DR TOSHINORI NAKAMURA (Orcid ID : 0000-0002-5176-8277)

Article type : Original Article

Different definitions of atopic dermatitis: Impact on prevalence estimates and associated risk factors

Running head: Impact of Different definitions of Atopic dermatitis

T. Nakamura*[1], S. Haider*[1], S. Colicino[2], C.S. Murray[3], J. Holloway[4], A. Simpson[3#], P. Cullinan[2#], A. Custovic[1#]

on behalf of STELAR[5] investigators

*Equal contribution

[#]Joint senior authors

[1]Department of Paediatrics, Imperial College London, UK

[2]National Heart and Lung Institute, Imperial College, London, UK

[3]Division of Infection, Immunity & Respiratory Medicine, University of Manchester, Manchester

[4]Human Development and Health, Faculty of Medicine, University of Southampton, Southampton

Correspondence and requests for reprints:

Professor Adnan Custovic, Section of Paediatrics, Department of Medicine, Imperial College London, UK.

Tel: +44 20 7594 3274; fax: +44 20 7594 3984; Email: a.custovic@imperial.ac.uk

Funding sources: STELAR cohorts are funded by The UK Medical Research Council (MRC) Grants G0601361 and MR/K002449/1.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/bjd.17853

Disclosure of potential conflict of interest: The authors declare that they have no conflict of interests

1. What is already known about this topic?

There is no objective test that can unequivocally confirm the diagnosis of atopic dermatitis (AD) and no uniform clinical definition. This results in different definitions utilised in AD studies, raising concerns on the generalisability of the results and comparability across different studies.

2. What does this article add to our knowledge?

This study has shown that different definitions of "Cases" and those of "Controls" have major impacts upon prevalence estimates and associations with risk factors, including genetics, in two population-based birth cohorts. These findings suggest the importance of developing a consensus on AD definitions of "Controls" as well as "Cases" to minimise biases in the studies.

ABSTRACT

Background: There is no objective test that can unequivocally confirm the diagnosis of atopic dermatitis (AD), and no uniform clinical definition.

Objective. To investigate to what extent operational definitions of AD cause fluctuation in the prevalence estimates and the associated risk factors.

Methods: We first reviewed operational definitions of AD used in the literature. We then tested the impact of the choice of the most common definitions of "Cases" and "Controls" on

AD prevalence estimates and associated risk factors (including *filaggrin-FLG* mutations) among children aged 5 years in two population-based birth cohorts: Manchester Asthma and Allergy Study (MAAS) and Asthma in Ashford. Model performance was measured by the percentage of children within an area of clinical indecision (defined as having a posterior probability of AD between 25% and 60%).

Results: We identified 59 different definitions of AD across 45 reviewed studies. Of those, we chose 4 common "Case" definitions, and 2 definitions of "Controls". The prevalence estimates using different case definitions ranged between 22% and 33% in MAAS, and 12% and 22% in Ashford. The area of clinical indecision ranged from 32% to 44% in MAAS, and from 9% to 29% in Ashford. Depending on the case definition used, the associations with *FLG* mutations varied (ORs [95% CI]: 1.8 [1.1-2.9] to 2.2 [1.3-3.7] (MAAS) and 1.7 [0.8-3.7] to 2.3 [1.2-4.5] (Ashford)). Associations with *FLG* mutations also differed when using the same "Case" definition, but different definitions of "Controls".

Conclusion: Use of different definitions of AD results in substantial difference in prevalence estimates, the performance of prediction models, and association with risk factors.

INTRODUCTION

Although atopic dermatitis (AD) is one of the most common skin diseases[1], there is no universally accepted definition of this condition for epidemiological and genetic studies[2], and no objective test that can unequivocally confirm the diagnosis[3]. Despite efforts to reach a consensus on nomenclature, two terms (AD and eczema) currently coexist to describe a clinically defined, pruritic, inflammatory skin condition, characterized by chronic and relapsing dermatitis in common anatomical sites[4], and are are often used interchangeably[5]. Further denominations such as atopic eczema/dermatitis syndrome (AEDS)[6] have also been proposed. Kantor *et al.* have shown that AD is currently the most commonly used term, but

that the term use differs between literature in different languages and scientific disciplines[5]. However, even when the same term (e.g. AD) is used in epidemiological[7] and genetic[8] studies, children are assigned as "Cases" and "Controls" using a variety of different definitions[7-10]. This may hinder the generalisability of the results and comparisons across different studies and geographical areas[7,11,12], and may impact on estimates of the magnitude of the effects of potential risk factors and on study conclusions. Such impact has been shown in asthma, in which the variation in the definition of the primary outcome had a considerable impact on the estimated prevalence and on results of prediction models[13].

