair responsiveness and therefore we can confidently use these outcome measures in clinical trials.

The MCID was 8.7 point for the SCORAD, 8.2 for the objective SCORAD, 6.6 for the EASI and 3.4 for the POEM. The quantification of the MCID will offer advantages to analyse and design clinical trials. Based on the MCID, accurate power calculations can be made and it can be decided if statistically significant differences reflect a real change in disease severity to deserve consideration in clinical practise.

Chapter 5 aims at novel, off-label therapy strategies for AD as well as for some other difficult to treat and recalcitrant diseases.

We performed three studies in this context. The first is a single-blinded investigator-initiated randomized controlled trial in which adult patients with severe AD were randomized (ratio 1:1) to receive methotrexate or azathioprine. Of the 45 patients that were screened, 42 patients were included. At week 12, patients in the methotrexate group changed from 57.2 (SD 11.8) to 34.4 (SD 13.0) on mean SCORAD, representing a relative reduction of 42% ($p<0.001$). Patients randomized to the azathioprine group changed from 58.4 (SD 10.4) to 36.3 (SD 16.9) on mean SCORAD, representing a relative reduction

of 39% ($p<0.001$). P-value for the absolute difference between the groups is 0.89. Proportion of patients achieving at least mild disease, reduction on impact of quality of life, symptoms and levels of thymus and activation-regulated chemokine were similar in both groups at week 12 and 24. No statistically significant differences were found in number and severity of adverse events. Abnormalities in blood count were more common in the azathioprine group. No serious adverse events occurred.

With this study we have shown that both methotrexate and azathioprine are generally well-tolerated and effective treatments options for patients with severe AD. Hereby we have added new evidence-based therapies to the treatment armentarium.

In the second study we have shown the available evidence on the off-label use of efalizumab by performing a systematic review and we illustrated the time path in which drugs are offered off-label to dermatological patients. We found some evidence supporting effectiveness in palmoplantar pustulosis, AD, lichen planus and cutaneous lupus erythematosus. In total twelve serious adverse events occurred, one due to an infection. Efalizumab was given off-label within half a year after registration of the drug for plaque psoriasis. In reaction to the marketing suspension recommended by the EMEA (European Medicines Agency), Genentech announced the voluntary withdrawal of efalizumab. After June 8, 2009, efalizumab was no longer available. Despite the fact that the use of efalizumab is ceased, results of studies with efalizumab are still relevant. With this review, we like to create awareness among health care professional about the rapid introduction of efalizumab in off-label setting and the safety issues concerned. The withdrawal of efalizumab makes clear that the use of new and potentially harmful drugs in an off-label setting is not without risk. With admission of drugs in unregistered indications, there is an unknown balance between dose, efficacy and safety pattern.

In the third study we summarized the evidence regarding effectiveness, efficacy and safety of the off-label use of azathioprine in dermatology by performing a systematic review. An extensive search through multiple databases retrieved 3870 articles of which 148 articles were selected for detailed review. Forty-three studies matched the predefined in- and exclusion criteria. Per indication, safety and efficacy data were given and overall quality assessment was performed by using the GRADE methodology. Taken into account all these aspects, clinical recommendations for off-label use were given per indication. We gave strong clinical recommendations for use of azathioprine in AD and parthenium dermatitis (a form of contact dermatitis). Weak clinical recommendations were given for the use of azathioprine in bullous pemphigoid,

chronic actinic dermatitis, cutaneous vasculitis, cutaneous lupus erythematosus, erythema multiforme, hand dermatitis, leprosy type 1 reactions, lichen planus, pityriasis rubra pilaris and psoriasis. Weak recommendations against the use of azathioprine in vitiligo and scleroderma were given. Conclusions concerning safety in off-label setting could not be drawn due to scarce and incomplete data. Long-term registries and prospective studies could add to the existing evidence and support us legally for off-label drug use in dermatology.

SAMENVATTING EN CONCLUSIES

De algemene introductie in ***Hoofdstuk 1*** biedt een overzicht van de verschillende aspecten van atopisch eczeem (AE). Het belicht de pathogenese, klinische kenmerken en patronen, nomenclatuur, epidemiologie en de behandeling van AE met bijbehorende uitdagingen.

