Atopic dermatitis

The epidemiology, causes and prevention of atopic eczema

Edited by

Hywel C. Williams

Foundation Professor of Dermato-Epidemiology
University of Nottingham, Queen's Medical Centre,
Nottingham

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK http://www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011–4211, USA http://www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcon 13, 28014 Madrid, Spain

First published 2000

Printed in the United Kingdom at the University Press, Cambridge

Typeface Utopia 8½/12. System QuarkXPress® [SE]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data

The epidemiology of atopic dermatitis / edited by Hywel C. Williams.
p. cm.
ISBN 0 521 57075 1
1. Atopic dermatitis – Epidemiology. I. Williams, Hywel C.
RL243.E65 2000
614.5'9521–dc21 99–31355 CIP

0 521 57075 1 hardback

Contents

1

What is atopic dermatitis and how should it be defined in epidemiological studies?

Hywel C. Williams

Developing reliable diagnostic criteria may be as tedious as filling in muddy holes with concrete but both provide the foundation on which all else depends (Professor R.E. Kendell, 1975)

What is atopic dermatitis?

A distinct 'entity' or a continuum?

A particular problem hindering understanding of disease classification in dermatology today is 'binary thought disorder'. Binary thought disorder is a state whereby individuals are unable to appreciate that most biological phenomena do not fit neatly into all-or-nothing 'either/or' categories. Ever since Pickering shook the medical world by daring to suggest that essential hypertension, a major cause of death, was a graded characteristic which shaded insensibly into normality (Oldham et al., 1960), many physicians still have difficulties in viewing diseases as a quantitative or multidimensional process. Yet in a population setting, even with diseases like hepatitis, which might at first appear to conform well to a dichotomous disease definition, one sees a gradation of sickness ranging from those who are apparently healthy (many of whom will have subclinical infection), those who have mild gastrointestinal symptoms (some of whom are not infected), some who are moderately ill and some who are moribund or dead. Similarly, in atopic dermatitis (AD) one sees some children with normal skin (but with high IgE and positive skin prick tests to allergens), children with mucosal atopy and dry skin only, some with one episode of itching and erythema in just one flexure, and others with classical persistent flexural disease. Perhaps the most appropriate question therefore is not to ask 'has he/she got atopic dermatitis, yes/no?' but rather 'how *much* atopic dermatitis does he/she have?' (Barker & Rose, 1979).

It is still not clear whether 'degree of atopic dermatitis' (if it can truly be expressed on a quantitative scale) is normally distributed in populations, or whether a bimodal distribution exists, the trough of which indicates a point of rarity or cut-off between 'disease' and 'normality'. Particular care has to be given to small population studies which claim disease bi- or trimodality, as artefactual peaks and troughs can easily be produced by chance or by manipulating the way in which individual features are scored. Two population-based studies in Germany (Figure 1.1) using an AD 'score' (Diepgen & Fartasch, 1992) suggest that 'degree of AD' could well be part of an underlying Gaussian distribution (Diepgen, T., personal written communication, 1998). It is possible that genetic factors, such as IgE hyper-responsiveness, and environmental triggers, such as high concentrations of house dust mite, shift the whole distribution of individuals to the right (Figure 1.2a), thereby increasing the proportion of individuals within the threshold whereby AD becomes manifest. The corollary of such a theory is that *any* individual could develop the clinical syndrome of 'AD' under the right circumstances, and that there is no ceiling to the prevalence of AD that could be theoretically achieved under appropriate adverse conditions.

Another viewpoint is that there exists in any one

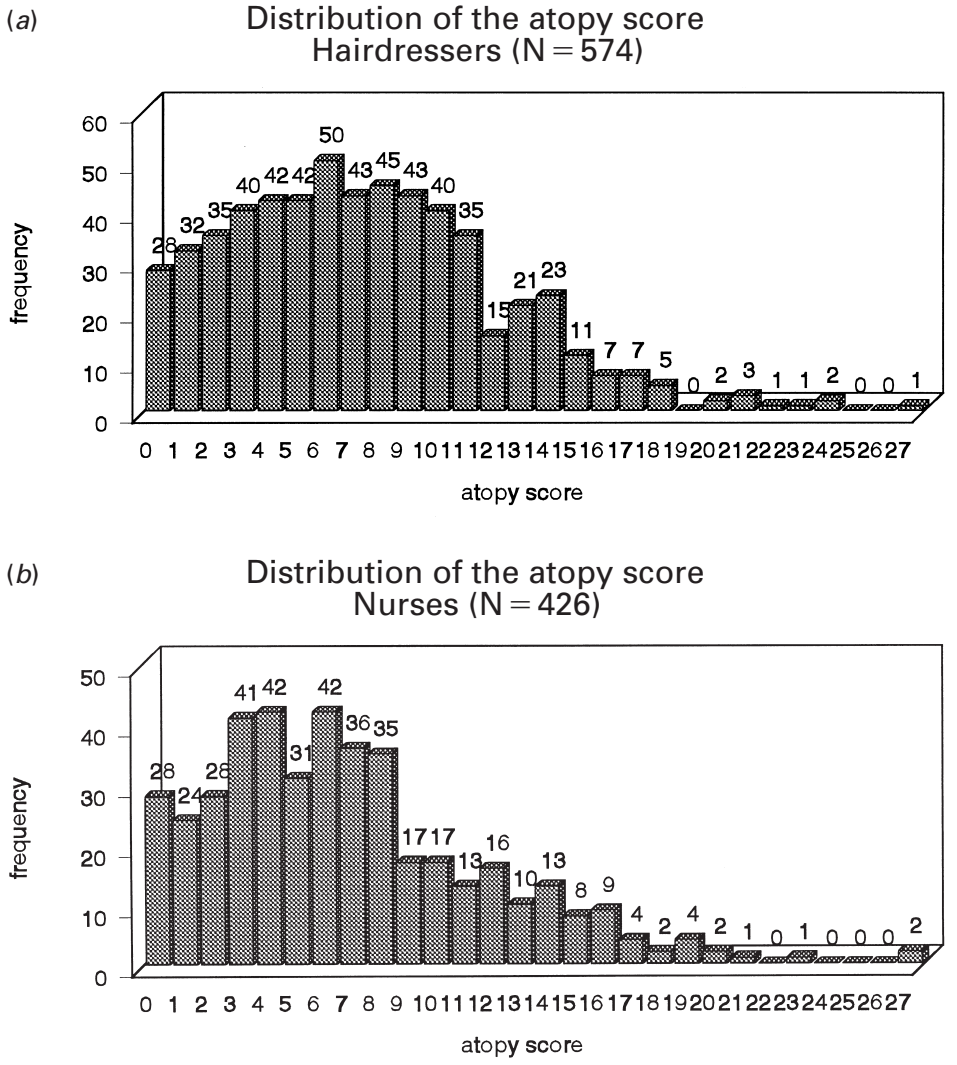

Fig. 1.1. Distribution of score of atopic skin diathesis amongst an unselected population of 574 hairdressers in Germany (Diepgen, T., written communication, 1998). A similar distribution is seen for 426 junior nurses

population a finite proportion of people who are genetically predisposed to AD, with additional perinatal or environmental factors determining the proportion of such people who will express disease at any one given time (Figure 1.2b). This concept could be one possible explanation of why the prevalence of AD has appeared to remain stable at around 20% in Japanese cities over the last 20 years (Sugiura et al., 1997), whereas it has increased two- to threefold at levels below 20% in Northern Europe (Williams, 1992). In other words, Japan has already witnessed its maximum prevalence in AD due to exposures correlated with rapid industrial and social development ahead of Western cultures, so that a 'state of saturation' has now been reached whereby nearly all predisposed subjects express disease. Such a notion would appear to fit well with the idea that a genetic

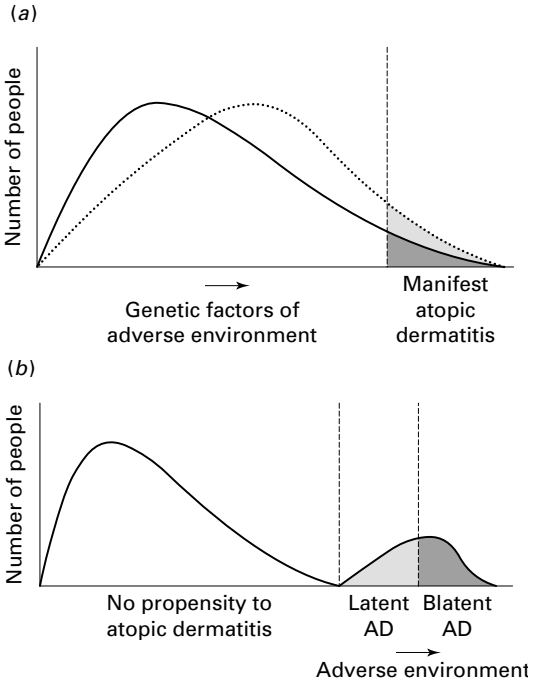

Fig. 1.2. (a) and (b) Is 'degree of atopic dermatitis' a continuum that is normally distributed in populations, with factors that enhance predisposition (genes) or precipitancy (allergenic environment) shifting the whole distribution to the right (Figure 2a)? Or are AD scores distributed bimodally, with only a fixed proportion of the population capable of expressing a manifest disease (Figure 2b)?

factor such as atopy or IgE responsiveness is that necessary predisposing influence but, as is discussed later, IgE responsiveness is neither necessary nor sufficient to diagnose AD. Until the genetic basis for AD and its subtypes becomes clearer, it would be wise not to make any assumptions on where normality ends and AD begins.

