Symptom trajectories in infancy for the prediction of subsequent wheeze and asthma in the BILD and PASTURE cohorts: a dynamic network analysis

Uri Nahum, Olga Gorlanova, Fabienne Decrue, Heide Oller, Edgar Delgado-Eckert, Andreas Böck, Sven Schulzke, Philipp Latzin, Bianca Schaub, Anne M Karvonen, Roger Lauener, Amandine Divaret-Chauveau, Sabina Illi, Caroline Roduit, Erika von Mutius, Urs Frey, BILD study group, PASTURE study group**

Summary

Background Host and environment early-life risk factors are associated with progression of wheezing symptoms over time; however, their individual contribution is relatively small. We hypothesised that the dynamic interactions of these factors with an infant's developing respiratory system are the dominant factor for subsequent wheeze and asthma.

Methods In this dynamic network analysis we used data from term healthy infants from the Basel-Bern Infant Lung Development (BILD) cohort (435 neonates aged 0–4 weeks recruited in Switzerland between Jan 1, 1999, and Dec 31, 2012) and replicated the findings in the Protection Against Allergy Study in Rural Environments (PASTURE) cohort (498 infants aged 0–12 months recruited in Germany, Switzerland, Austria, France, and Finland between Jan 1, 2002, and Oct 31, 2006). BILD exclusion criteria for the current study were prematurity (<37 weeks), major birth defects, perinatal disease of the neonate, and incomplete follow-up period. PASTURE exclusion criteria were women younger than 18 years, a multiple pregnancy, the sibling of a child was already included in the study, the family intended to move away from the area where the study was conducted, and the family had no telephone connection. Outcome groups were subsequent wheeze, asthma, and healthy. The first outcome was defined as ever wheezed between the age of 2 years and 6 years. Week-by-week correlations of the determining factors with cumulative symptom scores (CSS) were calculated from weeks 2 to 52 (BILD) and weeks 8 to 52 (PASTURE). The complex dynamic interaction between the determining factors and the CSS was assessed via dynamic host–environment correlation network, quantified by a simple descriptor: trajectory function $G(t)$. Wheeze outcomes at age 2–6 years were compared in 335 infants from BILD and 437 infants from PASTURE, and asthma outcomes were analysed at age 6 years in a merged cohort of 783 infants.

Findings CSS was significantly different for wheeze and asthma outcomes and became increasingly important during infancy in direct comparison with all determining factors. Weekly symptoms were tracked for groups of infants, showing a non-linear increase with time. Using logistic regression classification, $G(t)$ distinguished between the healthy group and wheeze or asthma groups (area under the curve>0·97, p<0·0001; sensitivity analysis confirmed significant CSS association with wheeze [BILD p=0·0002 and PASTURE p=0·068]) and $G(t)$ was also able to distinguish between the farming and non-farming exposure groups (p<0·0001).

Interpretation Similarly to other risk factors, CSS had weak sensitivity and specificity to identify risks at the individual level. At group level however, the dynamic host–environment correlation network properties ($G(t)$) showed excellent discriminative ability for identifying groups of infants with subsequent wheeze and asthma. Results from this study are consistent with the 2018 *Lancet* Commission on asthma, which emphasised the importance of dynamic interactions between risk factors during development and not the risk factors per se.

Funding The Swiss National Science Foundation, the Kühne Foundation, the EFRAIM study EU research grant, the FORALLVENT study EU research grant, and the Leibniz Prize.

Introduction

We and others[1–6] (*The Lancet*'s Commission on asthma; hypothesis 2)[7] have raised the hypothesis that the sum of the relative contribution of risk factors for asthma or wheezing disorders beyond infancy is composed of a dynamic interactive network of host and environment factors with the respiratory system, and that the dynamic interactions of these determining risk factors with an infant´s developing respiratory system are the dominant factor for subsequent wheeze. We previously identified in our cohort that sex, caesarean section, breastfeeding, maternal atopy, viral exposure (siblings or nursery care),

Lancet Digit Health 2024; 6: e718–28

*Members of the study groups are listed at the end of the Article

University Children's Hospital Basel, University of Basel, Basel, Switzerland (U Nahum PhD, O Gorlanova MD, F Decrue PhD, H Oller MSc, E Delgado-Eckert PhD, Prof S Schulzke PhD, Prof U Frey PhD); **Pediatric Respiratory Medicine, Department of Pediatrics, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland** (U Nahum, O Gorlanova, Prof P Latzin PhD); **Pediatric Allergology, Department of Pediatrics, Dr von Hauner Children's Hospital, University Hospital, Ludwig-Maximilians-Universität Munich Haunerschen Kinderklinik, Munich, Germany** (A Böck MD, B Schaub PhD, Prof E von Mutius MD); **University of Basel, Basel, Switzerland** (Prof U Frey); **Member of German Center for Lung Research—DZL, Ludwig-Maximilians-Universität Munich, Germany** (A Böck, B Schaub, S Illi PhD, Prof E von Mutius); **Centre for Cardiovascular Science, Queens Medical Research Institute, Edinburgh, UK** (F Decrue); **Department of Health Security, Finnish Institute for Health and Welfare, Kuopio, Finland** (Prof A M Karvonen PhD); **Christine Kühne Center for Allergy Research and Education, Davos, Switzerland** (Prof R Lauener MD, C Roduit PhD); **Children's Hospital of Eastern Switzerland, St Gallen, Switzerland** (Prof R Lauener); **Pediatric Allergy Department, University Hospital of Nancy, Nancy, France** (Prof A Divaret-Chauveau PhD); **UR3450 Développement Adaptation et Handicap,**

University of Lorraine, Nancy, France (Prof A Divaret-Chauveau); **UMR 6249 Chrono-environment, CNRS and University of Franche-Comté, Respiratory Disease Department, University Hospital of Besançon, Besançon, France** (Prof A Divaret-Chauveau); **Institute for Asthma and Allergy Prevention, Helmholtz Zentrum Munich, German Research Center for Environmental Health, Munich, Germany** (S Illi); **Comprehensive Pneumology Center Munich, Munich, Germany** (S Illi); **Division of Respiratory Medicine and Allergology, Department of Paediatrics, Inselspital, University of Bern, Bern, Switzerland** (C Roduit)

Correspondence to: Prof Urs Frey, University Children's Hospital Basel, University of Basel, 4056 Basel, Switzerland
urs.frey@ukbb.ch

Research in context

Evidence before this study

This literature review was based mainly on the knowledge and judgement of the authors, supported by selected references from a PubMed search for papers published from database inception to Nov 30, 2022, using two sets of search terms: ((((host) AND (environment)) AND (time)) AND (early life)) AND (asthma OR COPD) as well as ((((gene) AND (environment)) AND (on time)) AND (asthma OR COPD)) AND (early life). For the two searches, we retrieved 90 studies reporting time-dependent effects of host–environment interactions occurring in early life in childhood. Of these articles, 13 were reviews. Key papers from the authors' own files and online searches were also considered. Only a few papers were selected for this specialised topic, without language restrictions, based on relevance.

Added value of this study

Using machine learning and dynamic correlation network analysis, we showed in two birth cohorts that the correlations between early-life host and environment risk factors and respiratory symptoms in the first year of life undergo weekly temporal changes. These dynamic network-type correlation changes (cumulative symptom trajectories) are strong determinants of subsequent wheezing between 2 years and 6 years of life and asthma at age 6 years.