Hoofdstuk 2 draagt bij aan de huidige kennis over de pathogenese van AE door een analyse te geven van de beschikbare data over de prevalentie van AE in stedelijke versus landelijke gebieden. Het beantwoordt de vraag in welke van de twee gebieden AE vaker voorkomt.

In de loop der tijd zijn meerdere factoren gesuggereerd die mogelijk bijdragen aan de toenemende prevalentie van AE in de Westerse wereld. Een plausibele verklaring voor deze toename is de rol van omgevingsfactoren: 'de hygiëne hypothese'. Deze opvatting stelt dat leven in een hygiënische omgeving vroeg in de ontwikkeling resulteert in onvoldoende microbiële stimulatie van het immuunsysteem. Dit leidt tot een gebrek aan signalen voor een Th1-gemedieerde respons, waardoor de Th2-gemedieerde respons de overhand neemt en vermoedelijk eczeem veroorzaakt wordt. Deze hypothese zou door een verschil in prevalentie van AE in stedelijke versus landelijke gebieden worden ondersteund.

Door middel van een uitgebreid literatuuronderzoek zijn alle gevonden primaire studies met betrekking tot de prevalentie van eczeem tussen stedelijke en landelijke gebieden vergeleken. Er werden 26 geschikte artikelen gevonden. Negentien studies tonen een hoger risico op het hebben van eczeem in stedelijk gebied waarvan in elf studies dit verschil statistisch significant is. Zes studies tonen een lager risico op het hebben van eczeem in stedelijk gebied, waarvan slechts één statistisch significant is. Eén studie heeft een relatief risico van 1,00. Een meer homogeen beeld voor een hoger risico op eczeem in stedelijk gebied is te zien wanneer alleen de studies van goede methodologisch kwaliteit worden vergelijken. Als gevolg van heterogeniteit is het niet mogelijk om een meta-analyse van de resultaten uit te voeren.

Concluderend is er enig bewijs voor een hoger risico op het hebben van eczeem in stedelijke gebieden ten opzichte van landelijke gebieden. Dit kan erop wijzen dat de locatie van de leefomgeving een rol in de pathogenese van eczeem heeft. Welke factoren echter verantwoordelijk zijn voor dit effect en op welk tijdstip in de ontwikkeling (prenataal versus postnataal) dit van belang is zou onderwerp van nader onderzoek moeten zijn.

Hoofdstuk 3 is gericht op de diagnostische criteria voor AE. Over het algemeen wordt ervan uitgegaan dat elke dermatoloog AE direct herkent. Er zijn echter diagnostische criteria nodig om te dienen als gestandaardiseerd

diagnosticum. Zij zorgen voor homogene en vergelijkbare populaties in klinisch onderzoek (ziekenhuis populaties) waardoor onder andere bevolkingsbrede schattingen van de prevalentie van AE gemaakt kunnen worden.

Hoofdstuk 3.1 is een samenvatting van alle studies die over de validiteit van de tien bestaande diagnostische criteria gaan. Hiertoe is een systematisch zoekstrategie toegepast om alle mogelijk relevante studies te kunnen identificeren. Twintig artikelen, waaronder 27 validatie studies, voldeden aan de inclusiecriteria. Alle studies zijn beoordeeld op methodologische kwaliteit aan de hand van de Quality Assessment of Diagnostic Accuracy (QUADAS) methode.

De Hanifin en Rajka diagnostische criteria, de meest gebruikte criteria in klinische studies, worden slechts tweemaal gevalideerd: de specificiteit varieert van 87,9% tot 96,0% en de sensibiliteit van 77,6% tot 93,8%. Negentien validatie studies van de UK diagnostische criteria tonen een sensibiliteit en specificiteit, variërend van 10% tot 100% en 89,3% tot 99,1%. Andere diagnostische criteria worden slechts een- of tweemaal gevalideerd met variabele resultaten. De methodologische kwaliteit van de studies varieert daarbij aanzienlijk. Opmerkelijk is dat van de 10 bestaande diagnostische criteria, er vier nog nooit gevalideerd zijn. Op basis van deze systematische literatuurstudie is een noemenswaardige discrepantie zichtbaar tussen de meest gebruikte en de meest gevalideerde diagnostische criteria. Verder kan worden geconcludeerd dat, hoewel het de criteria in sommige studies ontbreekt aan sensibiliteit, de UK diagnostische criteria de meest gevalideerde criteria zijn en geschikt lijken voor het maken van prevalentieschattingen in grote populaties.