Measuring the total amount of disease in a population on a quantitative scale may sound attractive in that it provides us with information on all of the individuals in that population, but it also presents some serious difficulties for epidemiologists. There is a need to return to our main purpose of disease definition, i.e. to assist in the comparison of groups of people and to increase our predictive abilities about individuals. Thus, whilst a log odds score of AD of 3.27 might mean something to a researcher trying to predict the degree to which a hairdressing apprentice is likely to develop irritant hand dermatitis (Fartasch & Diepgen, 1994), such a score would have little meaning to the thousands of doctors in primary care who wish to describe the disease pattern in their population. Comparing mean AD scores between populations may be an interesting academic exercise, but its biological significance may be obscure. Another danger of quantitative scales is that they are open to statistical abuse on the erroneous assumption that such scales behave like other continuous variables such as height and weight. It is a natural reflex for workers to attempt mathematical manipulations when faced with a scale of numbers. Whereas it is true that a person who weighs 100 kg is twice as heavy as a person weighing 50 kg, it may not be assumed that a person with an AD score of 6 has twice the amount of AD as someone with a score of 3. In addition, the weights applied to individual disease features derived from regression models are highly dependent upon the population who were selected to derive the criteria (Wells, Feinstein & Walter, 1990), and ten different studies could produce ten different sets of criteria, each with different weighting, leading to international disputes on which weighting was 'correct' (Kendell, 1975).

Dichotomous or categorical disease definitions, on the other hand, require a line to be drawn between disease and nondisease. Even the word 'diagnosis', which is derived from the Greek words *διά* (the number two), and *γιγνώσκειν* (to perceive), implies a dichotomous outcome. Such dichotomous definitions are far more widely used and easily understood in public health settings, and are therefore logical choices for promoting international communication. Their main drawback is that the imposition of boundaries between those who are sick and those who are apparently healthy, *almost always results in the misclassification of some subjects.* Unless the disease in question has an abrupt natural cut-off between normal and abnormal, the imposition of an arbitrary dividing line will always

be subject to a trade off between sensitivity (proportion of true positives correctly identified by the test criteria) and specificity (proportion of genuine 'non-cases' correctly identified) (Sackett et al., 1991). Thus, very sensitive symptoms such as 'itchy skin' might include all subjects with AD, but it would also be highly nonspecific, including subjects with other pruritic skin diseases such as lichen planus or tinea pedis (Williams et al., 1994a). By contrast, very specific signs such as infra-auricular fissure (Tada et al., 1994), might exclude all other skin diseases in a population survey, but it would also exclude most cases of AD as the sign is encountered so infrequently in a population setting where mild cases predominate (Williams et al., 1994a).

Exclusion of those who have extremely mild or asymptomatic disease may be desirable in public health surveys, but it must be realized that drawing the line between disease and nondisease has to be an arbitrary process. Various techniques such as receiver–operator curves (Freiman et al., 1978) may be used to assist in deciding the optimal cut-off between sensitivity and specificity for continuous data, but these techniques need to be evaluated in the clinical context of the question being addressed, and not as a means of abrogating responsibility for decision making. As is seen later in this chapter, sometimes very specific criteria are needed at the expense of sensitivity, and using a cut-off derived from a receiver–operator curve may be inappropriate for this purpose. Despite its limitations, it is felt that a binary definition for AD would be far more readily understood and used by clinicians and epidemiologists throughout the world (Kendell, 1975).

More than one disease?

Some have suggested that more than one type of atopic dermatitis exists (Imayama et al., 1992; Wüthrich & Schudel, 1983). There are clinicians who, having observed individuals in a hospital setting, have favoured a division of AD into those with 'pure' AD limited to childhood and those with more chronic disease associated with respiratory atopy (Roth, 1987). Great care has to be taken in making inferences about such disease associations from hospital studies since disease co-occurrence and disease severity are positively associated with hospital referral. This selection bias can result in all sorts of misleading inferences (Gerber et al., 1982). Others have suggested an intrinsic and extrinsic form of AD based on the presence or absence of reactivity to allergens (Wüthrich & Schudel, 1983). Such a division may be practical when advising individual patients, but its validity is limited by our incomplete knowledge of which allergens to test for, which type of test one should use (e.g. skin prick test, aeroallergen patch test, oral challenge, or combinations of these), the relevance of such skin test results to clinical disease (David, 1991), and because allergen reactivity can fluctuate over time. In adults, further confusion may arise from irritant or allergic contact dermatitis mimicking or exacerbating AD.

Some workers have taken things much further by suggesting that there may be at least four different subtypes of AD based on different combinations of skin prick and aeroallergen testing (Imayama et al., 1992). Inevitably, the number of apparent subcategories of disease will increase according to the number of tests and cross-tabulations performed. For example, even in normal individuals, the probability of getting an abnormal serum biochemistry blood test result at the 5% significance level is 0.64 when 20 tests are performed. Data-driven post hoc subdivisions for AD are therefore only useful if they are subsequently shown to increase our predictive ability such as prognosis or responsiveness to treatment. No such studies have been performed to date.

An important consideration in relation to the subgroup issue in AD is the extent to which failure in recognizing subgroups can obscure important epidemiological disease associations. One indirect response to such a question might be that if misclassification was gross, important epidemiological disease associations would have been obscured. This has certainly not been the case to date for studies which have considered the clinical syndrome of AD (Williams, 1997a). Whilst it is true that perfect classification might have increased the magnitude of such associations, the fact that so many

relatively weak associations have been consistently shown for AD as it is currently classified, argues against major misclassification, at least in studies of children. The key question for researchers investigating AD in populations is not 'how can I be sure that all individual cases in my study have a homogeneous disease?' but 'is what is defined as atopic dermatitis in this study measuring a concept that is *useful* for health care workers?'

Is atopic dermatitis atopic?

The concept of 'atopy' has troubled many scientists since Coca and Cooke introduced the term in 1923 as meaning 'strange disease' (Coca & Cooke, 1923). Strange disease it certainly is, for whilst many physicians are content with the notion that 'atopy' represents a familial hypersensitivity of skin and mucous membranes against environmental substances associated with increased IgE production, the quest for consistent clinical, immunological or genetic markers that encompass all individuals fitting the above clinical picture has been fruitless. Some define atopy as the development of IgE antibody in response to antigen exposure (Turner, 1987), although individuals who make large amounts of IgE are not all atopic, e.g. those infected with parasites and, conversely, 20 to 50% of individuals with typical clinical AD exhibit normal values of total or specific IgE (Dotterud et al., 1995; Edenharter et al., 1998). It is also unclear whether the immediate hypersensitivity reactions encountered are relevant to the atopic dermatitis as concurrent mucosal allergy is often also present. Some have defined definite atopy on the basis of more than one positive skin prick test to common allergens, although such a definition could include 50% of the population (Barbee et al., 1987), most of whom will not have clinical disease. In addition, such 'atopy' may be inherited independently from the propensity to specific allergic disease (Sibbald, 1986). Ring has recognized the shortcomings of the traditional use of the term 'atopy' and has proposed that it should be redefined as a 'familial hypersensitivity of skin and mucous membranes against environmental substances, associated with increased IgE production and/or altered nonspecific reactivity' (Ring, 1991).