Implications of all the available evidence

The dynamic adaptive process of the respiratory system to the environmental risk factors in infancy effectively identifies infant groups with subsequent wheezing and asthma. These findings underline the crucial role of dynamic adaptation to the environment and its importance in disease evolvement in later life. Although not suited for risk assessment on an individual patient level, these novel early-life biomarkers are clinically and scientifically useful for identifying risk groups of infants with subsequent persistent airway disease. Our findings are consistent with the related hypothesis raised in the 2018 *Lancet* Commission on asthma, which emphasised the importance of dynamic interactions between environmental and host risk factors during development and not the risk factors per se.

and environmental tobacco exposure are important determinants for subsequent wheeze.[8,9] The relative importance of these determinants for subsequent wheeze might be age-dependent, because it is well known that the respiratory system undergoes rapid developmental changes in early life (eg, lung growth, immune, and anti-inflammatory system development). Thus, lung susceptibility to each of these host or environment factors might change at different ages. So far, limited attention has been given to the idea[7,10,11] (eg, the gene–environment–time model) of how the overall temporal evolution of health and disease is affected by the complex dynamic interplay of these determining factors.

The progression of a subsequent disease in early life might be considered as an impaired developmental process, given a set of host and environment factors.[7] We hypothesised that the way the respiratory system dynamically responds and adapts to these early-life environment factors is more relevant for subsequent wheeze than individual host and environment factors alone. This developmental adaptive process might be mathematically characterisable using a dynamic network approach. This approach would provide simple integrated dynamic early-life network predictors for the evolvement of subsequent wheeze and asthma. These predictors might have better predictive ability than the individual factors alone.

To characterise this dynamic, interactive, and integrative process we used data from the Basel-Bern Infant Lung Development (BILD)[12] and Protection Against Allergy Study in Rural Environments (PASTURE)[13] prospective birth cohorts. We aimed to make week-by-week correlations of the determining factors with the cumulative symptom score (CSS) obtained during the first year of life[8] and to investigate the dynamic changes of these correlations using mathematical network graphs. Because multivariate regression analysis is limited in capturing complex interactions, we aimed to use neural-network-based self-explaining machine learning to analyse the relative importance of the determining factors.[14]

The development of the methods was a step-by-step process directed by four aims. First, we aimed to determine how the week-by-week CSS change in the first year of life, and whether the dynamics of this change differ in infants with or without subsequent wheeze (ever wheezed between age 2 years and 6 years [first outcome] and asthma [second outcome]). Second, we aimed to examine the correlation of each determining factor on the CSS, week-by-week during the first year of life. At each week we determined whether the relative impact of a given factor on the CSS increased, decreased, or remained constant. To account for the complex network-type interactions between the factors and the symptoms during this adaptive process, we aimed to construct a correlation network model and to determine its dynamic behaviour over time, characterised by the trajectory function $G(t)$. Third, we aimed to analyse whether the CSS or $G(t)$ already assessed in infancy have the power to differentiate against the control group. Fourth, we aimed to investigate whether the CSS or $G(t)$ in the first year of life were dependent on the environmental context. We aimed to compare the cohorts' identical dynamic networks and also the groups of infants from PASTURE raised in farming and non-farming environments.

Methods

Study design

In this dynamic network analysis we developed the analytical methods for BILD[12] and then replicated the findings in PASTURE[13] (appendix pp 14–17). The BILD cohort comprised 435 neonates aged 0–4 weeks that were recruited antenatally in Switzerland between Jan 1, 1999, and Dec 31, 2012. The neonates were recruited randomly to represent a general population at birth and observed until the end of first year of life (infancy). The PASTURE cohort comprised 498 infants aged 0–12 months recruited in Germany, Switzerland, Austria, France, and Finland between Jan 1, 2002, and Oct 31, 2006, born in farming and non-farming environments. The BILD and PASTURE cohorts are two comparable birth cohort studies investigating the impact of early-life host and environmental risk factors contributing to the evolvement of asthma. BILD exclusion criteria for the current study were prematurity (<37 weeks), major birth defects, perinatal disease of the neonate, and incomplete follow-up period. PASTURE exclusion criteria were women younger than 18 years, a multiple pregnancy, the sibling of a child was already included in the study, the family intended to move away from the area where the study was conducted, and the family had no telephone connection. This study is an observational study with secondary data. The use of clinical data for the prediction of outcomes in both cohorts is part of the initial purpose of the original BILD and PASTURE cohorts and is thus covered by the BILD and PASTURE ethics.

Procedures

We calculated the week-by-week CSS (BILD 2–52 weeks of life exposures and PASTURE 8–52 weeks of life exposures). We first investigated the correlation between early-life host and environment risk factors and the week-by-week CSS in the first year of life, and then constructed a dynamic host-environment correlation network (DHECN) of these interactions quantified by $G(t)$. In subgroups of infants with available outcome measures at age 2–6 years, we then compared CSS in infancy to subsequent wheeze (first outcome) in 335 infants in BILD and $G(t)$ in infancy to subsequent wheeze in 437 infants in PASTURE. For the second outcome of asthma (lower prevalence), we merged the two cohorts (n=933) and analysed the correlation of CSS and $G(t)$ with subsequent asthma outcome (BILD + PASTURE) in 783 patients. In the PASTURE cohort we also tested the impact of farming environment (exposure) on the dynamic behaviour of CSS (DHECN). To test the clinical utility of the methods in a sensitivity analysis, we investigated the effect of missing data in the BILD cohort and symptom severity in the PASTURE cohort.

Based on previous BILD cohort studies,[9,15–21] we defined two types of classic risk factors: (1) host risk factors such as gestational age, sex (sex was determined by hospital record), maternal atopy, and caesarean section; and (2) environment factors such as low maternal education, parental smoking, prenatal and postnatal maternal smoking, prenatal and postnatal exposure to particulate matter air pollution (particulate matter with a diameter of 10 µM or less [PM_{10}] and NO_2), farm upbringing, nutritional factors (breastfeeding), and factors related to viral exposure (child care or siblings). Air pollution was calculated as described by Decrue and colleagues[16] and other factors were defined by Fuchs and colleagues.[12]

See Online for appendix

In the BILD cohort, any respiratory symptoms in the first year of life were prospectively recorded using severity scores,[8] resulting in nine symptom severity states (states 0–8; appendix p 5). To calculate the CSS, we added up the symptom severity states week-by-week, which is hypothesised to be a proxy for the integrated developmental process of the first year of life. This resulted in 52 CSS, expressed as $CSS=\Sigma_{symptoms}(t)$, whereby (t) is any given week. To calculate the cumulative effect of a determining factor, we computed each factor's sum up to a given week. This calculation then yielded, for each risk factor, the value $\Sigma_{risk\text{-}factor}(t)$. For example, $\Sigma_{breastfeeding}(t)$ is the sum of t weeks of breastfeeding. Further details of the analysis methodology, including the corresponding PASTURE cohort analysis, are in the appendix (pp 4–5). To compare with PASTURE, we repeated the analysis but with two differences: (1) using only the simple symptom scores (0 or 1; appendix p 5), and (2) starting from week 8 (not week 2, as in BILD).