We propose that research findings may differ substantially if different definitions of AD are used. Our aim was not to tackle which definition may be the most appropriate, but to investigate the potential consequence of using different definitions on the results of AD studies. As a first step, we reviewed the definitions of AD used in literature. We then tested the impact of the choice of the commonly used definitions of "Cases" and "Controls" on AD prevalence estimates, and associated risk factors (including *filaggrin-FLG* mutations[14,15]), among children aged five years in two UK birth cohorts.

METHODS

Definitions and operationalisations of AD

We reviewed the case definitions of AD in 26 studies included in a meta-analysis of genome-wide association studies (GWAS)[8] and 45 studies included in a systematic review of AD persistence[7]. More recent studies published between 2015 and 2017 were also included through a MEDLINE search, using PubMed. Studies which fulfilled the following criteria were included: 1) Prospective cohort design; 2) AD as the primary or secondary outcome; 3) Participants aged between 0 and 18 years; 4) Published in English. We extracted the

following information: 1) Definition of AD; and 2) Data sources used to diagnose AD (questionnaire, physical examination or medical records).

As some definitions consisted of a combination of several data-sources (e.g., both questionnaires and physical examination as in "Parent-reported AD confirmed by physical examination"), we decomposed those data-sources for each case definition (Figure E1). Questionnaire-based definitions were further categorised to either "Physician-confirmed AD", or "Parent-reported AD". As many of the questionnaire-based definitions utilised several clinical features of AD, such as types of symptoms or treatment used, definitions were further categorised as "No specific features", "Chronic skin condition", "Itchy skin condition", "Skin condition affecting skin creases", "Treatment", and "Other" (e.g. age of onset). The definition of "Control" included children who did not fulfil the case definitions, unless studies explicitly stated the definition.

Prevalence estimates and associated risk factors using different definitions

For the analysis of the impact of different AD "Case" definitions, we applied four commonly used definitions of current AD identified in the literature review (Table 1) to the data from two population-based birth cohorts: Manchester Asthma and Allergy Study (MAAS)[16] and Ashford cohort[17] from the UK STELAR consortium[18]. Detailed description of the cohorts is provided in the Online Repository. Both studies were approved by local research ethics committees. Written informed consent was obtained from all parents. For this analysis, we used data collected at review clinics at comparable follow-up age of five years. Validated questionnaires were interviewer-administered to collect information on parentally-reported symptoms, physician-diagnosed illnesses and medication usage. We assessed allergic sensitization by skin prick tests (SPT).[19] Genotyping was performed for two *FLG* mutations

(online supplement), and children with *FLG* loss-of-function were defined as those with either non-sense mutation R501X or frame-shift mutation 2282del4[14,20].

In the prediction modelling, we used the following set of established predictors of AD: *FLG* genotype, parental AD, allergic sensitization (age 5), and physician-confirmed asthma (age 5) (for definitions, see Online Supplement).

Statistical methods

First, we compared prevalence estimates for the four different "Case" definitions. We then used bivariate logistic regression analysis to assess the impact of the four AD "Case" and the two "Control" definitions on associations with *FLG* mutations and other risk factors. Finally, we constructed prediction models using multivariable logistic regression analysis and assessed the patterns of distributions of the posterior probabilities and the performance of prediction models following the study of van Wonderen *et al.*[13]. Performance was measured using the percentage of children whose posterior probability was in an area of clinical indecision (25-60%)[13], assuming that a posterior probability of 25% or less predicts a low risk of the disease and a posterior probability above 60% indicates a high risk. A sensitivity analysis was also undertaken by comparing the area of clinical indecision between 25% and 50%. The analyses of prediction models were conducted in children with complete data for the included variables. We used STATA 14.2 for all analyses (StataCorp, College Station, USA).

RESULTS

Search for definitions of AD in the literature

We reviewed 45 studies (Figure E2) and identified 59 different operational definitions of AD (summarised in Table E1). A total of 32 studies included a cumulative estimate of AD

(lifetime period), 26 used current AD (defined as the presence of AD in the previous 6, 12 or 24 months), and no time-period was specified in one study. Within each definition, there was further heterogeneity (for example, within the category of physician-confirmed AD for cumulative prevalence we found six different definitions, Table E1). After definitions which consisted of a combination of several data-sources were decomposed, further heterogeneity became apparent (e.g. 31 were derived from a single data-source, 24 from two, and 4 from three or more). Of these, 41 definitions were based on physician-confirmed AD, 43 on parent-reported AD, 7 on physical examination, and two on data from medical records.