De tweede studie van dit hoofdstuk heeft als doel het valideren en verder verfijnen van de Millennium criteria (MC). De MC zijn ontworpen door professor J.D. Bos en bestaan uit een unieke set van criteria die niet alleen kunnen differentiëren tussen AE en andere huidziekten, maar ook tussen AE en atopiform eczeem. Dit laatste wordt mogelijk door het gebruik van het criterium 'aanwezigheid van allergeen specifiek IgE'. In tegenstelling tot de UK diagnostische criteria, zijn de MC gericht op gebruik in ziekenhuispopulaties en dus klinische trials.

Ter validatie en verbetering van de MC werden de afzonderlijke sensitiviteit en specificiteit van alle losse criteria van de MC, de UK criteria en de Hanifin en Rajka criteria getest op 210 achtereenvolgende patiënten die zich presenteerden op onze polikliniek huidziekten en waarbij in de differentiaal diagnose gedacht werd aan atopisch of atopiform eczeem.

Na logistische regressie van de afzonderlijke criteria werd een set van vijf criteria geïdentificeerd die het hoogst onderscheidend vermogen laat zien, bestaande uit: typische morfologie, jonge leeftijd van ontstaan van de klachten, Dennie-Morgan

oogplooien, actuele aanwezigheid of verleden van eczemateuze afwijkingen in de plooien. De vernieuwde MC werden uit deze criteria opgebouwd.

Vervolgens werden de vernieuwde MC vergeleken voor diagnostische accuratesse met de UK criteria en de Hanifin en Rajka criteria. In deze studie toonde de vernieuwde MC een relatieve waarde van 0.81, de UK criteria van 0.71 en de Hanifin & Rajka criteria van 0.51.

Deze studie toont aan dat de vernieuwde MC een valide diagnosticum zijn voor het diagnosticeren van AE en goed toegepast kunnen worden in ziekenhuispopulaties en/of klinische studies. Tevens zijn de MC in staat om een onderscheid te maken tussen AE en atopiform eczeem. Dit faciliteert het uitvoeren van subgroepanalyses.

Indien er uniformiteit is in het gebruik van gevalideerde diagnostische criteria, zullen toekomstige studies beter met elkaar vergeleken kunnen worden en wordt heterogeniteit in studiepopulaties verminderd.

Hoofdstuk 4 beschrijft de resultaten van een studie die werd uitgevoerd om de kwaliteit van uitkomst metingen in de kliniek en in studies te verbeteren.

In deze studie worden de responsiviteit en de 'minimal clinically important difference' (MCID), van de vier meest gebruikte en volgens een recente review meest betrouwbare klinische uitkomstmaten gemeten: de Severity Scoring of Atopic Dermatitis (SCORAD), de objectieve SCORAD, de Eczema Area and Severity Index (EASI) en de Patient-Oriented Eczema Measurement (POEM).

De responsiviteit van een meetinstrument geeft aan of essentiële veranderingen in de activiteit van de ziekte ook als zodanig worden gemeten met het meetinstrument. De MCID is de kleinste verandering van de uitkomst parameter die een echte klinisch relevante verandering in de ernst van de ziekte aangeeft.

De klinische data van drie recent uitgevoerde RCT's werden gecombineerd om deze eigenschappen te bepalen. De responsiviteit werd berekend met behulp van de oppervlakte onder de curve (AUC) van de 'receiver operating characteristic' (ROC) curve. De MCID werd vastgesteld door de absolute veranderingen in het meetinstrument te analyseren wanneer een individuele patiënt een relevante verandering in ziekte activiteit doormaakte. Een verandering van de globale ziektestatus (investigator or patient global assessment) werd als referentiewaarde voor relevante verandering gezien.

