

Identification of eczema clusters and their association with filaggrin and atopic comorbidities: analysis of five birth cohorts

Sadia Haider^{1,2}, Raquel Granell³, John A. Curtin⁴, John W. Holloway^{5,6}, Sara Fontanella^{1,2}, Syed Hasan Arshad^{6,7,8}, Clare S. Murray⁴, Paul Cullinan¹, Stephen Turner^{9,10}, Graham Roberts^{5,6,7}, Angela Simpson⁴ and Adnan Custovic^{1,2}

¹National Heart and Lung Institute, Imperial College London, UK

²NIHR Imperial Biomedical Research Centre (BRC), London, UK

³MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

⁴Division of Infection, Immunity and Respiratory Medicine, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, NIHR Manchester Biomedical Research Unit, Manchester University NHS Foundation Trust, Manchester, UK

⁵Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK

⁶NIHR Southampton Biomedical Research Centre, University Hospitals Southampton NHS Foundation Trust, Southampton, UK

⁷David Hide Asthma and Allergy Research Centre, Newport, Isle of Wight, UK

⁸Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK

⁹Royal Aberdeen Children's Hospital, NHS Grampian, Aberdeen, UK

¹⁰Child Health, University of Aberdeen, Aberdeen, UK

Correspondence: Adnan Custovic. Email: a.custovic@imperial.ac.uk

S.H. and R.G. are co-first authors.

S.T., G.R., A.S. and A.C. are co-senior authors.

Abstract

Background Longitudinal modelling of the presence/absence of current eczema through childhood has identified similar phenotypes, but their characteristics often differ between studies.

Objectives To demonstrate that a more comprehensive description of longitudinal pattern of symptoms may better describe trajectories than binary information on eczema presence.

Methods We derived six multidimensional variables of eczema spells from birth to 18 years of age (including duration, temporal sequencing and the extent of persistence/recurrence). Spells were defined as consecutive observations of eczema separated by no eczema across 5 epochs in five birth cohorts: infancy (first year); early childhood (age 2–3 years); preschool/early school age (4–5 years); middle childhood (8–10 years); adolescence (14–18 years). We applied Partitioning Around Medoids clustering on these variables to derive clusters of the temporal patterns of eczema. We then investigated the stability of the clusters, within-cluster homogeneity and associated risk factors, including *FLG* mutations.

Results Analysis of 7464 participants with complete data identified five clusters: (i) no eczema (51.0%); (ii) early transient eczema (21.6%); (iii) late-onset eczema (LOE; 8.1%); (iv) intermittent eczema (INT; 7.5%); and (v) persistent eczema (PE; 11.8%). There was very-high agreement between the assignment of individual children into clusters when using complete or imputed ($n = 15\,848$) data (adjusted Rand index = 0.99; i.e. the clusters were very stable). Within-individual symptom patterns across clusters confirmed within-cluster homogeneity, with consistent patterns of symptoms among participants within each cluster and no overlap between the clusters. Clusters were characterized by differences in associations with risk factors (e.g. parental eczema was associated with all clusters apart from LOE; sensitization to inhalant allergens was associated with all clusters, with the highest risk in the PE cluster). All clusters apart from LOE were associated with *FLG* mutations. Of note, the strongest association was for PE [relative risk ratio (RRR) 2.70, 95% confidence interval (CI) 2.24–3.26; $P < 0.001$] followed by INT (RRR 2.29, 95% CI 1.82–2.88; $P < 0.001$).

Conclusions Clustering of multidimensional variables identified stable clusters with different genetic architectures. Using multidimensional variables may capture eczema development and derive stable and internally homogeneous clusters. However, deriving homogeneous symptom clusters does not necessarily mean that these are underpinned by completely unique mechanisms.

Accepted: 1 September 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of British Association of Dermatologists. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

What is already known about this topic?

- Eczema is heterogeneous, but there is no general consensus on what the different subtypes are.
- Techniques such as latent class analysis have been used to disentangle the long-term course of eczema using categorical longitudinal data.
- Eczema phenotypes assigned the same name in different studies often differ in age of onset, temporal trajectory, distributions within a population and associated risk factors, making comparisons difficult and clinical application uncertain.

What does this study add?

- Clustering a set of multidimensional indicators that describe the temporal development of eczema more holistically than binary repeated measures may lead to more internally homogeneous phenotypes robust to missing data imputation.
- Persistent and intermittent clusters showed the strongest associations with established risk factors of eczema, including *FLG* mutations.
- Using multidimensional variables may capture temporal patterns in the development of eczema symptoms and derive stable and homogeneous clusters, but this does not guarantee that these are underpinned by completely unique mechanisms.

Eczema (or atopic dermatitis) is the most common chronic inflammatory skin disease that affects infants and children, with variable symptom expression, therapeutic responses and disease progression. The longitudinal course of eczema is not routinely incorporated into clinical management, practice guidelines or clinical trial design.¹ Over the last 5 years, there has been a developing literature on the clustering of longitudinal data on eczema, using data-driven approaches (summarized in Table S1; see [Supporting Information](#)).^{2–10} In most studies, repeated information on the presence or absence of current eczema was modelled to derive classes (phenotypes), mostly using latent class analysis (LCA).¹¹ It is generally assumed that such phenotypes are internally homogeneous (i.e. that most children within each class have the same or very similar longitudinal pattern of symptoms) and stable (i.e. that the same individual clusters into the same phenotype in multiple runs of the model), and that they reflect underlying mechanisms.^{11,12} However, although eczema phenotypes derived in different studies are usually described using the same name (e.g. transient, persistent and late-onset), they often differ across studies in the time of onset/resolution of symptoms and associated risk factors.¹¹ Furthermore, there are notable differences in the prevalence of each phenotype, and inconsistent associations with *FLG* – the most important genetic risk variant for eczema (Table S1). In our previous study of LCA in a population-based birth cohort, we demonstrated that the confidence with which individuals were assigned to an eczema class/phenotype varied across phenotypes, and that a substantial number of children were classified imprecisely, in part due to the presence of noisy data arising from the intermittent reporting of eczema.⁶ Consequently, within-class individual patterns of eczema do not always match the phenotype label; for example, some children assigned to transient eczema report symptoms in mid-school age, some with similar symptom patterns are assigned to different phenotypes and some who reported eczema at certain time points are assigned to no eczema.⁶ This within-class heterogeneity and inaccurate individual allocation to phenotypes/classes may be partly responsible for a lack of consistent associations with risk factors, and may adversely affect the ability to identify phenotype-specific genetic associates and mechanisms.