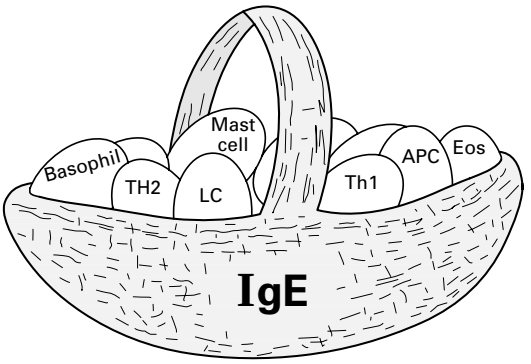

Fig. 1.3. We should not put all our atopic dermatitis eggs in the IgE basket

Recent research suggests that a type IV delayed-type hypersensitivity response involving different subsets of sensitized T-helper lymphocytes may be an important mechanism of allergic response in atopic dermatitis (Leung & Geha, 1986; Bos et al., 1992). Altered nonspecific skin reactivity such as increased α and decreased β adrenergic responsiveness and abnormalities in vasoactive mediators may also be key abnormalities underlying AD (Hanifin, 1992a). Another school of thought proposes that the crucial underlying problem of AD is that there is a primary defect in the barrier function of the epidermis, leading to a constellation of changes such as inflammation, itch and enhanced allergen–irritant penetration (Ogawa & Yoshiike, 1993).

Thus, although AD is strongly *associated* with increased total or specific IgE responsiveness, the role of classical immediate hypersensitivity in AD as a *necessary* phenomenon may have been overemphasized (Figure 1.3). As further research at a cellular level highlights the interaction between mast cell, eosinophil, Langerhans cell and T-lymphocyte in AD (Hanifin, 1992a), perhaps it would be wiser not constraining ourselves into the Gell and Coombs classification of hypersensitivity phenomena (Coombs & Gell, 1963) or a discussion of whether AD is atopic or not, but rather to ask ourselves to what degree is AD atopic?

In immunological terms, therefore, some might feel that the word 'atopy' when used in the term 'atopic dermatitis' is inappropriate or does not have a precise meaning. Although raised total and specific IgE levels and skin prick tests are frequently abnormal in atopic dermatitis subjects, their precise role in the pathogenesis of atopic dermatitis is still far from clear. The main argument for retaining the word 'atopic' in atopic dermatitis is to assist in separating our clinical concept of AD, a chronic pruritic disorder of early onset with inflammatory skin changes favouring flexural sites in individuals with a propensity to develop concomitant inhalant allergy, from other forms of dermatitis such as seborrhoeic, discoid, asteatotic, irritant and allergic contact dermatitis.

Dermatitis or eczema?

A detailed argument of the pros and cons of each term is beyond the scope of this chapter and may be found elsewhere (Ackerman, 1982). It is a sad reflection on modern dermatology that so much useful scientific energy has been wasted on arguing whether the term eczema or dermatitis should be used. Such debates have generated more heat than light on our understanding of the condition. Internationally, perhaps the term atopic dermatitis is more widely used than atopic eczema. The author accepts that the terms atopic eczema and atopic dermatitis are synonymous, and that in some countries such as the UK, others might prefer to use terms such as atopic eczema in order to avoid connotations of an occupationally acquired dermatosis.

Regressive and progressive nosology of disease

Based on the above discussion, some would argue that 'atopic dermatitis' is neither 'atopic' nor 'dermatitis'. Although the term 'atopic dermatitis' may have a scientific and objective ring to it, in practice it may not increase our predictive ability much more than the phrase 'itchy red rash in the skin folds'. Attaching a name to a condition can sometimes create a spurious impression of understanding so that we cease to investigate the nature of the disease further (Kendell, 1975). Hardin coined the word panchreston (meaning 'explain-all', by analogy with panacea, or 'cure-all') to draw attention to the ways in which jargon is used to provide comforting but meaningless explanations for things we do not really understand (Hardin, 1956). Pearce has suggested that many fashionable 'new' diseases, such as post-traumatic syndrome, posttraumatic stress disorder, chronic fatigue syndrome and repetitive strain injury, are simply labels which hinder appropriate treatment and further research (Pearce, 1994). Such regressive nosology was highlighted by Abrams (1994), who pointed out that the term 'prostatism' has been used for many years to imply a prostatic cause for urinary symptoms when, in reality, almost no evidence exists for such a cause. Nosology is not simply a matter of semantics, as many men with 'prostatism' without bladder outflow obstruction are still being subjected unnecessarily to prostatectomy. Other terms such as 'benign prostatic hyperplasia' carry a spurious diagnostic authority, which may be translated into treatment without a proper diagnosis. Both Abrams and Pearce suggest that we would be better advised to adhere to established phenomena, and to be unashamed at honest diagnoses such as 'facial pain of unknown aetiology' or 'lower urinary tract symptoms' – terms which at least prompt further description, consideration and research. The situation is summed up nicely by Pearce who points out that 'diagnoses are not diseases, but are ever changing representations of disease to permit convenient communication and to allow brief descriptive insights into their nature'.

Progressive nosology, on the other hand, defines disease on the basis of a hierarchy of external evidence ranging from clinical descriptions to aetiological agents. As Scadding (1963) points out, myxoedema was originally defined as a clinical syndrome, but came to be defined as a disorder of function – a disorder of deficiency or utilization of thyroxine. This new definition will include some patients such as those with hypopituitarism who were not embraced by the original syndrome, and will exclude others with localized myxoedema in the

absence of hypothyroidism, who were included in the original description. This is an example of progressive nosology, and similar examples are to be found in dermatology, such as the division of 'pemphigus', which formally referred to several diseases in which blistering was a feature (Pye, 1986) into pemphigoid, pemphigus and linear IgA disease on the basis of immunological discoveries. Changes of this sort are not a problem providing they are explicit, and that they confer benefits to patients (Kendell, 1975). By analogy, what we recognize as a clinical syndrome of atopic dermatitis today may in time be shown to be caused by three or four different agents. This does not imply that the original older criteria were 'wrong' at the time, provided they measured something useful or that they were instrumental in stimulating further research into the aetiology of that syndrome.

The need for a disease definition

Trying to define one of the most common skin diseases is not easy. Quite apart from the formidable difficulties of trying to define a disease which is variable in morphology, distribution and periodicity, and which lacks a laboratory reference standard, attempts to propose diagnostic criteria may be viewed as an imposition by other experienced physicians who are perfectly happy with the way in which they diagnose atopic dermatitis in individuals. Therein lies the crux of the matter. Diagnosis by physicians based on many years of clinical pattern recognition is entirely appropriate when dealing with individual patients. Problems begin, however, when *groups* of patients have to be described and compared. Whether this be the comparison of different prevalence rates from around the world, or comparison of therapeutic regimens, it is essential to know that different workers all refer to the same entity. Disease definition is essentially an aid to communication. Without it, all scientific communication would be impossible and our professional journals would be limited to case reports, anecdotes and statements of opinion.

There is always the possibility that the methodology for developing disease definitions becomes an end in itself. Disease definitions have meaning only in context to the biological question which is being asked. Different types of studies may require different types of definition. Disease definition is an evolutionary process which should be modified in the light of new knowledge.

Ways of defining atopic dermatitis

Various strategies can be employed in empidemiological studies for defining a disease dichotomy. For ordinal data (e.g. atopic dermatitis score) a statistical approach may be suitable. For example, any subject displaying a value above or below two standard deviations of a range of values of AD scores obtained from a representative population may be considered as abnormal. The biological meaning of such definitions may be obscure, however, and definitions based on two or more standard deviations from the mean also presupposes that the prevalence of all disease is 2.5% in each tail.

Prognostic definitions utilize elements of the condition which are associated with impaired outcome, such as sleep loss. Such an approach is useful for excluding asymptomatic or trivial disease, but the precise effects of disease on functional ability in many skin diseases is unknown.

Operational definitions are based on defining features for which action (in the form of cost effective treatment) is preferred to inaction. These are highly dependent on available resources and competing needs. This approach may be useful for implementing public health policies such as treatment of infestations in individual countries, but would be of little use in prevalence or aetiological studies.

On balance, a clinical approach of summarizing a constellation of symptoms and signs seems to be the most relevant to studying the epidemiology of AD today.

What is a good disease definition?

Before describing the various definitions for AD which have been used in epidemiological studies, it

Table 1.1. A good epidemiological definition for atopic dermatitis

1. Valid (sensitive and specific)
2. Repeatable (between and within observer)
3. Acceptable to the population
4. Rapid and easy to perform by field workers
5. Coherent with prevailing clinical concepts
6. A reflection of some degree of morbidity
7. Comprehensive in its applications
8. Comparable with other studies

is wise to consider what constitutes a good disease definition. These are summarized in Table 1.1 and are discussed in detail elsewhere (Williams, 1997b).