The correlation between a given risk factor and the CSS at week t was calculated resulting in a time series of correlation coefficients as a function of time: $k_{risk\text{-}factor}=corr(\Sigma_{symptoms}(t), \Sigma_{risk\text{-}factor}(t))$. Here, the function corr denotes the Pearson correlation.

Because associations of all risk factors with symptoms in the first year of life should be considered week-by-week in a comprehensive and interacting manner, a dynamic correlation network best visualises our hypothesis. A detailed description of the construction in the appendix (pp 8, 19). The network consists of the previously described factors (represented in the outer ring) and the weekly CSS (which sits in the middle of the circle, representing the centre of gravity of all $k_{risk\text{-}factor}$). This centre of gravity changes week-by-week as a function of time t and is denoted by the trajectory function $G(t)$. The position of $G(t)$ is a measure for the interactions (strength of correlation) between each factor and the CSS. To quantify the trajectory function $G(t)$, we denoted the week-by-week summed Euclidian distance between $G(t)$ from the beginning to end of the first year of life (which corresponds to the length of the two-dimensional trajectory line). We defined this summed distance as the trajectory measure.

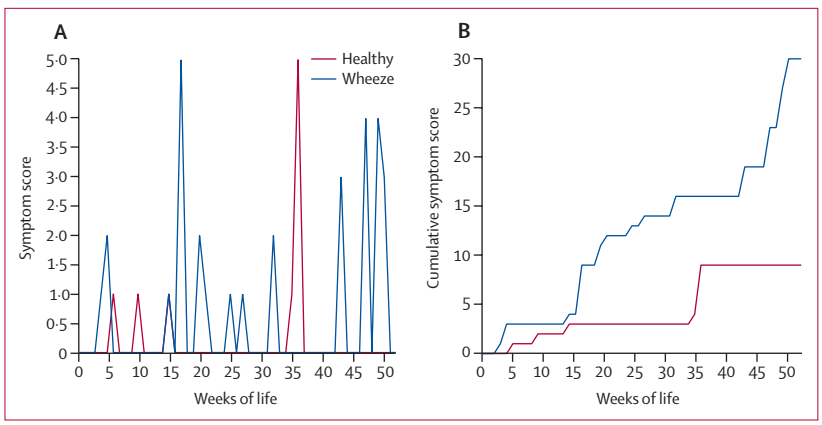

Figure 1: **Non-cumulative (A) and cumulative (B) weekly symptom scores from two infants**
(A) Representative time series of weekly symptom scores over the first year of life using the non-cumulative symptom score for an individual infant from the wheeze group and the healthy group. (B) Corresponding, week-by-week, cumulative symptom score time series, $\Sigma_{symptoms}$ (t) for weeks t=2–52, shown in (A).

Definition of outcomes

Outcomes were assessed by questionnaire during a visit at age 6 years. In accordance with the ISAAC III consortium[22] the primary outcome was defined as ever wheezed between the age of 2 years and 6 years. For simplicity, we referred to the group that ever wheezed from the age of 2 years to 6 years as the wheeze group and the group without wheeze as the healthy group. The comparison of *G(t)* measures for different outcome groups and the receiver operating curve (ROC) analysis are in the appendix (p 23).

Statistical analysis

First, we compared classic statistics (χ^2 or Kruskal–Wallis, logistic regression) between the two outcome groups and the time of contribution of each individual risk factor, $k_{risk\text{-}factor,}$ was calculated for each available risk factor. Then, by computing the p values and ROCs, and using self-explaining machine learning analysis, we determined whether the CSS interaction with risk factors is a better risk classifier for wheeze and asthma than risk factors alone. Self-explaining machine learning analysis (appendix p 7) accounts for all risk factor interactions, thus enabling the relative importance of all risk factors to be directly compared with the CSS. Self-explaining machine learning analysis identifies the risk factors carrying the most weight and shows how this dynamically changes over the first year of life.

We also tested whether the trajectory function *G(t)* in infancy differs between the wheeze, asthma, and healthy outcome groups. To statistically quantify *G(t)*, logistic regression classification was used (input was the trajectory measure; and the outcome was the wheeze group or the healthy group). Finally, we replicated the predictive power of the CSS and *G(t)* in the PASTURE cohort and compared the farming and non-farming PASTURE groups. Detailed statistical and sensitivity analysis information is in the appendix (pp 7–9, 11, 24). Of note, CSS is an individual classifier, whereas *G(t)* is a group classifier, thus area under the ROC (AUC) analysis (AUC_{group} of *G(t)*) is not directly comparable with $AUC_{individual}$ of CSS. The imputation strategy[23,24] is described in the appendix (p 5). Data were analysed using Python (version 3.9.13) and the Pandas (version 1.5.1), Seaborn (version 0.11.2), Scikit-learn (version 1.0.2), and NumPy (version 1.22.4) packages. Data visualisation was performed using Python packages Plotly (version 5.11.0) and NetworkX (version 2.7.1). The distribution of risk factors and symptoms in the first year of life, and missingness of data for both cohorts were processed in R (version 4.0.3) and Pandas (version 1.5.1; appendix p 7).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

435 participants were included from the BILD cohort, of whom 230 (53%) were male, 205 (47%) were female, and 435 (100%) were White middle European. 498 participants were included from the PASTURE cohort, of whom 248 (50%) were male and 250 (50%) were female. Ethnicity for the entire PASTURE cohort was 98·3% European (German, Swiss, Austrian, French, and Finnish) and 1·7% non-European. We showed the weekly symptoms as a function of time in the first year of life for one representative infant from the wheeze group and the healthy group and the infant's corresponding week-by-week CSS (figure 1). The median non-cumulative symptom scores and corresponding CSS for the whole subsequent wheeze group compared with the healthy group are in figure 2 (for PASTURE see appendix p 21). When comparing the CSS of the wheeze and the healthy groups (figure 2B), the infants with subsequent wheeze showed a non-linear steeper increase than the healthy group CSS. The findings in BILD were replicated using PASTURE (appendix p 22). A further analysis with asthma as the outcome confirmed that the use of CSS showed congruent behaviour (appendix p 23).

We performed a sensitivity analysis to test whether the ordering of the weekly CSS time series was a relevant finding and not found by chance. By replacing the observed non-cumulative symptom scores and CSS with scores calculated based on randomly shuffled symptom values, visible differences between the groups disappeared (appendix p 24).