Similar heterogeneity has been reported in childhood wheezing illness, which has been extensively investigated using data-driven approaches to derive subgroups for genetic, mechanistic and therapeutic studies.¹³ Several studies have suggested that LCA using binary information on current wheeze (the most common methodology for wheeze phenotype derivation) may classify individuals imprecisely,^{14–16} and that children with similar wheezing patterns can be assigned to different phenotypes.¹⁴ Recently, we developed a novel method to derive wheeze clusters by modelling multidimensional variables of wheezing spells rather than binary information on the presence of current wheeze.¹⁷ Spells were defined as observations of current wheeze on consecutive follow-ups separated by no eczema. This approach was much more robust in dealing with missing data, and the derived clusters were stable and internally homogeneous.¹⁷

We propose that a similar framework using the spells approach on longitudinal data in eczema may produce clusters/phenotypes that are internally homogeneous and stable, with individual patterns that are consistent within phenotypes and distinct between phenotypes. To this end, we used longitudinal data on eczema from infancy to adolescence from five UK birth cohorts to derive a set of multidimensional variables, and clustered these to describe the temporal variation of eczema. We investigated the robustness of these models to data imputation, which is of key importance for longitudinal studies in which data missingness is inevitable, and for pooled genetic analyses in which multiple cohorts with large sample sizes are essential. We then explored the associations of the derived clusters with early-life factors and tested the hypothesis that clusters have differential associations with *FLG* loss-of-function mutations and allergic comorbidities.

Patients and methods**Study design, setting and participants**

We used data from five UK population-based birth cohorts: the Avon Longitudinal Study of Parents and Children

(ALSPAC),¹⁸ Ashford,¹⁹ Isle of Wight (IOW)²⁰ and Aberdeen (SEATON)²¹ cohorts, and the Manchester Asthma and Allergy Study (MAAS).²² The cohorts and methods are described in detail in Appendix S1 (see [Supporting Information](#)). Data were harmonized to facilitate pooled analyses.²³

Data sources and definition of variables

Validated questionnaires were completed on multiple occasions from infancy to adolescence. Current eczema was generally defined as a positive response to the question 'Has your child had an itchy rash/eczema in the last 12 months?' The exact questions used in each cohort are provided in Table S2 (see [Supporting Information](#)), and the cohort-specific timepoints and sample sizes in Table S3 (see [Supporting Information](#)). For pooled analyses, we defined five epochs based on shared follow-up points: infancy (first year of life); early childhood (2–3 years); pre-school/early school age (4–5 years); middle childhood (8–10 years); adolescence (14–18 years). Using data on the presence/absence of current eczema at each epoch, we derived six indicators for each child: (i) age at the first observation; (ii) age at the last observation; (iii) total number of separate records over the observation period; (iv) duration of the longest spell (number of consecutive points with current eczema); (v) total number of separate spells; and (vi) spell type [0 = no eczema, 1 = single spell, 2 = intermittent spells (at least 2 nonconsecutive spells of eczema of any length)].

Spells were defined as consecutive eczema observations separated by no eczema. An example of the derivation of the indicators is shown in Table S4 (see [Supporting Information](#)).

Skin testing was carried out during school age in all cohorts, and at six follow-up points in MAAS.

FLG genotyping

Genotyping was performed using probes and primers as previously described. For R501X mutation, we used a TaqMan-based allelic discrimination assay. Mutation 2282del4 was genotyped by the sizing of a fluorescently labelled polymerase chain reaction fragment on a 3100 or 3730 DNA sequencer (Applied Biosystems, Birchwood, UK). The data were analysed as combined carriage of an *FLG* null allele.²⁴

Statistical analysis

Detailed descriptions of statistical and sensitivity analyses are provided in Appendix S1.

Eczema clusters (phenotypes) from infancy to adolescence derived from six indicators

To derive longitudinal clusters captured by the six indicators, we used the Partitioning Around Medoids algorithm coupled with the Wishart distance for mixed data.^{25,26} We adopted the framework of Basagaña *et al.*,²⁷ which integrates multiple imputation into cluster analysis among 15 848 participants with at least 2 observations of eczema across the 5 epochs.²⁸ The algorithm was run for 2–7 clusters, and the optimal number was chosen using the average of the silhouette width (ASW) criterion and Pearson gamma coefficient.

Model stability was assessed by comparing the optimal number of clusters using the Wishart and Gower distance matrices, and with and without the inclusion of the total number of separate spells indicator. To determine the stability to changes in different sample sizes, we ran multiple iterations of the model while sampling random subsets of varying sample size with decrements of 10% from the full set until 50% were included.

Association of clusters with FLG, early-life risk factors and clinical outcomes in adolescence

We used multinomial logistic regression to ascertain associations of clusters with early-life risk factors and other associates (including *FLG* mutations). The results are reported as relative risk ratios (RRRs) with 95% confidence intervals (CIs). Associations with comorbidities and allergic sensitization in adolescence were analysed using logistic regression. In adjusted models, multicollinearity was assessed using a variance inflation factor (ensuring that values did not exceed 10). Cluster analysis was conducted in R (version 3.6.3; R Foundation for Statistical Computing, Vienna, Austria) using the *cluster* package (version 2.1.0).²⁹ The *ggplot2* package was used for data visualization.³⁰ Multiple imputation was performed using the *mi* suite of commands in Stata 17 (StataCorp, College Station, TX, USA).