De SCORAD toont een AUC van 0,70 (95% betrouwbaarheidsinterval (BI) 0.61-0.78). De AUC van objectieve SCORAD is 0.73 (95% BI 0.70 - 0.77). De AUC van de EASI is 0.67 (95% BI 0.60 - 0.76) en de AUC van de POEM is 0.67 (95% BI 0.59 - 0.75). Hoewel de AUC's van de EASI en POEM lager lijken dan de AUC van beide SCORADs, overlappen de betrouwbaarheidsintervallen

grotendeels. Scores boven de 0.70 vertegenwoordigen voldoende responsiviteit. De MCID bedraagt 8.7 punt voor de SCORAD, 8.2 voor de objectieve SCORAD, 6.6 voor de EASI en 3.4 voor de POEM. De kwantificering van de MCID zal voordelen bieden voor het analyseren en ontwerpen van klinische studies. Op basis van de MCID kunnen nauwkeurige power berekeningen gemaakt worden voor klinische trials en kan men vaststellen of een statistisch significante verandering in een uitkomstparameter ook een reële verandering in de ziektestatus van de patiënt betekent.

Hoofdstuk 5 is gericht op nieuwe, off-label behandelstrategieën voor AE, evenals voor enkele andere moeilijk te behandelen en hardnekkige ziekten. In deze context zijn drie studies uitgevoerd.

De eerste studie is een enkel geblindeerde gerandomiseerde gecontroleerde trial waarin volwassen patiënten met ernstig AE werden gerandomiseerd (verhouding 1:1) voor methotrexaat of azathioprine behandeling. Van de 45 patiënten die werden gescreend voor dit onderzoek, zijn 42 patiënten daadwerkelijk gestart met de studie. In week 12 bleken de patiënten in de methotrexaat groep verbeterd van 57.2 (standaard deviatie (SD) 11.8) naar 34.4 (SD 13.0) punten op de gemiddelde SCORAD score. Dit betekent een relatieve verbetering van 42% ($p < 0.001$). Patiënten die gerandomiseerd werden naar de azathioprine groep veranderde van 58.4 (SD 10.4) naar 36.3 (SD 16.9) punten op de gemiddelde SCORAD score, wat vertaald kan worden als een relatieve reductie van 39% ($p < 0.001$). De P-waarde voor het absolute verschil tussen de groepen is 0.89. Er was ook geen verschil tussen de groepen in week 12 en 24 voor wat betreft het deel van de patiënten dat tenminste een milde ziektestatus op de globale beoordeling behaalden, de verbetering in de kwaliteit van leven en symptomen als jeuk en slaapverlies. Er zijn ook geen statistisch significante verschillen gevonden in het aantal en de ernst van de bijwerkingen. Afwijkingen in het bloedbeeld kwamen vaker voor in de azathioprine groep. Ernstige bijwerkingen zijn niet opgetreden.

Met deze studie tonen we aan dat zowel methotrexaat als azathioprine op de korte termijn goed verdragen worden en effectieve behandelopties kunnen zijn voor patiënten met ernstig AE. Met deze studie hebben we nieuw wetenschappelijk bewijs voor het off-label gebruik van deze therapieën en kunnen we ze toevoegen aan het behandelarsenaal van AE.

De tweede studie laat de beschikbare gegevens over off-label gebruik van efalizumab zien door middel van een systematische literatuurstudie. Tevens wordt de tijdslijn geïllustreerd waarin nieuwe medicijnen buiten de registratie aangeboden worden aan dermatologische patiënten.

We vinden enig bewijs voor effectiviteit van efalizumab in AE, palmoplantaire pustulose, lichen planus en cutane lupus erythematosus. In totaal werden 12 ernstige bijwerkingen beschreven in de geincludeerde studies, waaronder één infectie. Meldingen van off-label efalizumab werden binnen een half jaar na de registratie voor psoriasis vulgaris al gepubliceerd. In verband met het voorkomen van enkele zeer ernstige bijwerkingen, werd efalizumab vanaf 8 juni 2009 van de markt gehaald door de producent. Met het toepassen van geneesmiddelen in niet-geregistreerde indicaties, is er een onbekend evenwicht tussen de dosis, werkzaamheid en veiligheid.

De derde studie geeft een systematisch overzicht van alle data met betrekking tot de effectiviteit en veiligheid van azathioprine bij dermatologische patiënten met niet geregistreerde indicaties.