The CSS at 52 weeks of life was significantly associated with subsequent wheeze in BILD: p=0·0002 (PASTURE p=0·068). However, to see how the predictive power of CSS and all other considered risk factors changed week-by-week during the first year of life and thus determine the best predictor, we used self-explaining machine learning (figure 3; appendix

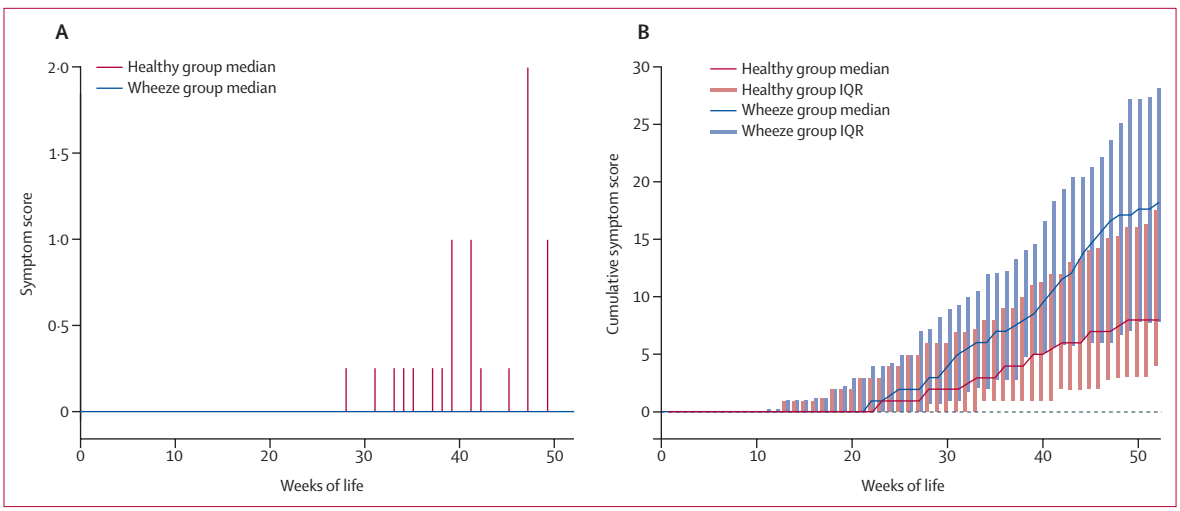

Figure 2: **Group medians and IQRs of non-cumulative symptom scores (A) and cumulative symptom scores (B)**
(A) The median values together with IQR for the number of symptoms over time for the wheeze group and the healthy group. Note the overlapping medians. (B) The median values of the corresponding cumulative symptom scores together with IQR for the wheeze group and the healthy group. From week 22, the two group medians increase with different slopes, supporting the relevance of the cumulative symptom scores as a predictor.

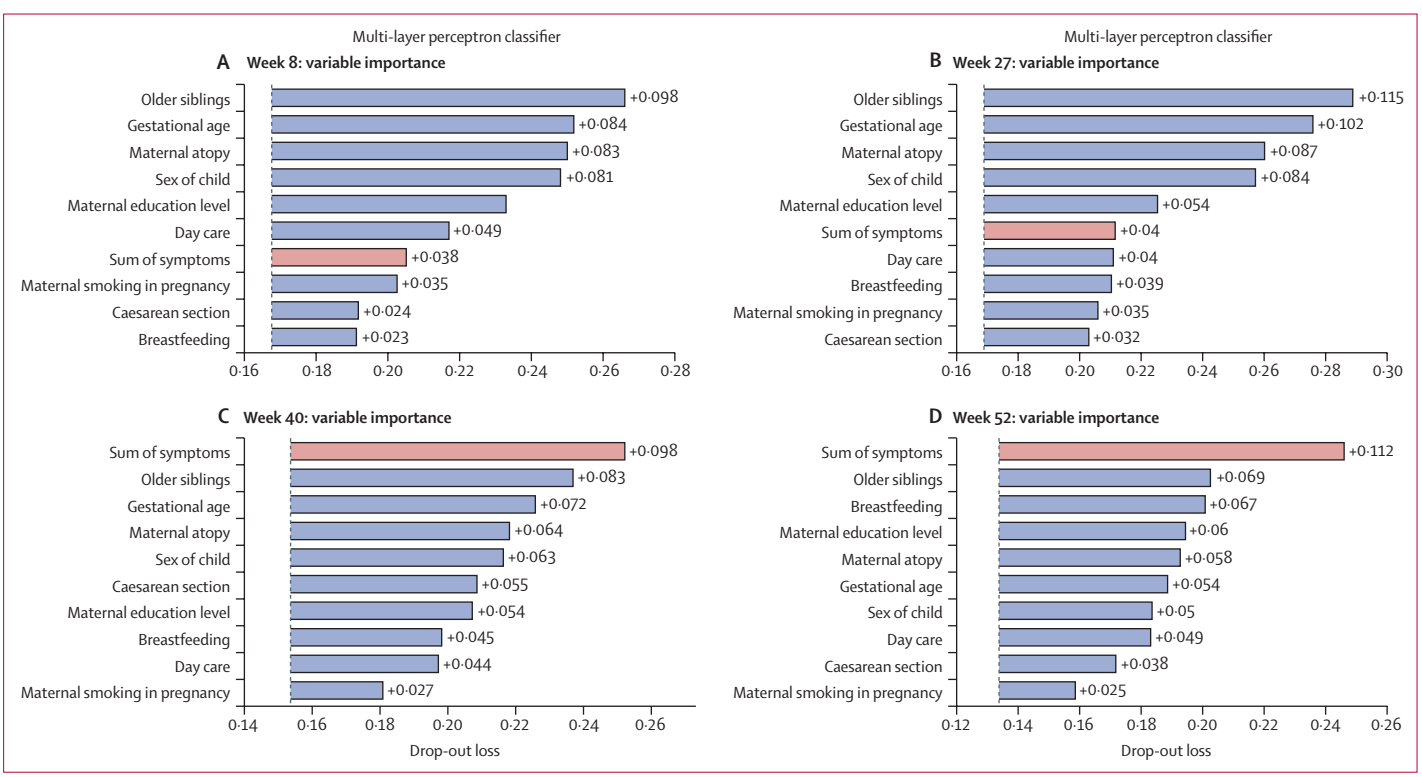

Figure 3: **Self-explaining machine learning showing the relative importance of risk factors over the first year of life**
Comparison of the relative importance of risk factors and the cumulative symptom score across multiple timepoints in the first year of life on subsequent wheeze using self-explaining machine learning. A multi-layer perceptron builds a model for prediction of subsequent wheeze (appendix pp 7, 25–28). This model is then used to evaluate the predictive power of each factor. Predictive power is determined by the difference between the loss function of the multi-layer perceptron model and two cases: (1) when each factor is dropped out (the difference is represented by the length of the bar; values are given in the horizontal axis); and (2) when the values of each factor are shuffled between the infants (the difference is represented by the value at the end of the bar). The importance of the cumulative symptoms rises over time and becomes the most important factor from week 34 until week 52. The full year animation is in video 1.

pp 25–28). When directly compared with the other determining factors, the relative importance of the CSS (sum of symptoms in figure 3) steadily increases week-by-week, becoming the most dominant factor from week 40 onwards (week 37 for PASTURE; appendix pp 25–26). Of note, the sensitivity and specificity of CSS

See **Online** for video 1

at week 52 for the clinical risk assessment on an individual patient ($AUC_{individual}$ for BILD was 0·65 and $AUC_{individual}$ for PASTURE was 0·61) was still weak. Similarly, the more specific outcome of asthma in the merged cohorts was 0·64.

The correlation coefficients ($k_{risk\text{-}factor}$) between all considered risk factors and the CSS as a function of time over the first year of life in BILD are in the appendix (p 29). During the first year of life, the influence of some factors generally decreased (eg, breastfeeding, smoking in pregnancy, and farming environment) and the impact of other factors increased (sex, day care, and siblings). The temporal influence of these factors on the CSS was tested in a sensitivity analysis, which suggested a true temporal influence (appendix p 30). For other risk factors (eg, NO_2 and PM_{10}), the influence showed a close-to-zero linear correlation throughout the first year of life.