Results

Characteristics of the study population

A total of 15 848 participants had data on eczema for at least 2 epochs and 7464 had complete data (50.6% male). Demographic characteristics are provided in Table 1. The proportion of parental smoking was higher and breastfeeding lower in those with complete data, with no difference in any of the clinical outcomes. At the follow-up in adolescence, 41.2% reported eczema ever; the prevalence was similar (42.6%) among 8384/15 848 individuals with incomplete data. The prevalence of current eczema was 18.1% in infancy; it peaked at 26.1% in early childhood and declined to 13.7% in adolescence (Table S5; see [Supporting Information](#)).

Clusters of eczema from infancy to adolescence

A five-cluster solution was optimal based on statistical fit (Figure S1; see [Supporting Information](#)). Based on the age of onset and duration of eczema in each cluster (Figure 1), the clusters (phenotypes) were characterized as: (i) no eczema (NOE; 51.0%); (ii) early transient eczema (ETE; 21.6%), where participants had eczema in infancy, which peaked in early childhood and remitted by mid-childhood; (iii) late-onset eczema (LOE; 8.1%), with eczema observed from pre-school age to a peak in mid-childhood and a gradual decline by adolescence; (iv) intermittent eczema (INT; 7.5%), where eczema was observed recurrently throughout the observation period; and (v) persistent eczema (PE; 11.8%) in which eczema was observed throughout the observation period.

Table 1 Demographic characteristics of the study population

	Participants with complete data (<i>n</i> =7464)	Participants with data at 2–4 points (<i>n</i> =8384)
Eczema ever (reported in adolescence)	2601/6309 (41.2)	1435/3372 (42.6)
Perinatal characteristics		
Male sex	3780/7464 (50.6)	4406/8374 (52.6)
Low birthweight (≤ 2500 g)	302/7357 (4.1)	449/7989 (5.6)
Parental characteristics		
Maternal age at delivery, mean (SD)	29.1 (4.6)	27.6 (5.1)
Maternal asthma ever (recruitment)	899/7412 (12.1)	996/7786 (12.8)
Paternal asthma ever (recruitment)	811/6256 (13.0)	730/5855 (12.5)
Environmental characteristics		
Breastfeeding ever	5810/7331 (79.2)	4986/6942 (71.8)
Maternal smoking (recruitment)	1242/6516 (19.1)	1933/6783 (28.5)
Paternal smoking (recruitment)	2074/7366 (28.1)	2801/7642 (36.6)
Presence of cat in home (recruitment)	2219/7316 (30.3)	2217/7742 (28.6)
Presence of dog in home (recruitment)	1422/6712 (21.2)	1783/7130 (25.0)
Outcomes in adolescence (age 14–18 years)		
Asthma ever	1648/6306 (26.1)	913/3092 (29.5)
Current asthma	696/5614 (12.4)	360/2417 (14.9)
Current asthma medication	637/5641 (11.3)	294/2501 (11.8)
Cohort		
ALSPAC	5120/11 158 (45.9)	6038/11 158 (54.1)
MAAS	669/1135 (58.9)	466/1135 (41.1)
SEATON	471/1488 (31.7)	1017/1488 (68.4)
IOW	784/1455 (53.9)	671/1455 (46.1)
Ashford	420/612 (68.6)	192/612 (31.4)

Data are presented as *n* (%) unless otherwise stated. Valid percentages displayed using available data for each variable. ALSPAC, Avon Longitudinal Study of Parents and Children; MAAS, Manchester Asthma and Allergy Study; SEATON, Aberdeen; IOW, Isle of Wight.

Within-cluster homogeneity

Individual patterns of current eczema within each cluster were consistent, with no overlap between the clusters (Figure 2). Table S6 shows the distribution of the derived indicators stratified by cluster, and Table S7 shows pairwise comparisons of quantitative indicators (see [Supporting Information](#)). INT was the only cluster in which children experienced three separate spells of eczema. In contrast,

PE was characterized by longer/single spells lasting 3–5 consecutive epochs.

Model stability and impact of missing data

Table S8 shows the distribution of the ASW across the samples and confirms that a five-class solution was consistently identified (see [Supporting Information](#)). There was a high agreement of partitions obtained between each cohort and the pooled data [adjusted Rand index (ARI) > 0.88] (Table S9; see [Supporting Information](#)). Missing data increased from 8% in infancy to 37% by adolescence (Figure S2; see [Supporting Information](#)). The optimal solution from the model among 15 848 participants with ≥ 2 observations was very similar to that from 7464 with complete data (Table S10, Figure S3; see [Supporting Information](#)). There was high agreement between the cluster assignment of individual participants (ARI = 0.99); only 53/7464 (0.71%) changed allocation (Table S11; see [Supporting Information](#)). The proportions of participants assigned to clusters were similar (Table S12; see [Supporting Information](#)). The optimal solution was stable to the exclusion of ALSPAC (Figure S4; see [Supporting Information](#)).

Associations with *FLG* mutations and early-life risk factors

The results of multinomial regression analysis are summarized in Figure 3 and Table S13 (see [Supporting Information](#)). Compared with NOE, all eczema clusters apart from LOE were associated with *FLG* mutations. The strongest association was for PE (RRR 2.7, 95% CI 2.24–3.26; $P < 0.001$). Breastfeeding was associated with membership of all eczema clusters, with the highest risk for ETE (RRR 1.35,

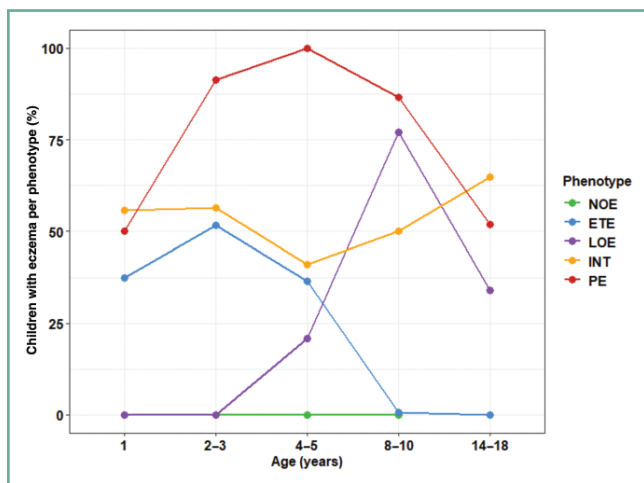


Figure 1 Trajectories of five clusters obtained with the Partitioning Around Medoids algorithm [no eczema (NOE) and four eczema clusters]: percentage of participants with reported eczema in the five cohorts (complete dataset *n* = 7464): NOE, *n* = 3809 (51.0%); early transient eczema (ETE), *n* = 1611 (21.6%); late-onset eczema (LOE), *n* = 604 (8.1%); persistent eczema (PE), *n* = 878 (11.8%); intermittent eczema (INT), *n* = 562 (7.5%).