Diagnostic criteria for use in epidemiological studies

The dark ages

Although disease definition is perhaps the most fundamental step in any form of medical research, at least 12 synonyms for atopic dermatitis (AD) were in widespread use in Northern Europe (Table 1.2) up until the late 1970s (Sulzberger, 1983). Even dermatology texts use reflexive statements to define atopic dermatitis such as 'atopic dermatitis is the characteristic clinical type of dermatitis usually associated with atopy' (Champion & Parish, 1986), or 'eczema is a disease which shows eczematous features'. Such problems can be viewed in terms of nominalistic versus essentialist classification of disease (Burton, 1981). Nominalistic disease definitions imply that diseases have no real existence outside the individual patient. Even infectious agents such as the tubercle bacillus, which can be 'captured' and kept in a culture bottle like some demon, can produce a very wide range of clinical manifestations ranging from commensal existence to acute miliary tuberculosis. Similarly, atopic dermatitis does not conform to an essentialistic disease model (i.e. the disease is an entity in itself which 'attacks' patients), but rather a syndrome of related clinical features arising in response to a number of endogenous and exogenous factors. The classification of a disease such as atopic dermatitis is thus the classification of patients, all of whom are different. 'Dis-ease' implies a complex interaction between external agents and host which will depend on a range of factors such as genetic predisposition, previous exposure to sensitizing agents and irritants, age, nutrition, hygiene, emotional and social well being and access to medical services.

Table 1.2. Synonyms for atopic dermatitis

- 'Eczema'
- Atopic eczema
- Infantile eczema
- Eczéma constitutionnel
- Flexural eczema
- Prurigo Besnier
- Allergic eczema
- Childhood eczema
- Lichen Vidal
- Endogenous eczema
- Spätexudatives Ekzematoid
- Neurodermatitis (constitutionalis)

Such a nominalistic approach can be taken to the extreme, however, for if we maintain that every patient is unique, then there could be as many diseases as there are patients. Whilst tailoring treatment to suit a unique constellation of problems in a particular individual might have some advantages in a clinical setting, as might have been the case in the 'dark ages', it is of little use in an epidemiological context where *groups* of patients need to be compared. Although some degree of nominalism is to be encouraged in order to reflect host factors, it is important that any patients defined by such an approach should behave similarly, so that we are able to communicate our findings on the morbidity and causes of the condition described by such a disease label.

The Hanifin, Lobitz and Rajka diagnostic criteria

The unsatisfactory situation of the dark ages came to an end with the suggestion by Rajka, Lobitz and

Table 1.3. The Hanifin and Rajka diagnostic criteria for atopic dermatitis

Must have three or more basic features

- Pruritus
- Typical morphology and distribution:
  Flexural lichenification or linearity in adults. Facial and extensor involvement in infants and children
- Chronic or chronically relapsing dermatitis
- Personal or family history of atopy (asthma, allergic rhinitis, atopic dermatitis)

Plus three or more minor features

Xerosis
Ichthyosis/palmar hyperlinearity/keratosis pilaris
Immediate (type I) skin test reactivity
Elevated serum IgE
Early age of onset
Tendency towards cutaneous infections
Tendency towards nonspecific hand or foot dermatitis
Nipple eczema
Cheilitis
Recurrent conjunctivitis
Dennie–Morgan infraorbital fold
Keratoconus
Anterior subcapsular cataracts
Orbital darkening
Facial pallor/facial erythema
Pityriasis alba
Anterior neck folds
Itch when sweating
Intolerance to wool and lipid solvents
Perifollicular accentuation
Food intolerance
Course influenced by environmental/emotional factors
White, dermographism/delayed blanch

Hanifin of a set of major and minor diagnostic criteria for atopic dermatitis (Rajka, 1975; Hanifin & Lobitz, 1977; Hanifin & Rajka, 1980) based on 24 clinical symptoms and signs (Table 1.3). In order to qualify as a case, subjects are required to have at least three out of four major features, or four out of five in a recent modification (Hanifin, 1992b), and at least three of the minor features listed in Table 1.3. These criteria undoubtedly represented a major step forward in ensuring some degree of uniformity of atopic dermatitis subjects in subsequent hospital studies and as a framework for further developments.

However, as Schultz Larsen and others have found out (Schultz Larsen & Hanifin, 1992; Seymour et al., 1987; Svensson, Edman & Möller, 1985; Visscher, Hanifin & Bowman, 1989; Diepgen & Fartasch, 1991) these criteria are unworkable in population-based studies. Many of the criteria, e.g. 'pityriasis alba', are not precisely defined (Hanifin, 1983), some (e.g. keratoconus) are very infrequent (Kennedy, Bourne & Dyer, 1986; Gelmetti, 1992), and some, such as white dermographism, are nonspecific (Svensson et al., 1985). They were derived in an empirical fashion in relation to clinical experience with predominantly white hospital-based cases of AD, and division into major and minor criteria was also empirical. More importantly, the criteria were not formally validated against the physician's diagnosis or tested for repeatability. In addition, the criteria contain invasive tests which are rarely used in routine clinical practice, and which might not be suitable for large studies involving children (Seymour et al., 1987).

Although the list of major criteria can usually be memorized, the list of over 30 minor criteria is difficult to assimilate into working practice, and introduces a large potential source of between- and within-observer variation. It has been shown that the human mind can process only about seven items of information simultaneously (Miller, 1956), and accuracy of diagnosis is usually diminished when physicians are presented with superfluous data (de Dombal et al., 1972). In addition, clinicians seldom incorporate arborizing strategies such as algorithms for diagnosis in clinical practice (Barrows et al., 1982). Although the Hanifin and Rajka criteria have been deployed in some population-based studies (Neame, Berth-Jones & Graham-Brown, 1993; Bakke, Gulsvik & Eide, 1990), the author suspects that what often happens in such cases is that physicians first decide whether or not a subject has AD using a pattern recognition approach (Sackett et al., 1991; Neufield et al., 1981), then seek confirmatory features from a wide choice of criteria in order to justify their initial clinical impression. Whilst the

Hanifin and Rajka diagnostic criteria may continue to be useful in some hospital studies because of their probable high sensitivity, their complexity and unknown validity makes them unsuitable for use in population-based studies or as a diagnostic aid to doctors in primary care.

The modern age

A number of groups have examined the usefulness of Hanifin and Rajka's diagnostic criteria (Mevorah et al., 1988; Diepgen, Fartasch & Hornstein, 1989; Kang & Tian, 1987; Kanwar, Dhar & Kaur, 1991; Kim, Chung & Park, 1993; Sehgal & Jain, 1993; Rudzki et al., 1994), but instead of addressing the crucial issues such as validation against the physician's diagnosis or deciding which should be designated major and minor criteria, these groups have tended to focus on small differences in the application of the 30 minor criteria mentioned in Hanifin and Rajka's original abstract. A recent editorial in *The Lancet* fuelled this preoccupation with minor criteria for AD (Rothe & Grant-Kels, 1996). These minor features may vary considerably following slight adjustments in their definition and interpretation. Features such as infraorbital folds, periorbital pigmentation and hyperlinear palms are probably highly dependent on the age, sex and ethnicity of the population under study, which may explain the large discrepancies between these studies (Mevorah et al., 1988; Williams et al., 1996a). The atopic dermatitis cases used in these studies have all been hospital-based, which might explain the high frequency of odd signs such as anterior neck folds, Hertoghe's sign, hyperlinear palms, etc. – which are probably more frequently seen in a severe or chronic subset of AD cases. In none of the studies has the observers' recording of the presence or absence of these signs been blinded to the exposure status of the patient, and repeatability (between- and within-observer) of signs seems to have been overlooked.

A notable exception to these studies is the work of Diepgen et al. who derived a scoring system of useful diagnostic features of AD based on χ^2 values (Diepgen, Fartasch & Hornstein, 1989). They compared established hospital-ascertained AD cases ($n = 428$) with normal young adults from the community who did not have AD ($n = 628$), with respect to a number of Hanifin and Rajka's diagnostic criteria. They used clinical evidence of recurrent flexural itching or lichenified dermatitis as a gold standard for cases. This implied that it was impossible to assess the usefulness of history or visible flexural involvement as a feature of AD since, by definition, this criterion had 100% specificity and 100% sensitivity. On the basis of their χ^2 results, Diepgen et al. showed that some features, such as personal or family history of atopy, which are considered as 'major' features in Hanifin and Rajka's original criteria, are not as useful as some previously designated 'minor' features such as xerosis. They also showed that raised total serum IgE (>150 units/ml) and a positive radioallergosorbent test (RAST) for inhalant allergens were neither particularly sensitive nor specific for AD, the corresponding χ^2 values being less than most of the anamnestic and clinical features. Using their scoring system, they demonstrated good separation between cases and controls, although the scoring system was tested on the same data set from which the criteria were derived, as opposed to an independent sample. It should also be noted that their scoring system refers to discrimination of hospital-based AD cases from community controls who do not have AD, and when tested against a sample of 329 adults with skin disease recruited from hospital outpatients, specificity dropped from 97% to 84% (Diepgen, Sauerbrei & Fartasch, 1994). More importantly, repeatability of individual features has not been taken into account. Their study is nevertheless by far the largest and most systematic analysis of diagnostic criteria of AD to date, and their scoring system in particular may prove to be useful in estimating the risk of unmasking AD in nonaffected individuals, as might be considered in preemployment examinations. The author agrees with their conclusion that the diagnosis of AD should be based on traditional anamnestic and clinical features.