The correlations of all risk factors with cumulative symptoms were represented by a DHECN. The DHECNs for weeks 4, 12, 26, and 52 show dynamically changing network characteristics over time (figure 4). An animation of these week-by-week changes is the first evidence of temporal changes of host–environment

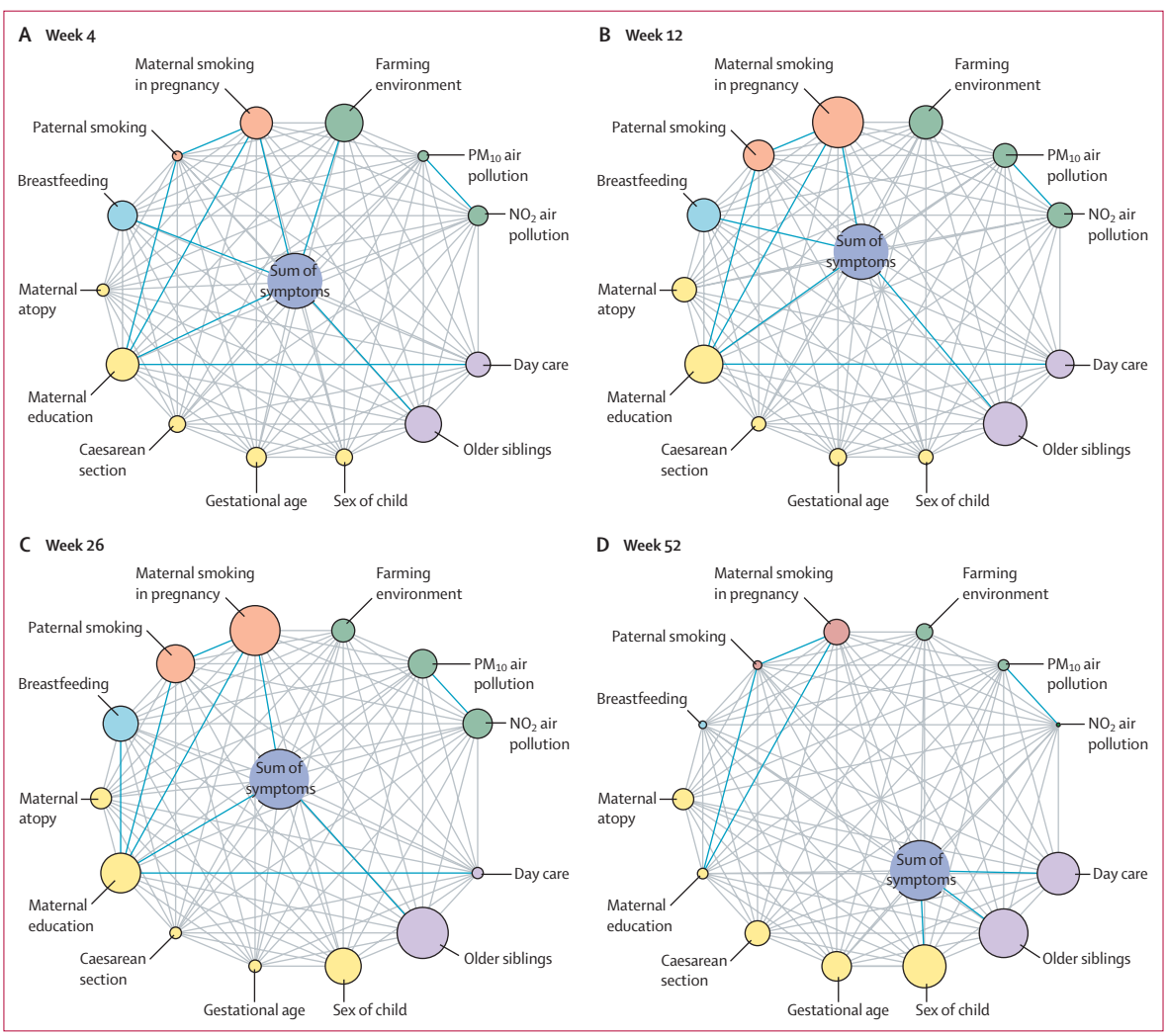

Figure 4: **Dynamic host–environment correlation network over the first year of life; the BILD cohort**

See **Online** for video 2

To simulate the overall adaptive behaviour of these interactions in the first year of life, a time series of networks from weeks 2 to 52 was generated. Here the network configurations at 4 weeks (A), 12 weeks (B), 26 weeks (C), and 52 weeks (D) are presented. The full animation for weeks 2 to 52 is in video 2. The different risk factors are represented by coloured circles with their size corresponding to their correlation to $\Sigma_{symptoms}$. The grey circle, (sum of symptoms), represents the cumulative symptom score for the given week ($\Sigma_{symptoms}$). Its position, $G(t)$, is determined by the strength of correlation with the various host and environment risk factors. Higher correlation to a risk factor shortens the distance between $G(t)$ and the risk factor's circle. The network line colour represents the correlation strength between individual risk factors and between risk factors and symptoms—stronger correlations are shown in blue. The position of the symptom circle is the two-dimensional result of the relative strength of the correlations (length of correlation lines) with all risk factors in the network graph. Across time, there is dynamic change in correlation and hence in $G(t)$, which indicates the adaptive process involving the competing influences of host and environment factors. BILD=Basel-Bern Infant Lung Development. PM_{10}=particulate matter with a diameter of less than 10 μm.

interactions in the first year of life (video 2). In the DHECN graphs, a higher correlation of a given risk factor to the CSS for a given week (ie, gravity point *G(t)*) shortens the distance between the two points, which results in a dynamic trajectory, *G(t)*, changing week-by-week. In early life, host factors play a more dominant role, however environmental factors, particularly those related to higher viral exposure (day care or siblings), start to dominate towards the end of the first year of life, as would be expected. In a correlation network representation, such influences are competing and determine the relative contribution to the developmental cumulative symptom trajectory (PASTURE appendix pp 31–32).

The DHECNs show differences in the cumulative symptom trajectories *G(t)* between the wheeze group ($G_{wheeze}(t)$) and the healthy group ($G_{healthy}(t)$; figure 5). The DHECNs show that the *G(t)* is distinct between the two groups at all times during the first year of life. As early as in the second week of life, *G(t)*, which at week 2 is *G(2)*, shows a major difference between the two outcome groups. Graphically, ($G_{wheeze}(t)$) and ($G_{healthy}(t)$) follow different dynamic trajectories, indicating that the relative impact of the various risk factors changes differently in the two outcome groups, even during the first year of life (PASTURE and asthma outcome are in the appendix pp 33–34). The trajectory measure is a comprehensive property and gives an AUC_{group} of 0·99 for BILD, 0·97 for PASTURE, and 0·98 for asthma outcome (figure 6).

In PASTURE we were able to study the influence of distinct environmental risk factors. From plotting the *G(t)* trajectories for PASTURE infants raised in farming and non-farming settings we saw clear differences in the two *G(t)* trajectories ($p<0·0001$; appendix p 35). Although the symptom trajectories of the farming group have different starting points, they converge towards a similar endpoint.