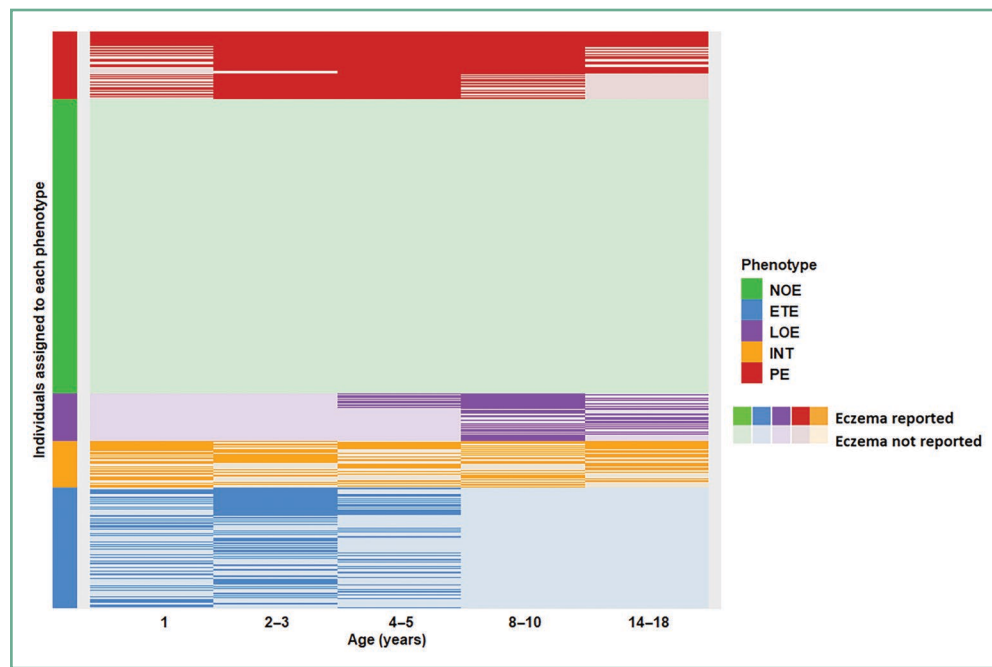


Figure 2 Within-individual patterns of eczema development stratified by clusters (phenotypes) among 7464 participants with complete data on eczema from infancy to adolescence. NOE, $n=3809$ (51.0%); early transient eczema (ETE), $n=1611$ (21.6%); late-onset eczema (LOE), $n=604$ (8.1%); persistent eczema (PE), $n=878$ (11.8%); intermittent eczema (INT), $n=562$ (7.5%).

95% CI 1.21–1.51; $P < 0.001$). Males were significantly less likely to experience LOE and PE but were at higher risk of ETE. Both findings were stable in an adjusted model that included risk factors in infancy (Table S14; see [Supporting Information](#)). Wheeze in infancy was associated with all clusters except LOE, with similar risks for INT and PE (approximate RRR 1.5). Early pet exposure was protective for PE in the univariate analysis but not in the adjusted multinomial model (Tables S13, S14).

Associations with allergic sensitization and atopic comorbidities

All eczema clusters were associated with allergic sensitization in early school age, with the risk being highest for those in the PE and INT clusters (Table S13). Trajectories of sensitization from infancy to adolescence in MAAS were similar for the PE and INT clusters, and differed from those in other clusters (Figure 4); differential patterns by cluster remained throughout childhood. Eczema preceded sensitization in the PE and INT clusters: at 1 year of age < 10% of children in these clusters were sensitized; by 16 years of age, approximately half were. In contrast, children in the LOE cluster developed sensitization before the onset of eczema.

Table 2 shows the adjusted associations of eczema clusters with asthma, rhinitis and allergic sensitization in adolescence. Relative to NOE, associations were highest for the PE cluster for all outcomes. Current asthma was associated with the INT and PE clusters but not the ETE or LOE clusters. All eczema clusters were associated with current rhinitis and with an increased risk of sensitization in adolescence, with the highest risk in the PE and INT clusters.

Individual allocation across eczema clusters and previously derived wheeze clusters using the same populations

and analytical approach is shown in Figure S5 and Table S15 (see [Supporting Information](#)).¹⁷ There was no clear pattern in the sequence of development between eczema and wheeze clusters.

Discussion

By pooling data on parent-reported eczema from birth to adolescence in five population-based birth cohorts, we have shown – using the spells approach – that we can achieve a potentially more accurate description of eczema trajectories. We identified four distinct, internally homogeneous and stable clusters of eczema. Three of the derived clusters (ETE, LOE and PE) appeared to be similar to those derived in previous studies. We identified a novel and clinically intuitive cluster (INT) in which children experienced episodic symptoms over time, and derived a pure NOE cluster, which comprised children who never reported eczema. Importantly, our approach was robust to missing data imputation, as demonstrated by the remarkably high agreement between assignment of individuals into clusters when using complete or imputed data, with < 1% of children changing cluster. This may facilitate genetic analyses in a much larger sample.