Schultz Larsen & Hanifin (1992) have proposed a questionnaire method for estimating AD, which includes many features which other workers have considered to be important in the diagnosis of AD. It is written using clear language, although the diagnostic label of 'eczema' (which might have many determinants) is mentioned throughout. Instead of using a binary disease definition (i.e. atopic dermatitis, yes/no), Schultz Larsen and Hanifin have chosen the categories of 'definite AD', 'possible AD' and 'no AD' as the main outcome measures based on a points system, the derivation of which is unclear. Such an approach is an attractive simplification of numerical estimations of the probability of AD in what may well be a disease continuum, but it is not clear how researchers comparing prevalences should deal with the 'possible AD' category, which could form the bulk of cases in community surveys. Defining opposite ends of the AD continuum is easy, but most prevalence or morbidity surveys will require a binary definition which offers a reasonable compromise between specificity of diagnosis and exclusion of asymptomatic disease.

Buser et al. have explored the validity of another questionnaire based on anamnestic criteria derived from the Hanifin and Rajka list against dermatologist's diagnosis in a sample of German schoolchildren (Buser et al., 1993). Although the questions have not yet been tested on an independent population or for repeatability, encouraging results were shown for a combination of three major and one minor feature. The authors chose to exclude ten children with equivocal diganosis from the main analysis, which perhaps defeats the purpose of the exercise.

The UK refinement of Hanifin and Rajka's criteria

In view of the absence of a definition for atopic dermatitis with known validity and repeatability, a UK working party set about the task of developing a minimum list of reliable discriminators for AD in 1990, using the Hanifin and Rajka list of clinical features as the building blocks. In addition to validity, repeatability and simplicity, a further requirement of the definition was that it should correspond well to our clinical concept of disease, be applicable to different ages and ethnic groups and be acceptable to subjects under study (Williams, 1997b). The detailed development of these criteria is to be found in six key papers published in the *British Journal of Dermatology* (Williams et al., 1994a, 1994b, 1994c, 1995a, 1996b; Popescu et al., 1998). Briefly, the study involved a national case-control study to examine the validity of specific symptoms and signs in relation to experienced physicians' diagnosis of AD. These physicians were consistent in their diagnosis of AD, and repeatability of signs was investigated in a separate study. Regression techniques and clinical consensus were used to derive a minimum list of reliable discriminators, which were then tested in independent validation studies. In order to capture the intermittent nature of AD and to minimize possible seasonal fluctuations in AD activity, the diagnostic criteria are recommended to be used as a 12-month period prevalence measure. The UK refinement of Hanifin and Rajka's criteria is shown in Table 1.4. Five out of the six UK diagnostic criteria are included as major features in a later refinement of the Hanifin and Rajka criteria (Hanifin, 1992b) which is a tribute to their original proposal. The

Table 1.4. The UK refinement of the Hanifin and Rajka diagnostic criteria for atopic dermatitis

In order to qualify as a case of atopic dermatitis with the UK diagnostic criteria, the child must have:
An itchy skin condition in the last 12 months
Plus three or more of:
1. Onset below the age of two*
2. History of flexural involvement
3. History of a generally dry skin
4. Personal history of other atopic disease†
5. Visible flexural dermatitis as per photographic protocol

*Not used in children under four years of age.
†In children aged under four years, history of atopic disease in a first-degree relative may be included.

exact wording of the questions is to be found in a manual that has been developed by the author for field studies (Williams, 1997c). This manual also contains a set of training photographs and a set of quality control photographs which can be checked centrally.

Performance

The UK criteria have performed well in subsequent independent hospital and community validation studies (Williams et al., 1994c, 1996b; Popescu et al., 1998). In a validation study of children attending hospital dermatology outpatients, the criteria were shown to have a sensitivity and specificity of 85% and 96%, respectively, when compared with a dermatologist's diagnosis (Williams et al., 1994c). When used as a one-year period prevalence measure in a community survey of London children aged 3–11 years of mixed ethnic groups where the prevalence of AD was approximately 10%, sensitivity and specificity were 80% and 97%, respectively (Williams et al., 1996b). Positive and negative predictive values in this survey were 80% and 97%, respectively. In an identical community validation study of 1114 Romanian schoolchildren (Popescu et al., 1998) the sensitivity and specificity of the criteria were 74% and 99%, respectively, when tested against the dermatologist's diagnosis. Acceptable repeatability has been demonstrated for the six features contained within the UK criteria (Williams et al., 1994b). The criteria appear to be equally applicable to children of different ethnic and socioeconomic groups. They have worked well in children down to the age of one year, but further evaluation in infants and adults needs to be done. The criteria are easy to ascertain (taking under two minutes per person, including examination for flexural dermatitis), and they have proven to be highly acceptable to children and adults because of their relatively simple and noninvasive nature. They correspond well to our clinical concept of atopic dermatitis in that they contain all of the key elements that previous researchers have emphasized. Several groups studying allergic diseases have used the UK criteria without any major problems. Further validation studies of these criteria in developing countries are currently underway.

The concept of using different versions of the same criteria

The idea that several versions of the UK refinement of Hanifin and Rajka's diagnostic criteria may be used at any one time to define atopic dermatitis may seem odd at first, considering that one of the driving forces to develop diagnostic criteria is to obtain one overall standardized definition so that groups of people can be compared. However, it should be pointed out that different studies have different requirements of their definitions, especially in terms of the relative importance of sensitivity and specificity. Suggestions for the most appropriate format of features for diagnosing atopic dermatitis in various study scenarios are given below and are discussed in more detail elsewhere (Williams, 1997c).

Simple prevalence survey to assess the burden of disease

Itch, plus three or more of the features shown in Table 1.4, is used as a one-year period prevalence measure to overcome potential problems with seasonal fluctuations. It will be noted that since the presence of an itchy skin condition is the sole necessary criterion, then only subjects responding affirmatively to this question need to be examined further for evidence of visible flexural dermatitis. Such a strategy might save considerable expense and time, although some researchers may also wish to examine a sample of those without a history of an itchy rash to assess the proportion of false negatives. If examining individuals is out of the question, then the questions-only version (itch plus two or more of the remaining four features in Table 1.4) should be used, or a single compound question that has been widely used in the International Study of Asthma and Allergies in Childhood (Williams et al., 1999). A

similar criteria format can be used for comparative prevalence surveys.

Mixed asthma–hay fever–atopic dermatitis surveys

One of the six criteria for diagnosing AD is a personal history of hay fever or asthma. It is possible that the inclusion of asthma–hay fever within the definition of AD may be undesirable in some surveys which wish to keep the elements of the three allergic diseases entirely separate. If the inclusion of asthma–hay fever as part of the diagnosis of AD is unacceptable to an investigator, it is recommended that AD should be defined in terms of single unambiguous items, such as history of flexural itchy rash or visible flexural dermatitis.

Case-control studies

Since only a fraction of cases are sampled in most case-control studies, more specific, less sensitive criteria formats, such as itch plus four or more features, might be more suitable to minimize the inclusion of false positive cases. Similarly, very specific definitions of AD should be used in economic assessments of AD cases, as the inclusion of costs for noncases would be very misleading.

Cohort studies

Measurement of disease incidence poses difficulties as AD is usually an intermittent disease. The criteria proposed to date all refer to prevalent cases since they all contain elements of past disease or chronicity. In a cohort study, it might be appropriate to use the one-year period prevalences as a measure of disease incidence if they are recorded annually. Alternatively, an incident case of AD could be defined as any person who develops an itchy skin condition for the first time which is also compatible with AD (i.e. visible flexural dermatitis with modifications for young infants as outlined in the protocol). It is also possible to use the criteria to measure lifetime prevalence of AD by using 'has your child *ever* had an itchy skin condition?' for question 1 of the questionnaire shown in Table 1.4.

Hospital-based studies of AD subjects

There is no reason why the preferred criteria format of itchy skin plus three or more features could not be used in hospital studies which seek to recruit a representative population of AD cases. This would permit the selection of cases who are not necessarily active at the time of recruitment, and this might reduce the tendency to record epiphenomena associated with disease activity and severity.