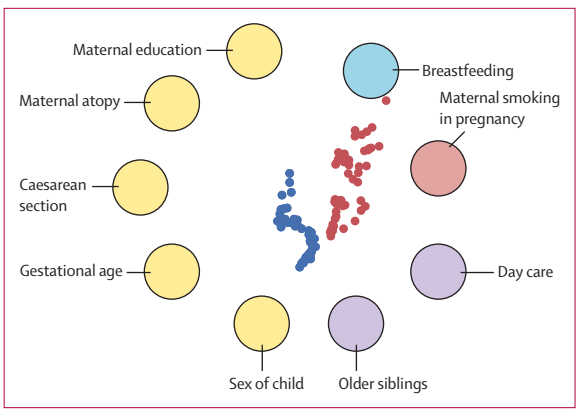

Figure 5: **Different networks for the wheeze and healthy groups in the BILD cohort**
The weekly trajectory function *G(t)* was computed for BILD for the wheeze group $G_{wheeze}(t)$ (red) and for the healthy group $G_{healthy}(t)$ (blue). We identified non-intersecting *G(t)* for both groups. For both groups the gravity points in the second week and in the last week (starting point, marked with a circle, 52nd week by triangle) show a large difference. The trajectory functions, *G(t)*, of the two groups approach over time. By computing the Euclidean distances between the following week of *G(t)*, a simple logistic regression algorithm can distinguish between both groups (PASTURE replication is in the appendix pp 33–35). BILD=Basel-Bern Infant Lung Development. PASTURE=Protection Against Allergy Study in Rural Environments.

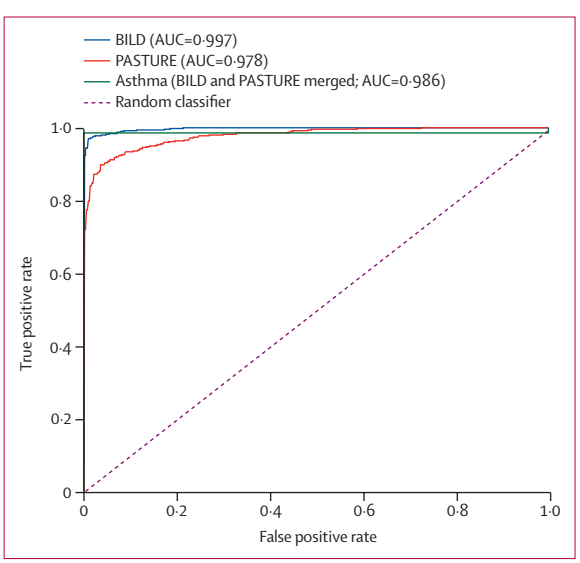

Figure 6: **ROCs for *G(t)* for wheeze at age 2–6 years and asthma**
The *G(t)* pathway shows a remarkable ability to predict the outcome group, shown by the ROCs. The ROCs for the trajectory measure for *G(t)* are presented for the outcome groups wheeze (BILD and PASTURE) and asthma (merged cohorts of BILD and PASTURE). The AUC was very high for both the wheeze outcome groups (0·997 for BILD and 0·978 for PASTURE) and the asthma outcome group (0·986). Each ROC analysis was performed by randomly taking 70% of the population from the healthy group and 70% from the wheeze or asthma group and computing the trajectory measure for *G(t)*. This process was repeated 500 times. AUC=area under the curve. BILD=Basel-Bern Infant Lung Development. PASTURE=Protection Against Allergy Study in Rural Environments. ROC=receiver operating curve.

Discussion

Infants with known early-life risk factors have a higher risk of wheezing disorders at school age[9,25] and, later in life, of respiratory morbidity and mortality.[26] Typically, the relative impact of each individual factor is small.[10] We present novel evidence that the impact of several early-life determinants for subsequent wheeze or asthma show temporal changes during the first year of life. These dynamic changes are complex and act in a network-type manner on the respiratory system. The resulting correlations between early-life determining factors and the respiratory system in infancy can be characterised and quantified with simple parameters such as the CSS and the trajectory function *G(t)* representing dynamic network characteristics. *G(t)* can be considered as a composite proxy for the dynamic interaction of the respiratory system with the environment during infancy, given a set of host factors. While the individual classifier CSS is moderately related, the group classifier *G(t)* is strongly related to subsequent wheeze and asthma.

Many epidemiological birth cohort studies use multivariate regression analysis to determine the effect of early-life risk factors on respiratory outcome in later age. However, according to our first hypothesis from *The Lancet's* Commission on asthma[7] that the dynamic interactions between risk factors during development are important and not the risk factors per se, such methods are limited by how they consider these complex interactions. We used novel self-explaining machine learning methods[14] to better capture these complex interactions. These methods confirmed that the week-by-week CSS (figure 2; appendix p 14) discriminated best between the subsequent wheeze group and healthy group, whereas for the asthma outcome maternal host factors dominated the comparison. These findings were no trivial mathematical phenomenon, because the effect disappeared in a sensitivity analysis using randomisation methods. Thus, these findings strongly support evidence that, even during the first year of life, the respiratory system's characteristics change differently in infants with subsequent wheeze or asthma. We also observed a change in slope of the group means of CSS after around 20 weeks of life. Thus, we hypothesised that this temporal pattern is likely related to the temporal change of the relative importance of the individual determining factors.

In both cohorts, we observed that the impact of some factors on the CSS during the first year of life changed with increasing age (eg, breastfeeding and farming environment impact decreases, and the impact of child care and siblings increases, in line with typical child-drearing behaviours). For some factors no significant linear correlation was found over the first year of life. Sensitivity analysis confirmed true age-dependent temporal changes in correlation profiles (appendix p 30). We further found that the overall correlation between risk factors and CSS at a given week can be represented in the DHECN, characterised by its centre of gravity $G(t)$ for each week, t.

We found that not only the CSS but also the dynamic changes in the network trajectory function $G(t)$ were different in the subsequent wheeze and asthma groups. Quantitative analysis showed that $G(t)$ was highly discriminating between the two groups.

$G(t)$, which best represents the respiratory system's dynamic and adaptive response to the environment, is a group classifier and cannot be used as an individual clinical biomarker. However, the temporal behaviour of CSS also inherits indirect characteristics of this dynamic and adaptive response. Its dynamically increasing relative importance for the prediction of subsequent wheeze (and asthma) in comparison to the other risk factors during the first year of life can be seen in figure 3 and the appendix (pp 25–28). This machine learning analysis directly compares CSS with other risk factors in a dynamic manner. The findings are consistent with the primary hypothesis of the relative importance of dynamic adaptive processes in infancy in comparison to other risk factors (*The Lancet* Commission on asthma).[7] Because we observed temporal changes in CSS and $G(t)$ in the first year of life, we propose a first model hypothesis that the dynamic changes of the correlations (relative impact) between determining factors and respiratory symptoms could be understood as an adaptive process. We hypothesise that this adaptive process was already different in infants with subsequent wheeze in comparison to the healthy group. Our data suggest that it is probably a dynamic integrating process involving both host and environment factors. The CSS could be seen as a proxy for a dynamic integration process, whereby the CSS predictor inherits the composite cumulative network-type impact of the known and, potentially, unknown risk factors.

Our second hypothesis is that this adaptive dynamic process during the first year of life might be more relevant for subsequent wheeze than the individual risk factors alone, as shown by the highly discriminative power of $G(t)$ and CSS dynamics.