Several recent studies have derived eczema phenotypes/classes using LCA (Table S1).^{2–10} There is considerable heterogeneity in study design (e.g. unselected, selected and high-risk cohorts), duration of follow-up, number of data collection points and age at collection (e.g. the age of the final data collection ranged from 6 to 50 years). Different definitions of eczema were used, including ISAAC (International Study of Asthma and Allergies in Childhood) definitions, parent-reported severity, doctor diagnosis, medication use and investigator-assessed and composite definitions. The

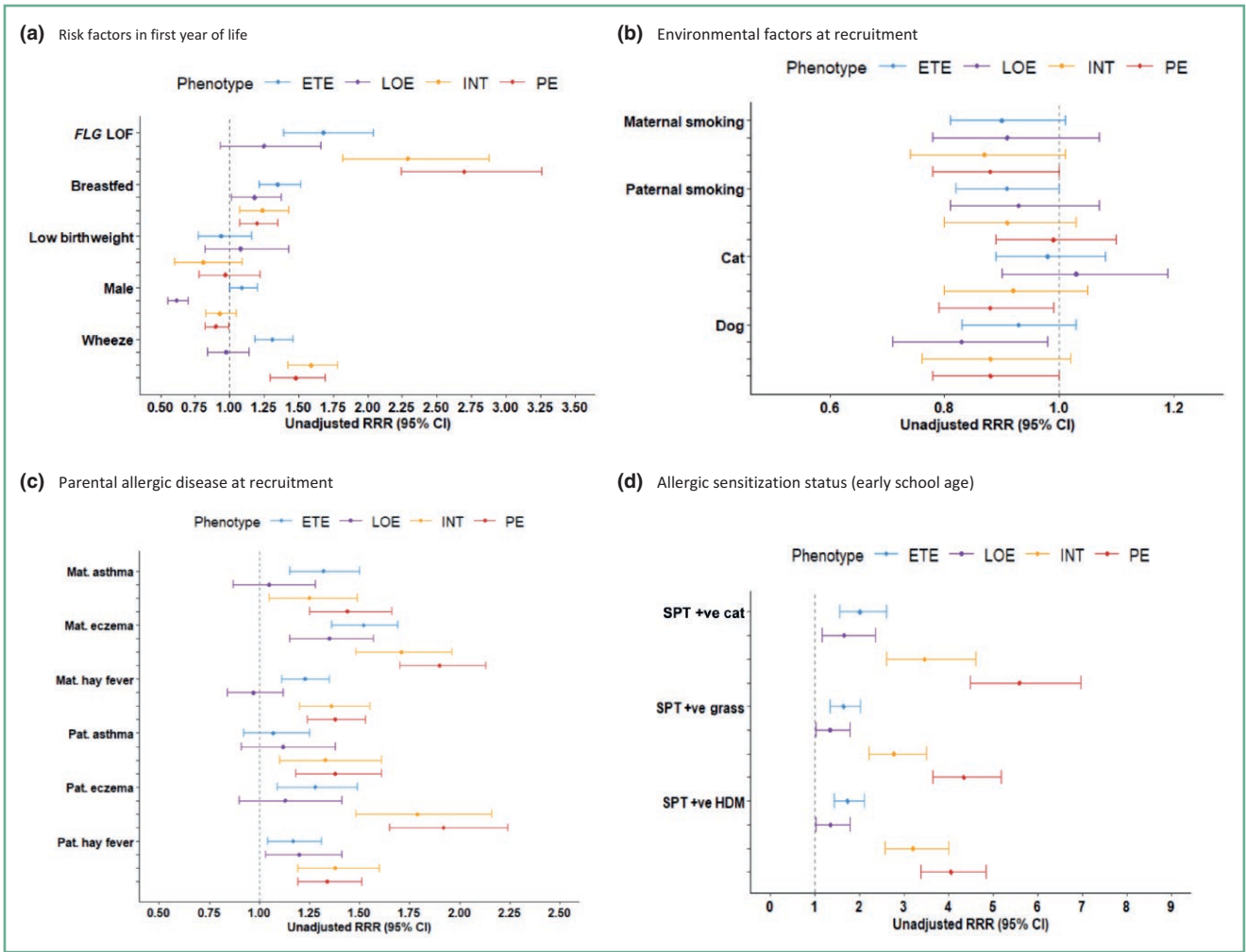


Figure 3 Associations of eczema clusters (phenotypes) with early-life risk factors and allergic sensitization in early school age: results from multinomial logistic regression in children with ≥ 2 observations on eczema [reference class: no eczema (NOE)] using weighted membership probabilities. Weights derived from probabilities of class membership across 10 imputation samples from the Partitioning Around Medoids model. Results are reported as unadjusted relative risk ratios (RRRs) with 95% confidence intervals (CIs). ETE, early transient eczema; HDM, house-dust mite; INT, intermittent eczema; LOE, late-onset eczema; LOF, loss of function; Mat., maternal; Pat., paternal; PE, persistent eczema; SPT, skin prick test.

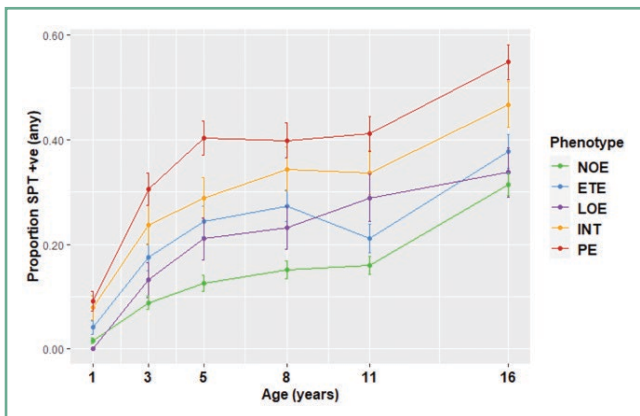


Figure 4 Proportion of children with allergic sensitization in each cluster (Manchester Asthma and Allergy Study); positive sensitization to at least one allergen (cat, grass, house-dust mite). ETE, early transient eczema; INT, intermittent eczema; LOE, late-onset eczema; NOE, no eczema; PE, persistent eczema; SPT, skin prick test.

number of derived phenotypes ranged from 3 to 6, and the size of the persistent class varied from 2% to 12% of the study populations. Although the nomenclature varied, all studies identified early transient and persistent phenotypes, while some did not observe LOE.^{4,6} Analyses in ALSPAC and the PIAMA study (Prevention and Incidence of Asthma and Mite Allergy) identified two resolving classes, which differed based on the timing of onset: 'early-onset, late-resolving' and 'late-onset, resolving'.³¹