Clinical trials

These would probably require subjects to have active disease on entry into a trial. Because cases referred to hospital are usually quite severe, it is likely that all hospital-ascertained cases of AD will have active dermatitis, and the normal criteria format of itch plus three or more features could be used.

Definite atopy

Some laboratory-based studies or clinical trials might require a stricter definition of the use of the word 'atopy' when defining subjects with atopic dermatitis. In such a context atopy, as defined by a positive skin prick test reaction (or raised allergen specific IgE) to one or more common environmental allergens, could be included as an additional necessary criterion for all cases.

As a diagnostic aid in the primary care setting

Although the UK refinements of Hanifin and Rajka's criteria were primarily designed for use in population surveys, they may be useful to family practitioners who wish to describe groups of subjects for audit studies. They may also be useful as a diagnostic aid to those less familiar with AD, but care must be taken in not interpreting failure to fulfil the diagnostic cri-

teria as proof of excluding AD, as opposed to the correct interpretation that AD is not *probable* within a certain degree of confidence at that moment. The use of likelihood ratios may help in this respect and these are discussed further in the manual (Williams, 1997c).

Thus, for one disease, there may be a range of definitions with slightly different validity indices, each of which may be better suited to specific study designs or requirements. They all define the *same* disease, but with differing precision and practical suitablility for different study designs and constraints.

It is important for the researcher to appreciate one further point which might influence the way in which the data are recorded and coded. It would be naïve to expect that the preferred format of itch plus three or more features could not be replaced by better criteria in the light of future discoveries on disease aetiology. Whatever new disease definitions emerge, the separate elements which make up the UK diagnostic criteria (e.g. history of flexural itchy rash) are still likely to be useful in describing the AD phenotype in future studies, especially for investigating secular trends and international comparative prevalence estimates. For this reason, in addition to composite measures such as atopic dermatitis 'yes/no', it is strongly recommended that the subjects' responses to individual criteria are retained separately on file.

Table 1.5. The relationship between true prevalence of AD, positive predictive value, prevalence of AD by UK criteria and systematic error using the validity indices (sensitivity 74% and specificity 99%) derived from a validation study in Bucharest

True prevalence of atopic dermatitis (%)	Positive predictive value (%)	Prevalence of AD by UK criteria (%)	Systematic error
1	34	1.8	1.8
2	58	2.6	1.3
3	67	3.3	1.1
4	73	4.0	1.0
5	79	4.7	0.9
6	81	5.4	0.9
7	84	6.2	0.9
8	86	6.9	0.9
9	87	7.7	0.9
10	88	8.4	0.8
11	89	9.1	0.8
12	90	9.9	0.8
13	91	10.6	0.8
14	91	11.4	0.8
15	92	12	0.8
16	92	12.7	0.8
17	93	13.5	0.8
18	94	14.2	0.8
19	94	15.0	0.8
20	94	15.7	0.8

Problem areas with the UK refinement of Hanifin and Rajka's criteria

Misclassification

As with most diagnostic tests, some degree of misclassification is inevitable. The gains and losses conferred by different sensitivities and specificities will depend very much on the nature of the study to which the criteria are applied. It is anticipated that one of the most common epidemiological uses for the proposed diagnostic criteria will be to compare prevalence rates between countries or in the same population at different points in time. Even though the sensitivity and specificity of the UK criteria for AD appear quite high, the underlying prevalence of AD has a critical influence on the positive predictive value (proportion of all those who fulfil the criteria and who are genuine cases) as shown in Table 1.5.

The effects of misclassification error in comparative prevalence studies can be examined directly by calculating the error for prevalence differences likely to be encountered in such studies. Take, for example, two populations of 1000 people in two different countries A and B. Suppose that the one-year period prevalence of AD in country A is 20% (200/1000) and in country B it is 10% (100/1000). Thus, the relative

risk of AD is twice as high in country A when compared with country B (95% confidence intervals 1.6 to 2.5, χ^2 for difference between the two proportions 38.4, $p<0.001$). If it is assumed that the sensitivity and specificity of our criteria for AD are 80% and 97%, respectively, when applied to these countries, then the prevelance of AD in country A will become 18.4% compared with 10.7% in country B. This represents a slightly lower relative risk of 1.72 (95% confidence interval of 1.38 to 2.15) of AD in country A when compared with country B, and a fall in the χ^2 value from 38.4 to 23.2 (still highly statistically significant). This reduction in the risk estimate towards unity is to be expected with such nondifferential misclassification, but this example illustrates how the new criteria are unlikely to obscure the true prevalence differences of the magnitude specified in these two populations. If the true prevalence of AD in countries A and B is 10% and 5%, respectively, then the estimated prevalence with the new criteria will be 10.7% and 6.8% in countries A and B, respectively. This difference is still highly significant ($p=0.003$), but the χ^2 value has fallen from 17.3 to 9.1.

Difficulties are likely to be encountered when very low disease prevalences occur. Although low prevalences for AD are unlikely in temperate climates, they could occur in tropical developing countries. Thus, a true prevalence difference of 5% and 2.5% in two countries may be obscured when the diagnostic criteria are applied. In addition, since the positive predictive values in these scenarios is more dependent on specificity, more specific alternative criteria formats, such as itch plus four or more features, could be used in populations with low disease prevalence.

Conversely, if significant differences are shown in a study where the new criteria are applied, then nondifferential misclassification is unlikely to have been a serious problem. Thus, in our community study of 695 schoolchildren in West Lambeth (Williams et al., 1995b), it was found that the prevalence of atopic dermatitis based on examination by a dermatologist was almost twice as high (16.3%) in Black children as in White children (8.7%, $p=0.03$). When the UK criteria for atopic dermatitis were used, almost identical findings were observed, suggesting that nondifferential misclassification introduced by our criteria did not obscure important ethnic group variations in the prevalence of AD, even in a study as small as this.

For prevalence surveys which simply wish to assess accurately the total burden of disease caused by AD, it may be imperative to establish the right count of cases. The degree to which the total number of true cases differs from the total number ascertained by the criteria can be expressed by a ratio called the 'systematic error', which refers to the ratio of the total number of cases positive to the survey and the reference tests. In a hypothetical prevalence survey of a community of 1000 children where, say, 100 children (10%) have had AD in the last year then, using the newly proposed criteria, 107 children will be described as having AD, representing a very similar prevalence rate (10.7%) and a very low systematic error of 1.07 when applied to these absolute counts. Although on an individual basis, a physician might take great exception to one child with keratosis pilaris being classified as atopic dermatitis in such a study, in epidemiological terms such misclassification is less serious when comparing populations. The relationship between systematic error, predictive values and prevalence of atopic dermatitis is shown in Table 1.5, assuming a sensitivity of 80% and specificity of 97% for the criteria. It can be seen that the systematic error is lowest when the prevalence of AD is around 10 to 12%, which fortunately also happens to be the most likely value of AD prevalence in most modern studies. True prevalence is underestimated by the criteria when the true prevalence is over 15%, and overestimated under this value. Systematic error is unacceptable when the true prevalence of AD is very low (<4%).

Even in a clinical situation where, for example, the diagnostic criteria are used by family practitioners as a diagnostic guide, the effects of misclassification of cases do not seem too serious. As our validation study showed (Williams et al., 1996b), nearly all of the false negative cases were inactive or asymptomatic cases of AD, and the consequences of a family practitioner not treating these cases with emollients

or very mild topical corticosteroids is unlikely to cause problems. Most of the false positive cases were mild forms of AD who were considered inactive by the validator. Even the nonAD false positive cases were composed mainly of other mild inflammatory dermatoses such as keratosis pilaris, frictional lichenoid dermatosis and pityriasis alba, all of which have been considered as being possible variants of AD, and all of which are treated using a similar therapeutic approach to mild AD. Thus at an operational level, the consequences of misclassification produced by the criteria in a clinical setting seem quite minor.

Asymptomatic disease

It is possible to derive diagnostic criteria in a systematic manner which are clinically meaningless and bear little resemblance to our clinical concept of atopic dermatitis. The UK Atopic Dermatitis Working Party believes that what is defined as a case of AD by the UK criteria is an entity which is worthy of measurement, reflecting a 'typical average case of AD'. That the criteria may miss the occasional mild cases of inactive AD is not a bad thing as most population surveys are interested in quantifying the distribution of *clinically important* as opposed to subclinical disease. Further studies examining the severity threshold of cases defined by the criteria, such as by measuring disability (Finlay & Khan, 1994), may be useful in this respect. It would be very easy to define 'definite AD' by making the diagnostic criteria more stringent, but this might exclude many cases of mild yet symptomatic diseases, which may form the bulk of community cases. It is recommended that researchers therefore stick to a binary as opposed to ordered categorical classification, and accept that the dividing line will always represent some form of compromise between excluding other itchy dermatoses and asymptomatic cases.