Our third hypothesis concerns the contextual aspects of the DHECN network in this adaptive process in infancy. We found clear differences in the dynamic behaviour of $G(t)$ in the group of infants between BILD and PASTURE (eg, the relative contribution of sex, breastfeeding, maternal education, and farming environment). Within PASTURE, the DHECN representation showed differences in the dynamic properties of the trajectory $G(t)$ in infants born on farms.

CSS can be used as an individual risk classifier for clinical purposes already in the first year of life. CSS can easily be assessed by parents using diaries or in an e-health and telemonitoring setting. As a cumulative biomarker, CSS is robust in terms of observation intervals—it works for complex (eg, BILD) and simpler symptom score schemata (eg, PASTURE) and for overall symptoms as well as severe symptoms alone. The discriminative power of CSS for identifying the risk of individual infants is comparable to other risk scores described in the literature (AUC of various methods: 0·66–0·87).[27] In the direct comparison of the different risk factors using self-explaining machine learning the relative importance of the CSS dynamically increased week-by-week in the first year of life and is the dominant determinant at the end of infancy. Although all known risk factors (including CSS) in our study have only a modest sensitivity and specificity to identify individual infants with subsequent wheeze and asthma; on a group level, it becomes clearer that dynamic network-type interactions (represented by $G(t)$) between respiratory system and host and environmental risk factors are very important. $G(t)$ is useful in a clinical and research setting for identifying groups of infants at risk of later wheeze and asthma. $G(t)$ is

suitable for detecting the effect of dynamically changing risk factors during development or for detecting network-type complex interactions between risk factors and the respiratory system. Based on *G(t)*, future studies could develop novel statistical and algorithmic methods to quantify the risk profile of mixed groups of infants or prognostic individual patient risk in research settings—eg, to guide clinical trials, preventive and therapeutic strategies, and to develop clinically relevant early-life biomarkers. In real-life clinical practice, the machine learning findings suggest that clinicians treating infants in the first weeks of life can identify those at risk of subsequent wheeze or asthma by considering factors such as breastfeeding, caesarean section, and maternal atopy. For older infants, clinicians can be better guided by cumulative symptoms.

Our study has some limitations. The analyses are explorative in nature and aim to raise hypotheses for future studies. We used subsequent wheeze as our outcome, albeit a heterogeneous entity, but it was chosen due to sample size considerations. Future prospective studies with sufficiently large data samples (of both healthy and wheeze or asthma populations) would then strengthen our results, helping to further explore the importance of the DHECN in infancy for more specifically defined asthma outcomes. The dynamic behaviour of the correlation networks depends on the choice of risk factors. Our choice in this proof-of-concept study was based on previous publications but theoretically, other determining factors could have been chosen (eg, prematurity, pet exposure, and preventive or therapeutic measures). Future studies might even include asthma-related biomarkers (eg, genetic polymorphism). Furthermore, the ordering of the risk factors within the network circle might affect both the graphical representation and the two-dimensional statistical analysis. Additionally, we chose a high resolution of symptom scores during the first year of life, assessed prospectively with resource-intensive interviews. To increase feasibility for clinical prediction tools, future studies need to show whether similar dynamic effects can be obtained with reduced observation or home-monitoring telemetric devices. Additionally, even though we could replicate our results using PASTURE, some of the effects appear later than in BILD (towards the end of the first year of life), which might happen because of the differing symptom scoring systems and the different starting points for symptom data collection. Another point of limitation is the small differences between the parameter and outcome definitions (eg, maternal education and asthma). Finally, we compared *G(t)* on only pure groups—ie, in individuals who have only wheeze, only asthma, or are only healthy. Analysis of mixed groups and individuals would require a larger sample size of wheeze or asthma population.

These novel findings highlight the importance of the dynamic development of the respiratory system and correlation networks in the context of a given environment for subsequent disease evolvement. We propose a paradigm shift in the understanding of the evolvement of persisting disease in early life, thus we propose novel mathematical methods to quantify the dynamic development of disease. For the first time we show evidence that such quantitative developmental network measures in infancy can effectively be used to predict the risk groups of infants with subsequent wheeze and asthma. Although gene–environment–time models have been described in the literature,[7,10] such quantitative predictors open a novel field of dynamic network physiology and developmental disease research. In future studies, the proposed dynamic correlation network analysis might help to investigate and quantify such temporal disease evolvement of any chronic disease, even with different sets of clinically relevant determining factors (eg, genes and preventive or therapeutic measures) and biomarkers.

Basel Bern Infant Lung Development (BILD) study group
Basel Fiona Beck, Fabienne Decrue, Urs Frey, Amanda Gisler, Olga Gorlanova, and Jakob Usemann. *Bern* Oliver Fuchs, Insa Korten, Philipp Latzin, Marc-Alexander Oestreich, Yasmin Salem, Ruth Steinberg, Andràs Soti, and Sophie Yammine.

Protection Against Allergy Study in Rural Environments (PASTURE) study group
Juha Pekkanen, Pirkka V Kirjavainen, Marjut Roponen, Martin Täubel, Ophélie Ritter, Marie-Laure Dalphin, Cindy Baring, Johanna Theodorou, Andreas Böck, Sonali Pechlivanis, Markus Ege, Martin Depner, Sabina Illi, Francesco Foppiano, Claudia Beerweiler, Guilia Pagani, Remo Frei, and Josef Riedler.

Contributors
UN: conceptualisation, formal analysis, investigation, methodology, project administration, resources, software, supervision, data validation, visualisation, and writing (original draft, review, and editing). OG: data curation and writing (review and editing). FD: data curation and writing (review and editing). HO: formal analysis. ED-E: methodology and writing (review and editing). AB: data curation and writing (review and editing). SS: data curation and writing (review and editing). PL: data curation, funding acquisition, project administration, resources, and writing (review and editing). BS: data curation and writing (review and editing). AMK, RL, AD-C, and CR: data curation. EvM: data curation and writing (review and editing). UF: conceptualisation, data curation, funding acquisition, investigation, methodology, project administration, resources, supervision, and writing (original draft, review, and editing). All co-authors have critically reviewed the manuscript, had full access to the study data, and have approved and accept responsibility for the final manuscript. UN, HO, and SI accessed and verified the data.

Declaration of interests
PL reports board membership, funding grants, and speaking and lecture fees from Vertex and OM Pharma; speaking and lecture fees from Vifor Pharma Switzerland; and board membership with Polyphor, Santhera Pharmaceuticals Schweiz, Allecra Therapeutics, and Sanofi. AD-C reports consulting fees from Sanofi, Stallergens, ALK, and Aimmune Therapeutics; speaking fees and meeting attendance costs from Novartis and ALK; stock or stock options with Essilor Luxottica; a grant from Novartis and ARAIRLOR for asthma and cough research; received a grant from Don du Souffle, Fondation du Souffle, and ARAIRLOR for the PASTURE research; and received a contract with the French public agency ANSES. AMK reports that payments were made to their institution from the Juho Vanio Foundation, the Päiwikki and Sakari Sohlberg Foundation, the Finnish Cultural Foundation, the Academy of Finland, and the Finnish Institute for Health and Welfare.