Of the 59 operational definitions, 27 were derived based on questions referring to an "itchy skin condition", 23 on "skin condition affecting skin creases", and 17 on "chronic skin condition". Of the 43 case definitions which included "Parent-reported AD", 27 (63%) incorporated at least one of these three common features. Of these, 11 adopted all three features (Figure E3). Of 41 definitions which included "Physician-confirmed AD", 33 relied on a single or several questions pertaining to physician diagnosis (Figure E3). Only seven definitions incorporated the use of treatment, and the age of onset was considered in four.

We then chose four common operational case definitions (Table 1) to estimate prevalence, risk factors and predictive performance of prediction models in the two cohorts: *Definition 1*. Physician-confirmed AD; *Definition 2*. Physician-confirmed AD and parent-reported chronic itchy skin condition affecting skin creases; *Definition 3.* Parent-reported chronic itchy skin condition affecting skin creases; *Definition 4.* Physician-confirmed AD or parent-reported chronic itchy skin condition affecting skin creases. For these analyses, "Controls" were defined as children who did not fulfil the case definition.

Prevalence estimates, associates, and prediction model performance

We used data from 1069 children in MAAS and 604 in Ashford, of whom 771 (MAAS) and 405 (Ashford) had a complete data set. Table 2 shows the characteristics of children included in the analysis. Caucasian children accounted for 95% of the sample in MAAS and 99% in Ashford. *FLG* mutations were present in one-tenth of the children.

The Venn diagram in Figure 1shows that the prevalence was highest using Definition 4 (30% [95% CI 27-33] and 22% [18-25]), and lowest using Definition 2 (22% [19-24] and 12% [9-15], MAAS and Ashford respectively), mean difference [95% CI]: 8% [5-12], $p<0.001$ in MAAS, 10% [5-13], $p<0.001$ in Ashford. The prevalence estimates of AD were similar using Definitions 1 and 3 (25% [22-27] and 27% [25-30] in MAAS and 19% [15-22] and 15% [12-17] in Ashford). Transition between Cases and Controls in each definition is shown in Table E2: for example, among children assigned as Cases in Definition 4, 27% (MAAS) and 43% (Ashford) were assigned as Controls in Definition 2; Among those assigned as Cases in Definition 1, 12% (MAAS) and 36% (Ashford) were assigned as Controls in Definition 3.

The strength of the association with *FLG* genotype among Caucasian children differed between different definitions in both cohorts (ORs [95%CI]: 1.7 [1.0-2.8] to 2.1 [1.3-3.6], MAAS; and 1.7 [0.8-3.7] to 2.3 [1.2-4.5], Ashford, Table 3). Association with other risk factors is shown in Table E3.

Performance of prediction models

Figure 2 shows the distributions of posterior probabilities of the prediction models of current AD for the four definitions of "Cases". In both cohorts, the distribution of the probabilities varied depending on the definition. A consistent finding was that the posterior probability in Definition 2 was skewed to the lowest, and those in Definition 4 was skewed to the highest.

The percentages of children whose posterior probability was in the area of clinical indecision was the lowest in Definition 2 (32% in MAAS and 9% in Ashford) and the highest in Definition 4 (44% and 29%). Hence, in both cohorts, the prediction models had the best performance in Definition 2 and the worst performance in Definition 4.

The effect of different definitions of "Controls"

We then proceeded to ascertain the effect of different definitions of "Controls" on the pattern of the association with risk factors. From the literature search, we extracted two definitions of "Control" which comprised of the combination of responses to several questions ("Strict" and "Moderate", Table E4). Using the "Strict" control definition, 186 (18%) children in MAAS and 135 (22%) in Ashford were unclassifiable (i.e., could not be assigned to either the "case" or "control" due to a positive answer to one of the questions we used). The patterns of responses to three questions among "unclassifiable" children are shown in Table E5.

The associations of AD (using Definition 4) with *FLG* mutations were stronger when we used the "strict" Control definition (ORs [95%CI] 2.4 [1.5-4.0] and 2.2 [1.1–4.6]) than the "moderate" (ORs [95%CI] 1.8 [1.1-2.9] and 1.9 [0.99–3.8], MAAS and Ashford respectively, Table 4). We observed a significant association between the "unclassifiable" group with *FLG* mutations which was of a similar magnitude as that for the cases in MAAS (ORs [95%CI] 2.5 [1.3-4.7]), but not in Ashford (1.4 [0.7–3.2], Table E6). The association with other risk factors is shown in Table E7. In both cohorts, associations with sensitisation and asthma were stronger when we used the "strict" Control definition.

As the choice of control definition may have implications for sample size and power, we calculated the power for detecting an association between AD and *FLG* genotype using the strict and moderate control definitions in MAAS. Although the sample size reduced by

approximately one-fifth when moving from the moderate to strict definition, the power increased by ~50% from 0.58 to 0.85 by having a "purer" control as a comparator for AD cases. Consequently, there was a larger effect size using the strict version compared with the moderate one (Table 4).