Adults and infants

Some modification of diagnostic criteria for AD may be necessary for young children (Williams et al., 1994c). Since early onset and inhalant allergy are likely to be less useful in the young child, early onset are omitted, and personal history of atopic disease is replaced by family history of atopic disease for children under four years of age. This scheme resulted in a sensitivity of 85% and specificity of 96% when applied to the 38 children aged four years and under in the hospital paediatric validation study (Williams et al., 1994c).

The author is less confident about the validity of the UK criteria in the first year of life. Although most children at this age were correctly classified, numbers were small, and validation against physician's diagnosis is perhaps not so useful due to the widespread disagreement between experts of what constitutes a case of AD at this age (Yates, Kerr & MacKie, 1983). One approach of describing AD in the first year of life is simply to focus on recording the prevalence of symptoms such as itchy rash or signs such as flexural dermatitis, analogous to similar conventions in asthma research. Further research using longitudinal designs might then delineate which features best discriminate those individuals who later develop typical AD, although it does not overcome the problem of saying that those who do clear did not originally have AD.

Based on our own data and other studies which have tried to separate seborrhoeic dermatitis of infancy from AD (Yates et al., 1983), we suggest that in order to be classified as a case of AD, children under the age of one year should have a history of scratching or rubbing plus three or more of: history of involvement of outer arms or legs, family history of atopic disease in first degree relatives, history of a general dry skin, and visible dermatitis on the cheeks or outer arms or legs with absence of axillary involvement.

Although the effect of potential confounding by age on the usefulness of the six diagnostic criteria has been thoroughly explored in the development work (*a*) by performing separate regression analyses in children and adults, (*b*) by looking for interaction between the criteria and a dummy variable for 'age under 16', and (*c*) by separate analysis of adult and paediatric data in the hospital validation study,

further testing of the performance of the criteria in adults in a community setting is still desirable because of the low numbers of adults with AD in our study.

Applicability to other ethnic groups

Although clinical experience suggests that AD in Black children can appear very different in its propensity to follicular lichenification, extensor involvement, later onset and lower frequency of personal or family history of atopy, the UK diagnostic criteria have performed well in Afro-Caribbean children in our community validation study in London and in another study in Jamaica (Williams et al., 1995b; Burrell-Morris et al., 1997). Another study has found that the physical features of atopic dermatitis in Black children are very similar to those in Whites (Macharia, Anabwani & Owili, 1993). It is important to appreciate that some cultures may not have a direct translation of the word 'itching', although recognition of the word *scratching* is universal – hence its inclusion in the wording of the major diagnostic criterion.

Use of the criteria in other communities

In countries where there is a high prevalence of other skin diseases which could be confused with atopic dermatitis, such as scabies or onchocerciasis, it seems prudent to stipulate that *the eruption must lack specific features of that dermatosis*. In studies of atopic dermatitis in areas where scabies is endemic, it would be wise to stipulate the addition of 'absence of burrows or finger web lesions' as a necessary major criterion in these circumstances. This presupposes that those conducting the examination are capable of identifying burrows, but if the disease is particularly prevalent, it is highly likely that nurses and health workers will be very familiar with the signs. In a study of scabies off the coast of Panama for instance, mothers of children were so adept at spotting scabies burrows and extracting mites that they were employed as survey helpers (Taplin et al., 1991).

It is also important that the criteria are tested in tropical countries where the appearance of a 'typical' case of AD may be altered by environmental factors such as UV light and infection. It should also not be assumed that some of the key words used in the criteria will retain their meaning when translated into other languages (Williams, 1999). Any translation of the questions should be performed by a person with good knowledge of local terms, such as a schoolteacher, and the translated version should be translated back into English by another independent person to ensure that the meaning of the questions is not grossly altered (Asher et al., 1995). Further validation studies of the UK refinement of Hanifin and Rajka's criteria for AD are currently in progress in China, Germany and India.

Where do other 'variants' of atopic dermatitis fit in?

When discussing the diagnostic criteria with other dermatologists, the author is often asked how other dermatological conditions which have been considered as possible variants of atopic dermatitis – such as dyshidrotic eczema, discoid eczema in children, juvenile plantar dermatosis and follicular and papular forms (Wüthrich, 1991) – fit in with the criteria. The answer is that the criteria will not provide an easy way of saying whether these conditions are truly variants of atopic dermatitis. The criteria derived in our studies were only designed to discriminate between typical mild to moderate AD and other inflammatory conditions, as opposed to determining the degree of 'atopic dermatitis' in other purported atopic dermatitis variants. Thus, it would be quite wrong to say that an individual with an unusual pattern of dermatitis 'definitely does not have atopic dermatitis' simply because he/she does not fulfil the UK criteria for AD. It would be accurate, however, to state that such an individual has a 97% probability of not having 'typical atopic dermatitis'. The author accepts that vesicles on the sides of the fingers in summer months in someone predisposed to atopy, discoid eczema in a child and late-onset eczematous erythroderma with high IgE levels in an

adult, may all be related to atopic dermatitis but their precise relationship to AD would require a special study such as cluster analysis or numerical taxonomy, preferably backed by genetic marker studies.

Dangers of suggesting a disease definition

It is possible that the mere proposal of a disease definition can create a spurious impression of understanding that disease which could stagnate further research into disease aetiology. This is unlikely to occur providing the limitations of our definition are recognized. The author feels that the advent of a definition for AD that is designed for use in population studies is likely to stimulate rather than stifle further studies of disease aetiology.

Will the criteria change again in five years?

One must consider the possibility that the newly proposed criteria will not last very long. This does not unduly concern the author, providing that they are replaced by something better. The author recognizes that the criteria as proposed today are but a transient step in the process of progressive nosology (Kendell, 1975). Thus if a more rational basis for the classification of atopic dermatitis is found in the next 20 years, then the current definition based on a clinical syndrome might well lose its value. As Kendell (1975) commented, 'to the contemporary medical research worker, if not to every practising clinician, diseases are little more than convenient working concepts based on a variety of different defining criteria, anatomical, physiological or behavioural, and liable to change their defining characteristics, or even to be abandoned altogether, with advances in knowledge'.

The emergence of many similar 'rival' criteria based on arbitrary arrangements of clinical features alone would not be useful, however, unless they were shown to produce marked benefits over the UK criteria when put to the test in independent validation studies. Even in the absence of 'rival' diagnostic criteria, it is likely that the criteria will produce a number of 'boundary disputes'. Difficulties in establishing international agreement of diagnostic criteria are more likely to be encountered when defining the boundary or outer rim of what separates a condition from other adjacent categories as opposed to agreeing on what constitutes the core of typical clinical features. Thus, one group might insist that all subjects with AD must be atopic as defined by objective tests of immediate hypersensitivity, or that visible 'eczematous' skin changes must be visible in all subjects. Providing these modifications are explicit, and that the individual elements of the diagnostic criteria are recorded separately, then most of these 'boundary disputes' are unlikely to be insurmountable. Indeed, the different nature of the many types of study designs available to researchers means that some modification of the criteria for the purposes of a particular study is inevitable.

It should be pointed out that even if a genetic basis for atopic dermatitis is discovered in the future, there will always be a need to provide an adequate description of the disease phenotype with well defined clinical criteria. The UK diagnostic criteria for atopic dermatitis may be useful for ensuring a degree of comparability of subjects in future epidemiological surveys. The author views the diagnostic criteria for atopic dermatitis as an evolving instrument, and welcomes modifications and improvements in the light of further knowledge.

Conclusions

This chapter has challenged the way we think about defining atopic dermatitis for epidemiological studies. Having stated the desirable properties of diagnostic criteria for epidemiological studies, the development of a refined set of simple criteria has been described. These appear to work well, but more testing is needed in other communities. The validity of these criteria in infants and in adults is still not known, and a satisfactory definition of an incident case of AD is still lacking.

Although other definitions will undoubtedly emerge for AD, the key issue in epidemiological studies is the need to compare results from many

studies from around the world. For this, *standardization* is of paramount importance. It is better to have a less than perfect disease definition of known validity and repeatability than a definition that is claimed to be better but which is of unknown validity.

The point of defining a disease is to improve our understanding of the disease and to improve our predictive abilities (Burton, 1981), whether this be in relation to its causes and distribution, natural history, biology or treatment. As described in this chapter, even within the sphere of epidemiology, there may be a range of requirements for disease definition with different degrees of precision, depending on the nature of the study. Disease definition in epidemiology should be viewed as an instrument or tool of known validity which is only important in relation to the biological or social question that is being addressed.