Middels een uitgebreide zoekopdracht werden 3870 potentieel relevante artikelen gevonden. Na selectie op de titel en de abstract zijn daarvan 148 artikelen geselecteerd voor een gedetailleerde review. Drieënveertig studies komen overeen met de vooraf gedefinieerde in- en exclusiecriteria.

De Grading of Recommendations Assessment, Development and Evaluation GRADE methodiek werd toegepast voor het omzetten van conclusies over effectiviteit en veiligheid naar aanbevelingen in de dagelijkse praktijk. Er werd, rekening houdend met de kwaliteit en kwantiteit van de data en de grootte van het therapeutische effect, een aanbeveling geformuleerd voor potentieel off-label gebruik van azathioprine bij de desbetreffende indicatie. De werkgroepleden van de Richtlijn off-label geneesmiddelengebruik in de dermatologie zijn hierbij als experts betrokken.

We geven een sterke klinische aanbevelingen voor het gebruik van azathioprine bij AE en parthenium dermatitis (een vorm van contact dermatitis). Zwakke klinische aanbevelingen worden gegeven voor het gebruik van azathioprine bij bulleus pemfigoïd, chronische actinische dermatitis, cutane vasculitis, cutane lupus erythematosus, erythema multiforme, handeczeem, lepra type 1 reacties, lichen planus, pityriasis rubra pilaris en psoriasis. Zwakke aanbevelingen tegen het gebruik van azathioprine worden gegeven bij gebruik bij vitiligo en sclerodermie. Conclusies over de veiligheid van azathioprine in off-label setting, kunnen we niet trekken als gevolg van schaarste aan data en onvolledigheid van gegevens. Lange termijn registers en prospectieve studies kunnen een oplossing betekenen voor dit probleem. Dit zal zorgen voor een toename van bewijskracht en zal het toepassen van off-label medicatie juridisch ondersteunen.

7 GENERAL DISCUSSION

This thesis aims at unraveling some epidemiologic and treatment facets of atopic dermatitis (AD). To this end, we performed studies that focused on prevalence, diagnostic criteria, clinical outcome measures and off-label therapy of AD. This chapter will discuss the main findings of this thesis, puts it in the context of the current knowledge and presents recommendations for future research.

In the first part of this thesis (chapter 2), evidence concerning the urban/rural gradient in the prevalence of AD is summarized. As genetics alone cannot explain the rapid increase of the prevalence of AD in the last decades, environmental factors are also likely to play a role in the development of AD in an individual. We found evidence that indicated that AD is more prevalent in an urban than in a rural environment and thus urban lifestyle seems to be associated with a higher prevalence of AD. Which aspects of urban lifestyle are important should be subject of further research in which also time point and duration of exposure is included.

The second part of this thesis (chapter 3 & 4) aims at improving the diagnostic work-up as well as the outcome reporting on AD. The evidence on validity of diagnostic criteria is summarized in a systematic review, indicating the lack of uniformity in the use and lack of validity of some diagnostic criteria. Also, the millennium criteria were validated and refined. These criteria are especially aimed for hospital-based studies and enable sub-analysis by differentiating between atopic and atopiform dermatitis, which is a niche in the market. Concerning the outcome measures, we established the responsiveness of four clinical outcome measures, the Eczema Area and Severity Index, Patient-oriented Eczema Measurement and (objective) Severity Scoring of Atopic Dermatitis. Thereby, we have calculated the minimal clinically important difference. Based on this, accurate power calculations can be made and it can be decided if statistically significant differences reflect a real change in disease severity to deserve consideration in clinical practise.

In order to increase the impact of clinical research, it is inevitable to ensure adequate (and repeatable) diagnosis and outcome reporting. For both applies that many different tools are currently used and that most of them lack good methodological and clinimetric quality. This impairs the quality of evidence and disables comparison between studies, thereby prohibiting generating higher levels of evidence. As stated by Tugwell from the OMERACT initiative (Outcome Measures in Rheumatoid Arthritis Clinical Trials), 'clinical trials are only as credible as their endpoints'. We would like to add 'and as their diagnostic criteria'. In this light, harmonizing diagnostic criteria and clinical outcome reporting is of vital importance. Initiatives to establish consensus like HOME (Harmonizing

Outcome Measurement in Eczema) should be welcomed with enthusiasms by clinical researchers as well as by key players in the expert field.