DISCUSSION

We have described numerous different definitions of AD which have been used in epidemiological and genetic studies. By applying common definitions to two population-based birth cohorts, a consistent finding was that the use of different definitions of both cases and controls resulted in substantial differences in the prevalence estimates, the performance of prediction models, and the association with risk factors.

One limitation of this study is that our literature review was not systematic, hence relevant studies may have been missed. However, we reviewed studies encompassed within recent meta-analyses of AD persistence[7] and GWAS[8], and our results may contribute to a discussion about the extent to which the variability in the results of these studies arose from differences in the definition of primary outcome.

We assessed the impact of questionnaire-based definitions using three questions regarding AD features, but the questions were not identical in the two cohorts. This may account for some of the differences in findings between our cohorts. We acknowledge that physical examination may offer a more accurate way of defining AD[7]. The U.K. Working Party's[21-23] and Hanifin and Rajka[24] diagnostic criteria are excellent for case definition in case/patient studies, but are difficult to fully implement in large-scale epidemiological studies which are mostly questionnaire-based. Information from physical examination available in birth cohorts is usually available at only a few time points during the clinical follow-up (e.g. once every 2-

3 years). Given the temporal variability of AD symptoms, using this information would likely introduce bias towards more severe disease. However, it is of note that in any of the data-sources, there are currently no uniform definitions[25,26], and the variation of outcomes in observational studies of AD may well be more extensive than the findings reported in this study.

Further limitation is that we assessed children at age 5 years, and cannot infer that different definitions have similar impact in other age groups. A study which investigated the association between AD and cardiovascular disease in adults reported a poor agreement between questionnaire-based diagnostic criteria, thus hindering consistent conclusions about associations[27].

We have not taken into account the temporal pattern of AD during childhood. Identification of the individual trajectories over the life-course may contribute to understanding the disease heterogeneity[28], and latent class analysis has recently been used to assign children to different AD phenotypes based on longitudinal patterns of flexural rash[29,30]. It would be important to know how the different disease definitions impact on the identification of AD trajectories, but such analyses were beyond the scope of the current study.

We did not include all identified *FLG* mutations. However, we have previously shown in MAAS that there were no differences in results when FLG loss-of-function was defined using R501X and 2282del4, compared to using six mutations (R501X, S3247X, R2447X, 2282del4, 3673delC and 3702delG)[31].

When comparing the results of different cohorts, it is necessary to consider the study regions[32], languages[33], and the age of the subjects[34,35 36] as confounders affecting the prevalence of AD. We have carried out our analyses in two birth cohorts from the same

geographical area, which used similar questionnaires administrated at the same age. As a result, we anticipate the effect of these confounders to be minimal.

We confirmed a wide variety of definitions for AD in the literature. The most commonly used definition was questionnaire-reported physician-confirmed AD (our Definition 1). The second most common definition used three important features of AD, namely "itchy skin condition", "skin condition affecting skin creases", and "chronic skin condition" (our Definition 3), which may be influenced by the International Study of Asthma and Allergies in Childhood (ISAAC) core questionnaire[37]. The ISAAC questionnaire was established in 1995 to enhance the comparability of epidemiological research in asthma and allergic diseases[37]. However, our findings demonstrate that although many studies adopted the ISAAC questionnaire, a variety of definitions have been used (e.g. using questions on chronic itchy skin condition, but not the distribution affecting skin creases[38,39]). Williams *et al* cautioned that such modifications may result in a decrease in the specificity of the diagnosis[22].

FLG mutations are one of the most robust genetic risk factors for AD[15,40], but a number of factors can mediate this relationship, including race and age[34]. The heterogeneous patterns of associations with *FLG* mutations in our study populations, which are ethnically homogenous and assessed at the same age, indicate that different case/control definitions may have adverse impact on our understanding of the underlying pathophysiological mechanisms. We observed in both cohorts that some definitions (such as Definition 2) had stronger associations with *FLG* mutations than others. This definition included both physician-confirmed AD and parent-reported chronic itchy skin condition affecting skin creases. In addition, the prediction models had the best performance for Definition 2, with the lowest percentage of the area of clinical indecision. An implication of this is that a standardised definition of AD should capture multiple domains of the disease, including severity. Furthermore, the comparison between the "strict" and "moderate" control definitions demonstrated that the association of

AD with *FLG* mutations was stronger if the "strict" definition was used. When we used the "strict" definition, a fifth of children were unclassifiable (and thus eliminated from the analyses). However, despite this reduction in sample size, the power of the study to detect significant associations increased by ~50%, and with a larger effect size. It is of note that even though the choice of the definition of "Controls" for the analyses of genetic and environmental risk factors clearly influenced the study outcomes, in practice, of 28 studies utilising multiple case definitions, only 7 (25%) reported the definitions for the "Controls" expressly.