In a data-driven analysis in ALSPAC, Mulick *et al.* applied LCA to information on parent-reported eczema and self-reported severity,⁵ and identified a mild intermittent phenotype (14%). The most persistent eczema was characterized by the most severe symptoms, whereas clusters mostly characterized by intermittent eczema were reported to be milder.⁵ In our study, we did not use information on symptom severity and, unsurprisingly, the intermittent phenotypes in the two analyses differed. Inference from these two studies, which used different data sources and

Table 2 Associations of eczema clusters with allergic comorbidities and eczema in adolescence: results from multinomial logistic regression using children with ≥ 2 observations on eczema (reference class: no eczema) using weighted membership probabilities

Associations with allergic comorbidities in adolescence (age 14–18 years) ^a				
	Current asthma ^b		Asthma ever	Current rhinitis
NOE	Ref.		Ref.	Ref.
ETE	1.20 (0.95–1.50)		1.50 (1.28–1.75)	1.24 (1.06–1.45)
<i>P</i> -value	0.122		< 0.001	0.006
LOE	1.32 (0.95–1.50)		1.48 (1.19–1.83)	1.28 (1.04–1.58)
<i>P</i> -value	0.07		< 0.001	0.02
PE	3.29 (2.66–4.08)		3.47 (2.94–4.10)	2.44 (2.06–2.89)
<i>P</i> -value	< 0.001		< 0.001	< 0.001
INT	2.24 (1.71–2.94)		2.32 (1.89–2.85)	1.94 (1.58–2.38)
<i>P</i> -value	< 0.001		< 0.001	< 0.001
Associations with sensitization in adolescence (SEATON, MAAS, IOW) ^a				
	Grass	HDM	Cat	At least one positive SPT
NOE	Ref.	Ref.	Ref.	Ref.
ETE	1.49 (1.15–1.93)	1.35 (1.03–1.76)	1.37 (0.97–1.94)	1.48(1.16–1.89)
<i>P</i> -value	0.003	0.03	0.08	0.001
Late onset	1.35 (0.96–1.90)	1.55 (1.10–2.17)	1.31 (0.83–2.07)	1.36 (0.99–1.86)
<i>P</i> -value	0.08	0.01	0.24	0.05
Persistent	3.04 (2.29–4.05)	2.53 (1.90–3.38)	3.98 (2.87–5.51)	2.99 (2.25–3.99)
<i>P</i> -value	< 0.001	< 0.001	< 0.001	< 0.001
INT	2.67 (1.91–3.75)	1.88 (1.33–2.67)	2.59 (1.73–3.87)	2.38 (1.71–3.34)
<i>P</i> -value	< 0.001	< 0.001	< 0.001	< 0.001

Weights derived from probabilities of class membership across 10 imputation samples from the Partitioning Around Medoids model. Results are reported as adjusted odds ratios with 95% confidence intervals. Significant *P*-values are highlighted in bold. ETE, early transient eczema; HDM, house-dust mite; INT, intermittent eczema; LOE, late-onset eczema; NOE, no eczema; PE, persistent eczema; SPT, skin prick test. ^aModels adjusted for parental history of allergic diseases (recruitment), parental smoking (recruitment), sex, breastfeeding status, presence of a pet in the home (recruitment) and low birthweight. ^bAvailable at the latest follow-up [18 years in Isle of Wight (IOW), 16 years in Manchester Asthma and Allergy Study (MAAS), 15 years in SEATON (Aberdeen), 15 years in Ashford and 15 years in the Avon Longitudinal Study of Parents and Children (ALSPAC)].

analytical approaches, suggests that the temporality of symptoms is likely confounded by severity. This indicates that more clinically intuitive subgroups of eczema, which may better reflect underlying mechanisms, may be identified by applying data-driven techniques to detailed information on frequency, severity and triggers of different skin symptoms.

Another key difference between our study and previous LCA analyses is that we derived a 'pure' NOE group, where no participants experienced eczema. Previous analyses in two of our cohorts (ALSPAC and MAAS) have quantified or visualized the presence of eczema in the derived latent classes, and up to 36% of children in the 'unaffected' class experienced eczema at least once during the study period. Also, in contrast with previous studies, in the current

analysis ETE and LOE clusters had clearly demarcated ages at which symptoms resolved or began.

PE and INT clusters showed the strongest associations with established risk factors of eczema, including *FLG* mutations, comorbid wheeze, parental history of eczema and allergic sensitization. In terms of environmental factors, early cat and dog exposure appeared to be protective of persistent and intermittent eczema. Sex was differentially associated with clusters, with females being at higher risk of LOE and males at higher risk of ETE. All clusters showed evidence of an increased risk of asthma and rhinitis in adolescence, suggesting that the risk of comorbidities is pervasive throughout childhood, regardless of the age at eczema onset. However, the associations with current asthma and rhinitis in adolescence were strongest with the PE cluster.

Breastfeeding was associated with a significantly increased risk of all eczema clusters. A meta-analysis of 27 prospective studies demonstrated a lack of protection of breastfeeding on the incidence of eczema.³² Rather than breastfeeding having a causal effect on eczema, the association is likely a result of reverse causation (e.g. increased duration of breastfeeding among those with a family history or symptoms of eczema).

Consistent with previous studies, clusters derived using our spell-based indicators had differential association with *FLG* mutations. In general, the strongest association was with PE (in previous studies, the odds ratios ranged from 0.8 to 4.3). We have shown that the impact of *FLG* mutations may vary throughout childhood, with some children with *FLG* mutations experiencing resolving symptoms and others having persistent or recurrent symptoms. Analyses in other cohorts is required to demonstrate whether more homogeneous clusters will lead to the discovery of a clearer contribution of genetic variants to the development of specific eczema trajectories.