Summary of key points

- Atopic dermatitis (AD) should be viewed as a multidimensional phenomenon.
- It is unclear at present if *any* individual could develop AD under the right adverse circumstances, or whether genetic predisposition is a prerequisite for disease expression.
- Despite its limitations, a binary definition for AD is most readily understood by clinicians and epidemiologists throughout the world.
- Data-driven post hoc subcategories of AD are only useful if they are shown to increase our predictive ability on factors such as prognosis or responsiveness to treatment.
- Because many individuals with AD have normal IgE responsiveness, the word 'atopy' when used in the term 'atopic dermatitis' does not have a precise meaning.
- Other mechanisms such as altered nonspecific reactivity or T-lymphocyte dysfunction may be just as important in the pathogenesis of AD as IgE hyper-responsiveness.
- The terms 'atopic eczema' and 'atopic dermatitis' are synonymous.
- Although the term 'atopic dermatitis' may have a scientific and objective ring to it, it may not increase our predictive ability much more than the phrase 'itchy red rash in the skin folds'.
- Standardized diagnostic criteria are essential if valid comparisons are to be made between groups of people.
- Disease definitions have meaning only in context to the biological question which is being asked.
- A good disease definition is valid, repeatable, easy to use, applicable to a wide range of situations, acceptable to the population and contains elements that are coherent with prevailing clinical concepts.
- The Hanifin, Lobitz and Rajka diagnostic criteria represented a major milestone in summarizing the clinical concept of AD. However, because of their unknown validity and complexity, they are not suitable for epidemiological studies.
- A UK refinement of the Hanifin, Lobitz and Rajka diagnostic criteria has been developed for use in epidemiological studies. These criteria have good repeatability and contain elements which allow comparability with older studies.
- The UK criteria have a sensitivity of 80% and 74% and a specificity of 97% and 97% when tested against a dermatologist's diagnosis in population studies in the UK and Romania, respectively.
- Different arrangements of the UK criteria with different specificities may be used for different types of study requirements.
- Misclassification of disease is unlikely to obscure clinically important prevalence differences for disease prevalences above 5%.
- The validity of the UK diagnostic criteria in infants and in adults needs further study.
- There is currently no accepted method for defining an incident case of AD.
- It is better to have a less than perfect disease definition of known validity and repeatability than a definition that is claimed to be better with unknown validity.
- Disease definition is an evolutionary process which should be modified in the light of new knowledge.

- The possibility that what we recognize as the clinical syndrome of AD today will be subsequently shown to be caused by three or four different agents does not imply that older diagnostic criteria are 'wrong', providing they measure something useful or that they are instrumental in stimulating further research.

References

Abrams, P. (1994). New words for old: lower urinary tract symptoms for 'prostatism'. *Br Med J*, **308**, 929–30.

Ackerman, A.B. (1982). A plea to expunge the word 'eczema' from the lexicon of dermatology and dermatopathology. *Am J Dermatopath*, **4**, 315–26.

Asher, M.I., Keil, U., Anderson, H.R. et al. (1995). International study of asthma and allergies in childhood (ISAAC): rationale and methods. *Eur Resp J*, **8**, 483–91.

Bakke, P., Gulsvik, A. & Eide, G.E. (1990). Hay fever, eczema and urticaria in southwest Norway. *Allergy*, **45**, 515–22.

Barbee, R., Kaltenborn, W., Lebowitz, M. & Burrows, B. (1987). Longitudinal changes in allergen skin test reactivity in a community population sample. *J Allergy Clin Immunol*, **79**, 16–24.

Barker, D.J.P. & Rose, G. (1979). *Epidemiology in Medical Practice*, 2nd edn. Edinburgh: Churchill Livingstone.

Barrows, H.S., Norman, G.R., Neufield, V.R. & Feightner, J.W. (1982). The clinical reasoning of randomly selected physicians in general medical practice. *Clin Invest Med*, **5**, 49–55.

Bos, J.D., Wierenga, E.A., Smitt, J.H.S. et al. (1992). Immune dysregulation in atopic eczema. *Arch Dermatol*, **128**, 1509–12.

Burrell-Morris, C.E., LaGrenade, L., Williams, H.C. & Hay, R.J. (1997). The prevalence of atopic dermatitis in Black Caribbean children in London and Kingston, Jamaica. *Br J Dermatol*, **137** (Suppl. 50), 22.

Burton, J.L. (1981). The logic of dermatological diagnosis. *Clin Exp Dermatol*, **6**, 1–21.

Buser, K., von Bohlen, F., Werner, P. et al. (1993). The prevalence of neurodermatitis among schoolchildren in the Hannover administrative district. *Dtsh Med Wschr*, **118**, 1141–5.

Champion, R.H. & Parish, W.E. (1986). Atopic dermatatis. In: Rook, A.J., Wilkinson, D.S. & Ebling, F.J.G. (eds.) *Textbook of Dermatology*, 4th edn. Oxford: Blackwell Scientific.

Coca, A.F. & Cooke, R.A. (1923). On the classification of the phenomena of hypersensitiveness. *J Immunol*, **8**, 163–82.

Coombs, R.R.A. & Gell, P.G.H. (1963). The classification of allergic reactions underlying disease. In: Gell, P.G.H. & Coombs, R.R.A. (eds.) *Clinical Aspects of Immunology*. Philadelphia: Davis.

David, T.J. (1991). Conventional allergy tests. *Arch Dis Childh*, **66**, 281–2.

de Dombal, F.T., Horrocks, J.C., Staniland, J.R. & Guillou, P.J. (1972). Pattern recognition: a comparison of the performance of clinicians and non-clinicians – with a note on the performance of a computer-based system. *Methods Inf Med*, **11**, 32.

Diepgen, T.L. & Fartasch, M. (1991). Stigmata and signs of atopic eczema. In: Ring, J. & Pryzbilla, B. (eds.) *New Trends in Allergy*, 3rd edn. pp. 222–9. Berlin: Springer-Verlag.

Diepgen, T.L. & Fartasch, M. (1992). Recent epidemiological and genetic studies in atopic dermatitis. *Acta Derm Venereol (Stockh)*, Suppl. 176, 13–18.

Diepgen, T.L., Fartasch, M. & Hornstein, O.P. (1989). Evaluation and relevance of atopic basic and minor features in patients with atopic dermatitis and in the general population. *Acta Dermatol Venereol (Stockh)*, Suppl. 144, 50–54.

Diepgen, T.L., Sauerbrei, W. & Fartasch, M. (1994). Evaluation and validation of diagnostic models of atopic dermatitis. *J Invest Dermatol*, **102**, 619 (abstract).

Dotterud, L.K., Kvammen, B., Lund, E. & Falk, E.S. (1995). Prevalence and some clinical aspects of atopic dermatitis in the community of Sør-Varanger. *Acta Derm Venereol (Stockh)*, **75**, 50–3.

Edenharter, G., Bergmann, R.L., Bergmann, K.E. & Wahn, U. (1998). Definition of atopic dermatitis (AD) in infants at 5 years: combining atopic symptoms and signs with atopic biomarkers. *Proceedings of the EDEN Congress*, Bamberg, Germany, May 2–4, 1998.

Fartasch, M. & Diepgen, T.L. (1994). Atopic diathesis and occupational dermatoses. *J Invest Dermatol*, **102**, 620.

Finlay, A.Y. & Khan, G.K. (1994). The Dermatology Life Quality Index: a simple practical measure for routine clinical use. *Clin Exp Dermatol*, **19**, 210–16.

Freiman, J.A., Chalmers, T.C., Smith, H. & Kuebler, R.R. (1978). The importance of beta, the type 2 error and sample size in the design and interpretation of the randomized controlled trial. *N Engl J Med*, **299**, 690–4.

Gelmetti, C. (1992). Extracutaneous manifestations of atopic dermatitis. *Ped Dermatol*, **9**, 380–2.

Gerber, L.M., Wolf, A.M., Braham, R.L. & Alderman, M.H. (1982). Effects of sample selection on the coincidence of hypertension and diabetes. *J Am Med Ass*, **247**, 43–6.

Hanifin, J.M. (1983). Clinical and basic aspects of atopic dermatitis. *Sem Dermatol*, **2**, 5–19.

Hanifin, J.M. (1992a). Atopic dermatitis. In: Marks, R. (ed.) *Eczema*, pp. 77–101. London: Martin Dunitz.

Hanifin, J.M. (1992b). Atopic dermatitis. In: Moschella, S.L. & Hundey, H.J. (eds.) *Dermatology*, 3rd edn., pp. 441–64. Philadelphia: Saunders.