Given a significant association of the "unclassifiable" group with *FLG* loss-of-function mutations, some of these children are likely to have mild AD, or other condition such as ichthyosis. Some participants with *FLG* null mutations have fallen in the "unclassifiable" group because even though they were asymptomatic at age five years, a doctor had diagnosed AD in their infancy. This is consistent with a finding that the average duration of AD persistence in individuals with *FLG* mutation was 77 months[41].

Our findings suggest that large questionnaire-based studies, in which the primary outcome is usually defined using the lowest common denominator, may not be the most informative, and that it may be time to move on to clinical diagnosis. The international Harmonising Outcome Measures for Eczema (HOME) initiative suggested the use of a minimum standard of core features, such as clinical signs, symptoms, long-term control, and quality of life (QOL), for clinical trials[42], and similar approach is needed for epidemiological and genetic studies. In conclusion, there is a pressing need to develop a uniform definition for "Cases" and "Controls" of AD for epidemiology using a set of harmonised outcomes which comprise multidimensional information to facilitate comparison of study findings, better understanding of the AD heterogeneity, and minimise biases arising from the choice of definitions.

REFERENCES

1. Wollenberg A, Oranje A, Deleuran M, et al. ETFAD/EADV Eczema task force 2015 position paper on diagnosis and treatment of atopic dermatitis in adult and paediatric patients. *Journal of the European Academy of Dermatology and Venereology* 2016; **30**(5): 729-47.
2. Bieber T. Why we need a harmonized name for atopic dermatitis/atopic eczema/eczema! *Allergy* 2016; **71**(10): 1379-80.
3. Bieber T, D'Erme AM, Akdis CA, et al. Clinical phenotypes and endophenotypes of atopic dermatitis: Where are we, and where should we go? *Journal of Allergy and Clinical Immunology* 2017; **139**(4): S58-S64.
4. Silverberg JI, Thyssen JP, Paller AS, et al. What's in a name? Atopic dermatitis or atopic eczema, but not eczema alone. *Allergy* 2017; **72**(12): 2026-30.
5. Kantor R, Thyssen JP, Paller AS, Silverberg JI. Atopic dermatitis, atopic eczema, or eczema? A systematic review, meta-analysis, and recommendation for uniform use of 'atopic dermatitis'. *Allergy* 2016; **71**(10): 1480-5.
6. 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): 813-24.
7. Kim JP, Chao LX, Simpson EL, Silverberg JI. Persistence of atopic dermatitis (AD): A systematic review and meta-analysis. *Journal of the American Academy of Dermatology* 2016; **75**(4): 681-7.e11.
8. Paternoster L, Standl M, Chen C-m, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. *Nature genetics* 2011; **44**(2): 187-92.
9. Simpson EL, Keck LE, Chalmers JR, Williams HC. How should an incident case of atopic dermatitis be defined? A systematic review of primary prevention studies. *The Journal of allergy and clinical immunology* 2012; **130**(1): 137-44.
10. Vakharia PP, Chopra R, Silverberg JI. Systematic Review of Diagnostic Criteria Used in Atopic Dermatitis Randomized Controlled Trials. *American Journal of Clinical Dermatology* 2017.
11. Brenninkmeijer EEA, Schram ME, Leeflang MMG, Bos JD, Spuls PI. Diagnostic criteria for atopic dermatitis: a systematic review. *The British journal of dermatology* 2008; **158**(4): 754-65.
12. 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. *British Journal of Dermatology* 2005; **152**(4): 735-41.
13. Van Wonderen KE, Van Der Mark LB, Mohrs J, Bindels PJE, Van Aalderen WMC, Ter Riet G. Different definitions in childhood asthma: How dependable is the dependent variable? *European Respiratory Journal* 2010; **36**(1): 48-56.
14. Irvine AD, McLean WHI, Leung DYM. Filaggrin Mutations Associated with Skin and Allergic Diseases. *New England Journal of Medicine* 2011; **365**(14): 1315-27.
15. Rodríguez E, Baurecht H, Herberich E, et al. Meta-analysis of filaggrin polymorphisms in eczema and asthma: Robust risk factors in atopic disease. *Journal of Allergy and Clinical Immunology* 2009; **123**(6): 1361-70.e7.
16. Custovic A, Simpson BM, Murray CS, Lowe L, Woodcock A. The National Asthma Campaign Manchester Asthma and Allergy Study. *Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology* 2002; **13 Suppl 1**: 32-7.
17. Cullinan P, MacNeill SJ, Harris JM, et al. Early allergen exposure, skin prick responses, and atopic wheeze at age 5 in English children: a cohort study. *Thorax* 2004; **59**(10): 855-61.
18. Custovic A, Ainsworth J, Arshad H, et al. The Study Team for Early Life Asthma Research (STELAR) consortium 'Asthma e-lab': team science bringing data, methods and investigators together: Figure 1. *Thorax* 2015; **70**(8): 799-801.
19. Simpson A, Tan VY, Winn J, et al. Beyond atopy: multiple patterns of sensitization in relation to asthma in a birth cohort study. *Am J Respir Crit Care Med* 2010; **181**(11): 1200-6.