Besides clinical severity outcome measures, biomarkers, when established as reliable outcome measures in future studies, could provide us with real objective measurement.

The third part of this thesis (Chapter 5) was aimed at off-label systemic therapy for AD. Physicians run out of treatment options for patients with severe AD as registered treatment options can be contra-indicated or fail. In the absence of new major developments for therapeutics that specifically targets severe AD, more general immunosuppressive drugs are focus of investigator-initiated research. In a randomized controlled trial, evidence was generated that showed that methotrexate and azathioprine are effective treatment options for severe AD and thereby we have expanded the evidence-based treatment armentarium. Future studies should be performed to establish optimum dosages, to confirm the long-term safety profile of both methotrexate and azathioprine, to confirm their role in children and to compare both treatments with other therapies such as cyclosporin, oral corticosteroids and other disease modifying anti-inflammatory drugs.

In order to identify applicable off-label treatment options and to indicate new fields of research, available evidence for often off-label prescribed drugs should be summarized. We have provided a methodologically sound overview of two systemic drugs that have potential benefit in AD and some other recalcitrant skin diseases. Besides the need for long-term registries, those reviews are needed to provide evidence for guidelines to justify off-label therapy.

FUTURE CONSIDERATIONS

Without any doubt, a lot has changed over the last years concerning research in AD. It is a somewhat provocative comment to state that research in AD in the past years has been mostly dominated by a lack of uniformity, but not less true. However, times are changing and initiatives to harmonize are now put forward. Dermatologists worldwide seem to be more and more aware of the need to find consensus. If we want to proceed to generate high levels of evidence and thereby expanding our knowledge of the management of AD, it is conditional to combine our forces and clear the road.

8 BIBLIOGRAFIE EN DANKWOORD

BIBLIOGRAFIE

1. **Schram ME**, Tedja AM, Spijker R, Bos JD, Williams HC, Spuls PI. Is there a rural/urban gradient in the prevalence of eczema? A systematic review. *British Journal of Dermatology* 2010*;* 162:964-73.

2. **Schram ME**, Jensema AJ, Sillevis Smit JH. Atopisch eczeem. *Modern Medicine* 2010;34:33-7.

3. **Schram ME**, Spuls PhI, Bos JD. Off-label use of efalizumab in dermatology *Expert review of Dermatology* 2010; 5: 535-547.

4. **Schram ME**, Borgonjen RJ, Bik CM, van der Schroeff JG, van Everdingen JJ, Spuls PI; Off-Label Working and Project Group of Dutch Society of Dermatology and Venereology. Off-label use of azathioprine in dermatology: a systematic review. *Archives of Dermatology* 2011; 147:474-88.

5. **Schram ME**, Leeflang MM, den Ottolander JP, Spuls PI, Bos JD. Validation and refinement of the Millennium Criteria for atopic dermatitis. *Journal of Dermatology* 2011, in press

6. **Schram ME**, Roekevisch E, Leeflang MM, Bos JD, Schmitt J, Spuls PI. A randomized trial of methotrexate versus azathioprine for severe atopic eczema.*Journal of Allergy Clinical Immunology* 2011, in press

7. **Schram ME**, PhI Spuls, MMG Leeflang, R Lindeboom, JD Bos, J Schmitt. (Objective) SCORAD, EASI & POEM for atopic eczema: Responsiveness and minimal clinically important difference. *Allergy* 2011, in press

8. Bos JD, **Schram ME**, Mekkes JR. Dermatologists are Essential for Quality of Care in the General Practice of Medicine. *Actas Dermosifiliograficas* 2009;100:101-5.

9. Bos JD, **Schram ME,** Brenninkmeijer EE, Middelkamp-Hup MA, Spuls PI, Smitt JH. Atopic eczema or atopiform dermatitis. *Experimental Dermatology* 2010;19:325-31.

10. Brenninkmeijer EE, **Schram ME**, Leeflang MM, Bos JD, Spuls PI. Diagnostic criteria for atopic dermatitis: a systematic review. *British Journal of Dermatology* 2008;158:754-65.

11. Lubbert PHW, **Schram ME**, Leenen LPH. Is there a reason for spine board immobilization in the emergency department for patients with a potential spinal injury? *European Journal Traumatology* 2005;31:375-8.