Using wheeze clusters previously derived with the same methods in the same populations,¹⁷ we were able to glean insights into the patterns of development of eczema and wheeze. Our study does not support the existence of a specific trajectory from eczema to wheeze. A quarter of children with LOE had preceding wheeze, and children with ETE were more likely to have no wheeze than to develop subsequent wheeze. However, some children who outgrow eczema may still be at risk of developing subsequent wheeze, suggesting that the nature of the association between these symptoms and mechanisms underlying this may change over time. Similarly, analyses in ALSPAC and PIAMA found that while the PE phenotype showed the strongest association with asthma in adolescence, there was an increased risk across all eczema phenotypes.⁷ These findings confirm observations that there is not a clear directional relationship between eczema and asthma.³³

Our study had several limitations. We have previously shown that different definitions of eczema led to inconsistent associations with *FLG* genotype.^{6,34} We acknowledge that questionnaire-based parent reports are not sensitive enough to differentiate severity or true eczema from other pruritic skin conditions. Hanifin and Rajka proposed diagnostic criteria for eczema based on a physical examination;³⁵ however, there is no uniform definition of the disease for large-cohort studies in which regular physical examination is not feasible.^{36–38} This has resulted in numerous different definitions being used in epidemiological and genetic studies.^{34,39} Our previous study showed that the use of different eczema definitions significantly influenced prevalence estimates and associations with risk factors.³⁴ We did not have access to confirmed diagnosis by a physician in all five cohorts. However, diagnoses across clinicians may be heterogeneous and clinicians may rely upon history provided by parents for diagnosing the condition. While questionnaires provide useful information on the prevalence of self-reported symptoms, there may be a risk of overestimating the prevalence, as eczema may present similarly to other skin conditions (particularly common rashes in infancy).

We harmonized common data points across cohorts into epochs for pooled analyses, thereby reducing the number of time points. However, we acknowledge that the inclusion

of all available data may lead to more granular insights into disease architecture over time, owing to the increase in information.

Another limitation is that genetic analyses were carried out for White participants only, and our findings on genetic associations cannot be generalized to other groups. Recent evidence from a US-based multiethnic cohort observed different sequences for allergic diseases in Black vs. White children.⁴⁰ It is important to understand how patterns of eczema and their associates vary in different ethnic groups, and the mechanisms underlying these, through greater representation of under-represented groups in research. Therefore, further work in multiethnic cohorts is of critical importance.

We did not have access to more specific clinical markers of skin dysfunction; however, this is a limitation common to most population-based birth cohorts. We could not ascertain whether treatment adherence patterns varied by cluster. While there could be an association between received treatment, disease severity, duration and patterns of eczema, reported prescribed medication does not equate to accurate information on whether study participants were taking the drug. This is a limitation of birth cohort studies in general. Finally, our findings were exploratory in nature and hypothesis generating.

In conclusion, we applied a methodology that derived four eczema clusters that were robust to data imputation, each comprising individuals with homogeneous symptom sequences. We did not identify risk factors that uniquely predicted the risk of a particular cluster, suggesting that the clusters, even when internally homogeneous and stable, are not necessarily underpinned by completely unique mechanisms (i.e. are not endotypes). It would be important to assess whether modelling different disease domains, including markers of immune system dysregulation and epidermal barrier dysfunction (which themselves are affected by genetic and environmental factors), gives rise to a similar phenotypic structure to strengthen the evidence for the clinical importance of distinct patterns of eczema development.

Funding sources

The UNICORN consortium was funded by the Medical Research Council Programme Grant MR/S025340/1; the STELAR consortium was funded through the Medical Research Council grants G0601361 and MR/K002449/1. Infrastructure support for this research was provided by the National Institute for Health and Care Research Imperial Biomedical Research Centre (BRC). The UK MRC and Wellcome (grant ref: 217065/Z/19/Z), and the University of Bristol provide core support for the Avon Longitudinal Study of Parents and Children (ALSPAC). This publication is the work of the authors and Adnan Custovic, and R.G. will serve as guarantor for the contents of this paper. A comprehensive list of grant funding is available on the ALSPAC website (<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>). The Manchester Asthma and Allergy Study (MAAS) was supported by Asthma UK grants no. 301 (1995–1998), no. 362 (1998–2001), no. 01/012 (2001–2004) and no. 04/014 (2004–2007); the BMA James Trust (2005) and the JP Moulton Charitable Foundation (2004–2016); the North West Lung Centre Charity (1997–current); and MRC

grant MR/L012693/1 (2014–2018). The Isle of Wight cohort was supported by the Isle of Wight Health Authority, the National Asthma Campaign, UK (grant no. 364) and National Institutes of Health grants R01 HL082925-01, R01 AI091905 and R01 AI121226.

Conflicts of interest

A.C. reports personal fees from Novartis, Thermo Fisher Scientific, Philips, Sanofi, Stallergenes Greer and AstraZeneca outside the submitted work. A.S. reports lecture fees from Chiesi. C.S.M. reports personal fees from Novartis and Boehringer Ingelheim and grants and personal fees from GSK. The other authors declare no conflicts of interest.

Data availability

The data underlying this article will be shared upon reasonable request to the corresponding author.

Ethics statement

All studies were approved by research ethics committees. Informed consent was obtained from parents, and participants gave their assent/consent when applicable.

Supporting Information

Additional [Supporting Information](#) may be found in the online version of this article at the publisher's website.