20. Bisgaard H, Simpson A, Palmer CN, et al. Gene-environment interaction in the onset of eczema in infancy: filaggrin loss-of-function mutations enhanced by neonatal cat exposure. *PLoS Med* 2008; **5**(6): e131.
21. 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): 383-96.
22. Williams HC, Burney PG, Pembroke AC, Hay RJ. The U.K. Working Party's Diagnostic Criteria for Atopic Dermatitis. III. Independent hospital validation. *Br J Dermatol* 1994; **131**(3): 406-16.
23. 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.
24. Hanifin J, Rajka G. Diagnostic features of atopic dermatitis. *Acta Derm Venereol (Stockh)* 1980; **92(Suppl)**: 44-7.
25. Dizon MP, Yu AM, Singh RK, et al. Systematic review of atopic dermatitis disease definition in studies using routinely collected health data. *Br J Dermatol* 2018.
26. Bieber T. How to Define Atopic Dermatitis? *Dermatologic Clinics* 2017; **35**(3): 275-81.
27. Andersen YMF, Egeberg A, Hamann CR, et al. Poor agreement in questionnaire-based diagnostic criteria for adult atopic dermatitis is a challenge when examining cardiovascular comorbidity. *Allergy* 2018; **73**(4): 923-31.
28. Belgrave DC, Granell R, Simpson A, et al. Developmental profiles of eczema, wheeze, and rhinitis: two population-based birth cohort studies. *PLoS Med* 2014; **11**(10): e1001748.
29. Roduit C, Frei R, Depner M, et al. Phenotypes of Atopic Dermatitis Depending on the Timing of Onset and Progression in Childhood. *JAMA Pediatrics* 2017; **109**(2): 338-42.
30. Paternoster L, Savenije OEM, Heron J, et al. Identification of atopic dermatitis subgroups in children from two longitudinal birth cohorts. *The Journal of allergy and clinical immunology* 2017.
31. Brough HA, Simpson A, Makinson K, et al. Peanut allergy: effect of environmental peanut exposure in children with filaggrin loss-of-function mutations. *J Allergy Clin Immunol* 2014; **134**(4): 867-75 e1.
32. 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. *Journal of Allergy and Clinical Immunology* 1999; **103**(1): 125-38.
33. Popescu CM, Popescu R, Williams H, Forsea D. Community validation of the United Kingdom diagnostic criteria for atopic dermatitis in Romanian schoolchildren. *The British journal of dermatology* 1998; **138**(3): 436-42.
34. Leung DYM. Atopic dermatitis: Age and race do matter! *Journal of Allergy and Clinical Immunology* 2015; **136**(5): 1265-7.
35. Leung DYM, Guttman-Yassky E. Deciphering the complexities of atopic dermatitis: Shifting paradigms in treatment approaches. *Journal of Allergy and Clinical Immunology* 2014; **134**(4): 769-79.
36. Garmhausen D, Hagemann T, Bieber T, et al. Characterization of different courses of atopic dermatitis in adolescent and adult patients. *Allergy* 2013; **68**(4): 498-506.
37. Asher MII, Keil U, Anderson HRR, et al. International study of asthma and allergies in childhood (ISAAC): rationale and methods. *European Respiratory Journal* 1995; **8**(3): 483-91.
38. Semic-Jusufagic A, Gevaert P, Bachert C, Murray C, Simpson A, Custovic A. Increased serum-soluble interleukin-5 receptor alpha level precedes the development of eczema in children. *Pediatric Allergy and Immunology* 2010; **21**(7): 1052-8.
39. Ziyab AH, Raza A, Karmaus W, et al. Trends in eczema in the first 18 years of life: results from the Isle of Wight 1989 birth cohort study. *Clinical & Experimental Allergy* 2010; **40**(12): 1776-84.
40. Schuttelaar MLA, Kerkhof M, Jonkman MF, et al. Filaggrin mutations in the onset of eczema, sensitization, asthma, hay fever and the interaction with cat exposure. *Allergy* 2009; **64**(12): 1758-65.

41. Henderson J, Northstone K, Lee SP, et al. The burden of disease associated with filaggrin mutations: a population-based, longitudinal birth cohort study. *The Journal of allergy and clinical immunology* 2008; **121**(4): 872-7.e9.
42. Schmitt J, Apfelbacher C, Spuls PI, et al. The Harmonizing Outcome Measures for Eczema (HOME) roadmap: a methodological framework to develop core sets of outcome measurements in dermatology. *J Invest Dermatol* 2015; **135**(1): 24-30.

[5]***STELAR investigators***

Prof Syed Hasan Arshad

Prof Graham Devereux

Dr Raquel Granell

Professor John Henderson

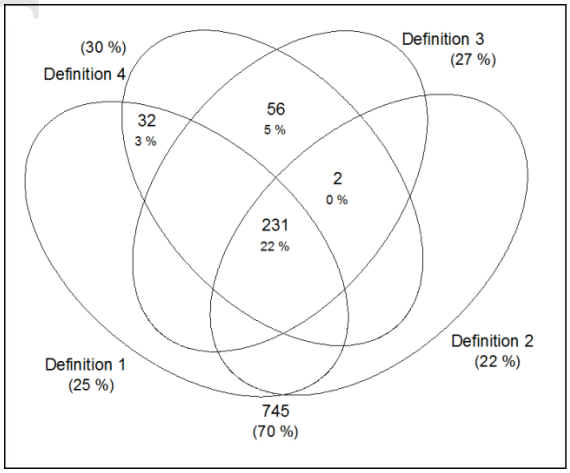
Prof Graham Roberts

Prof Steve Turner

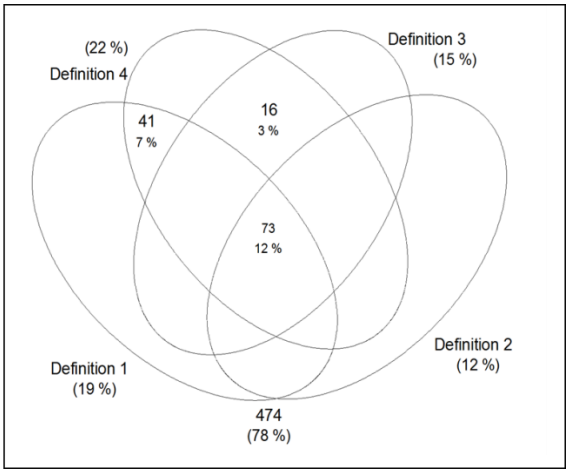
Figure 1. Overlap of each definition for current atopic dermatitis in MAAS and Ashford.

Definition 1: Physician-confirmed AD; Definition 2: Physician-confirmed AD and parent-reported chronic itchy skin condition affecting skin creases; Definition 3: Parent-reported chronic itchy skin condition affecting skin creases; and Definition 4: Physician-confirmed AD or parent-reported chronic itchy skin condition affecting skin creases.

MAAS (N = 1066*)



Ashford (N = 604)

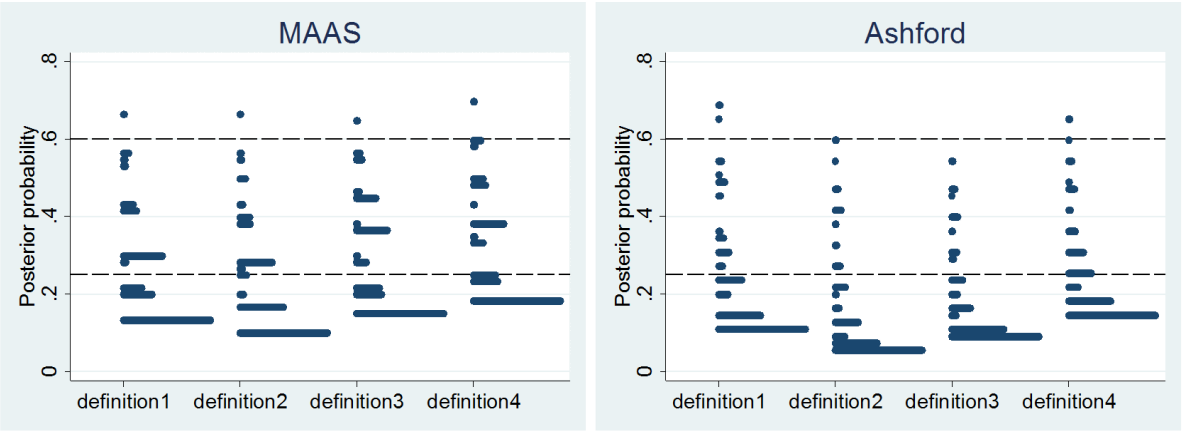


* 3 children had missing values in Definitions 1 and 2.

Figure 2. Performance of prediction models for four different "case" definitions in MAAS and Ashford.

Definition 1-4: See Table 1.