References

- Chovatiya R, Silverberg JI. Evaluating the longitudinal course of atopic dermatitis: implications for clinical practice. *Am J Clin Dermatol* 2022; **23**:459–68.
- Abuabara K, Ye M, Margolis DJ *et al.* Patterns of atopic eczema disease activity from birth through midlife in 2 british birth cohorts. *JAMA Dermatol* 2021; **157**:1191–9.
- Hu C, Duijts L, Erler NS *et al.* Most associations of early-life environmental exposures and genetic risk factors poorly differentiate between eczema phenotypes: the Generation R Study. *Br J Dermatol* 2019; **181**:1190–7.
- Lopez DJ, Lodge CJ, Bui DS *et al.* Establishing subclasses of childhood eczema, their risk factors and prognosis. *Clin Exp Allergy* 2022; **52**:1079–90.
- Mulick AR, Mansfield KE, Silverwood RJ *et al.* Four childhood atopic dermatitis subtypes identified from trajectory and severity of disease and internally validated in a large UK birth cohort. *Br J Dermatol* 2021; **185**:526–36.
- Nakamura T, Haider S, Fontanella S *et al.* Modelling trajectories of parentally reported and physician-confirmed atopic dermatitis in a birth cohort study. *Br J Dermatol* 2022; **186**:274–84.
- Paternoster L, Savenije OEM, Heron J *et al.* Identification of atopic dermatitis subgroups in children from 2 longitudinal birth cohorts. *J Allergy Clin Immunol* 2018; **141**:964–71.
- Roduit C, Frei R, Depner M *et al.* Phenotypes of atopic dermatitis depending on the timing of onset and progression in childhood. *JAMA Pediatr* 2017; **171**:655–62.
- Suaini NHA, Yap GC, Bui DPT *et al.* Atopic dermatitis trajectories to age 8 years in the GUSTO cohort. *Clin Exp Allergy* 2021; **51**:1195–206.
- Ziyab AH, Mukherjee N, Zhang H *et al.* Sex-specific developmental trajectories of eczema from infancy to age 26 years: a birth cohort study. *Clin Exp Allergy* 2022; **52**:416–25.
- Duverdier A, Custovic A, Tanaka RJ. Data-driven research on eczema: systematic characterization of the field and recommendations for the future. *Clin Transl Allergy* 2022; **12**:e12170.
- Custovic A, Custovic D, Kljaic Bukvic B *et al.* Atopic phenotypes and their implication in the atopic march. *Expert Rev Clin Immunol* 2020; **16**:873–81.
- Oksel C, Haider S, Fontanella S *et al.* Classification of pediatric asthma: from phenotype discovery to clinical practice. *Front Pediatr* 2018; **6**:258.
- Oksel C, Granell R, Mahmoud O *et al.* Causes of variability in latent phenotypes of childhood wheeze. *J Allergy Clin Immunol* 2019; **143**:1783–90.
- Hallmark B, Wegienka G, Havstad S *et al.* Chromosome 17q12-21 variants are associated with multiple wheezing phenotypes in childhood. *Am J Respir Crit Care Med* 2021; **203**:864–70.
- Oksel C, Granell R, Haider S *et al.* Distinguishing wheezing phenotypes from infancy to adolescence. a pooled analysis of five birth cohorts. *Ann Am Thorac Soc* 2019; **16**:868–76.
- Haider S, Granell R, Curtin J *et al.* Modeling wheezing spells identifies phenotypes with different outcomes and genetic associates. *Am J Respir Crit Care Med* 2022; **205**:883–93.
- Fraser A, Macdonald-Wallis C, Tilling K *et al.* Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol* 2013; **42**:97–110.
- 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**:855–61.
- Arshad SH, Holloway JW, Karmaus W *et al.* Cohort profile: the Isle Of Wight Whole Population Birth Cohort (IOWBC). *Int J Epidemiol* 2018; **47**:1043–4.
- Martindale S, McNeill G, Devereux G *et al.* Antioxidant intake in pregnancy in relation to wheeze and eczema in the first two years of life. *Am J Respir Crit Care Med* 2005; **171**:121–8.
- Custovic A, Simpson BM, Murray CS *et al.* The National Asthma Campaign Manchester Asthma and Allergy Study. *Pediatr Allergy Immunol* 2002; **13**(s15):32–7.
- 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. *Thorax* 2015; **70**:799–801.
- Simpson A, Brough HA, Haider S *et al.* Early-life inhalant allergen exposure, filaggrin genotype, and the development of sensitization from infancy to adolescence. *J Allergy Clin Immunol* 2020; **145**:993–1001.
- Kaufman L, Rousseeuw PJ. Partitioning Around Medoids (Program PAM). In: *Finding Groups in Data*. Wiley, 1990.
- Wishart D. k-Means Clustering with outlier detection, mixed variables and missing values. In: *Exploratory Data Analysis in Empirical Research* (Schwaiger M, Opitz O, eds). Berlin, Heidelberg: Springer, 2003; 216–26.
- Basagaña X, Barrera-Gomez J, Benet M *et al.* A framework for multiple imputation in cluster analysis. *Am J Epidemiol* 2013; **177**:718–25.
- Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, 2002.
- Maechler M, Rousseeuw P, Struyf A *et al.* *cluster: Cluster Analysis Basics and Extensions. R package version 2.1.0*. Vienna: R Foundation for Statistical Computing.
- Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer, 2009.
- Paternoster L, Savenije OEM, Heron J *et al.* Identification of atopic dermatitis subgroups in children from 2 longitudinal birth cohorts. *J Allergy Clinical Immunol* 2018; **141**:964–71.
- Yang YW, Tsai CL, Lu CY. Exclusive breastfeeding and incident atopic dermatitis in childhood: a systematic review and

- meta-analysis of prospective cohort studies. *Br J Dermatol* 2009; **161**:373–83.
- 33 Haider S, Fontanella S, Ullah A *et al.* Evolution of eczema, wheeze, and rhinitis from infancy to early adulthood: four birth cohort studies. *Am J Respir Crit Care Med* 2022; **206**:950–60.
- 34 Nakamura T, Haider S, Colicino S *et al.* Different definitions of atopic dermatitis: impact on prevalence estimates and associated risk factors. *Br J Dermatol* 2019; **181**:1272–9.
- 35 Hanifin JM, Rajka G. Diagnostic features of atopic dermatitis. *Acta Derm Venereol (Stockh)* 1980; **92**(Suppl.):44–7.
- 36 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**:1480–5.
- 37 Silverberg JI, Thyssen JP, Paller AS *et al.* What's in a name? Atopic dermatitis or atopic eczema, but not eczema alone. *Allergy* 2017; **11**:1–5.
- 38 Frainay C, Pitarch Y, Filippi S *et al.* Atopic dermatitis or eczema? Consequences of ambiguity in disease name for biomedical literature mining. *Clin Exp Allergy* 2021; **51**:1185–94.
- 39 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. *J Allergy Clin Immunol* 2012; **130**:137–44.
- 40 Biagini JM, Kroner JW, Baatyrbek Kyzy A *et al.* Longitudinal atopic dermatitis endotypes: an atopic march paradigm that includes Black children. *J Allergy Clin Immunol* 2022; **149**:1702–10.