Multivariate logistic regression analysis included *Filaggrin* mutations, parental history of AD, allergic sensitization at age 5, and physician-confirmed asthma at age 5 as predictors. The area of clinical indecision represents percentages of children whose posterior probability lies between 25% and 60% or 25% and 50%.



Clinical indecision								
25-60%:	36%	32%	35%	44%	18%	9%	10%	29%
25-50%:	31%	28%	31%	40%	15%	9%	9%	27%

Table 1. AD definitions for "Case" applied to the data in MAAS and Ashford cohorts.

Question 1 (physician-confirmed ever AD) "Has a doctor ever told you that your child had eczema?" and "Has a doctor ever told you that your son or daughter has eczema?";

Question 2 (current itchy skin condition) "Has your child had an itchy rash at any time in the last 12 months" and "In the last twelve months, has your child had an itchy skin rash? (by itchy we mean scratching or rubbing the skin)";

Question 3 (current flexural rash) "Has this itchy rash at any time affected any of the following places: the fold of the elbows, behind the knees; in front of the ankles, under the buttocks; around the neck, ear or eyes?" and "Has this skin condition at any time affected the skin creases in the past? (by skin creases we mean fronts of elbows, behind the knees, fronts of ankles".

		Question			
	Definitions for "cases"	1	2	3	Response to questions
1	Physician-confirmed AD	✓	✓		Yes to 1 and 2
2	Physician-confirmed AD and Chronic itchy skin condition affecting skin creases	✓	✓	✓	Yes to 1, 2 and 3
3	Chronic itchy skin condition affecting skin creases		✓	✓	Yes to 2 and 3
4	Physician-confirmed AD OR Chronic itchy skin condition affecting skin creases	✓	✓	✓	Yes to (1 and 2) or (2 and 3) or (1, 2, and 3)

Table 2. Characteristics of the study populations

The denominators stand for children without a missing value for each variable.

Variables	MAAS Frequency (%)	Ashford Frequency (%)
Gender (male)	581/1069 (54)	259/499 (52)
Parent history of AD	265/1068 (25)	174/593 (29)
Paternal history of AD	112/1068 (10)	82/593 (14)
Maternal history of AD	175/1069 (16)	110/596 (18)
Dog ownership at recruitment	174/1047 (17)	155/596 (26)
Cat ownership at recruitment	219/1047 (21)	223/596 (37)
Physician-confirmed ever AD	421/1058 (40)	214/604 (35)
Current itchy skin condition	344/1069 (32)	165/604 (27)
Current flexural rash	292/1069 (27)	89/604 (15)
Physician-confirmed asthma	248/1062 (23)	118/604 (20)
Atopic sensitization	291/954 (30)	78/551 (14)
Ethnicity (Caucasian)	971/1023 (95)	568 /574 (99)
Filaggrin null mutations	73/795 (9)	45/439 (10)

Table 3. Association between FLG null mutations and four "Case" definitions among the children of white European origin.

Definition 1: Physician-confirmed AD; Definition 2: Physician-confirmed AD and parent-reported chronic itchy skin condition affecting skin creases; Definition 3: Parent-reported chronic itchy skin condition affecting skin creases; and Definition 4: Physician-confirmed AD or parent-reported chronic itchy skin condition affecting skin creases.

ORs: odds ratios; C.I.: confident interval; Binary logistic regression

	MAAS		Ashford	
	ORs (95% C.I.)	P value	ORs (95% C.I.)	P value
Definition 1	1.9 (1.1 – 3.2)	0.02	2.3 (1.2 – 4.5)	0.02
Definition 2	2.2 (1.3 – 3.7)	0.003	2.2 (1.0 – 4.7)	0.045
Definition 3	1.8 (1.1 – 2.9)	0.027	1.7 (0.8 – 3.7)	0.15
Definition 4	1.8 (1.1 – 2.9)	0.02	1.9 (0.9 – 3.8)	0.052

Table 4. Odds ratios for the association between AD and *FLG* mutations in two different "control" definitions using same case definition among the children of white European origin (Definition 4: Physician-confirmed AD or parent-reported chronic itchy skin condition affecting skin creases).

ORs: odds ratios; 95% C.I.: 95% confident interval; Binary logistic regression

Control definition	MAAS			Ashford		
	N (%)	ORs (95% C.I.)	P value	N (%)	ORs (95% C.I.)	P value
Strict						
Controls	519 (64)	(reference)		315 (72)	(reference)	
Cases	286 (36)	2.4 (1.5 – 4.0)	0.001	123 (28)	2.2 (1.1 - 4.6)	0.03
Moderate						
Controls	685 (71)	(reference)		445 (78)	(reference)	
Cases	286 (29)	1.8 (1.1 – 2.9)	0.02	123 (22)	1.9 (0.99 - 3.8)	0.052