RL reports a research grant from the Kühne Foundation; and advisory board and lecture fees from ALK, Pfizer, Milupia, VIFOR, and Sanofi. BS reports research grants from Deutsche Forschungsgemeinschaft and consulting fees from GSK, Novartis, and Sanofi. EvM reports funding grants from the PASTURE study, the EFRAIM study, the FORALLVENT study, and the Leibniz Prize; research grants from the German Federal Ministry of Education and Research, the Bavarian State Ministry of Health and Care, OM Pharma, and the European Research Council; royalties or licences from Elsevier, Georg Thieme Verlag, Springer-Verlag, and Springer Nature Group; consulting fees from the Chinese University of Hong Kong, the European Commission, AstraZeneca, Imperial College London, and OM Pharma; lecture, speaker, or educational fees from ALK-Abello Arzneimittel, Japanese Society of Pediatric Allergy and Clinical Immunology, Klinikum Rechts der Isar, University of Colorado, Paul-Martini-Stiftung, AstraZeneca, Imperial College London, Children's Hospital Research Institute of Manitoba Kompetenzzentrum für Ernährung (Kern), OM Pharma, Swedish Pediatric Society for Allergy and Lung Medicine, Chinese College of Allergy and Asthma, Abbott Laboratories, Deutscher Apotheker Verlag, Socieded Chilena de Enfermedades Respiratorias, Japanese Society of Allergology,British Society for Asthma and Clinical Immunology, American Academy of Allergy, Asthma & Immunology, and the European Respiratory Society (ERS); meeting attendance support from Deutsches Zentrum für Lungenforschung, Fraunhofer ITEM Hannover, MCCA Institut für Immunologie Uni Wien, Karl-Landsteiner Privatuniversität für Gesundheitswissenschaften, Swiss Institute of Allergy and Asthma Research, Davos (Associated Institute of the University of Zurich), Medizinische Hochschule Hannover, Natasha Allergy Research Foundation, Deutsche Forschungsgemeinschaft, Gordon Research Conferences, Socieded Chilena de Enfermedades Respiratorias, Arla, Universität Leiden,OM Pharma, American Academy of Allergy, Asthma & Clinical Immunology, Deutsche Forschungsgemeinschaft, ERS, Deutsche Gesellschaft für Kinder- und Jugendmedizin, World Allergy Organization, and European Parliament; has a patent pending (PCT/EP2019/085016) for barn dust extract for the prevention and treatment of diseases; has royalties paid to ProtectImmun for a patent (EP2361632 for specific environmental bacteria for the protection from and/or the treatment of allergic, chronic inflammatory and/or autoimmune disorders, granted on March 19, 2014); has patents licensed to ProtectImmun (EP1411977 for composition containing bacterial antigens used for the prophylaxis and the treatment of allergic diseases, granted on April 18, 2007; EP1637147 for stable dust extract for allergy protection granted on Dec 10, 2008; and EP1964570 for pharmaceutical compound to protect against allergies and inflammatory diseases, granted on Nov 21, 2012); has a patent (EP21189353.2) for proteins identified from barn dust extract for the prevention and treatment of diseases; has a patent (PCT/US2021/016918) for therapeutic fractions and proteins from asthma-protective farm dust; has a patent (EP21189353.2) for proteins identified from barn dust extract for the prevention and treatment of diseases; and is a member of the EXPANSE (funded by European Commission) Scientific Advisory Board, the BEAMS External Scientific Advisory Board, the Editorial Board of *The Journal of Allergy and Clinical Immunology: In Practice*, the Scientific Advisory Board of the Children's Respiratory and Environmental Workgroup, the International Scientific & Societal Advisory Board of Utrecht Life Sciences University of Utrecht, the External Review Panel of the Faculty of Veterinary Science (University of Utrecht), the Selection Committee for the Gottfried Wilhelm Leibniz Programme, the International Advisory Board of Asthma UK Centre for Applied Research, the International Advisory Board of *The Lancet Respiratory Medicine*, the Scientific Advisory Board of the Canadian Healthy Infant Longitudinal Development study (McMaster University, Hamilton, Canada), the Asthma UK Centre for Applied Research, the Pediatric Scientific Advisory Board Iceland, and the Abbott Allergy Risk Reduction Advisory Board. All other authors declare no competing interests.

Data sharing

Data will be made available from the corresponding author upon request with study proposal and completion of a satisfactory data transfer agreement.

Acknowledgments

The Swiss National Science Foundation funds BILD (320030_204717). A European Union research grant and the Kühne Foundation support PASTURE (QRLT4-CT 2001-00250, KBBE-2-2-06). PASTURE was also supported by an EFRAIM study EU grant (211911), a FORALLVENT study EU grant (31708), and the Leibniz Prize (2013 to EvM). The authors thank the study parents and children for their participation, the BILD and PASTURE research teams for their work and collaboration, and Fiona Beck (Universitäts-Kinderspital beider Basel) for medical writing.

References

1 Régnier SA, Huels J. Association between respiratory syncytial virus hospitalizations in infants and respiratory sequelae: systematic review and meta-analysis. *Pediatr Infect Dis J* 2013; **32:** 820–26.

2 Xue M, Dehaas E, Chaudhary N, O'Byrne P, Satia I, Kurmi OP. Breastfeeding and risk of childhood asthma: a systematic review and meta-analysis. *ERJ Open Res* 2021; **7:** 00504-2021.

3 Owora AH, Zhang Y. Childhood wheeze trajectory-specific risk factors: a systematic review and meta-analysis. *Pediatr Allergy Immunol* 2021; **32:** 34–50.

4 Burke H, Leonardi-Bee J, Hashim A, et al. Prenatal and passive smoke exposure and incidence of asthma and wheeze: systematic review and meta-analysis. *Pediatrics* 2012; **129:** 735–44.

5 Makrinioti H, Hasegawa K, Lakoumentas J, et al. The role of respiratory syncytial virus- and rhinovirus-induced bronchiolitis in recurrent wheeze and asthma—a systematic review and meta-analysis. *Pediatr Allergy Immunol* 2022; **33:** e13741.

6 Dai R, Miliku K, Gaddipati S, et al. Wheeze trajectories: determinants and outcomes in the CHILD cohort study. *J Allergy Clin Immunol* 2022; **149:** 2153–65.

7 Pavord ID, Beasley R, Agusti A, et al. After asthma: redefining airways diseases. *Lancet* 2018; **391:** 350–400.

8 Latzin P, Frey U, Roiha HL, et al. Prospectively assessed incidence, severity, and determinants of respiratory symptoms in the first year of life. *Pediatr Pulmonol* 2007; **42:** 41–50.

9 Usemann J, Xu B, Delgado-Eckert E, et al. Dynamics of respiratory symptoms during infancy and associations with wheezing at school age. *ERJ Open Res* 2018; **4:** 00037-2018.

10 Agustí A, Melén E, DeMeo DL, Breyer-Kohansal R, Faner R. Pathogenesis of chronic obstructive pulmonary disease: understanding the contributions of gene-environment interactions across the lifespan. *Lancet Respir Med* 2022; **10:** 512–24.

11 Kyvsgaard JN, Chawes BL, Horner DLG, et al. Risk factors and age-related patterns of asthma-like symptoms in early childhood. *J Allergy Clin Immunol Pract* 2023; **11:** 1773–84.

12 Fuchs O, Latzin P, Kuehni CE, Frey U. Cohort profile: the Bern infant lung development cohort. *Int J Epidemiol* 2012; **41:** 366–76.

13 von Mutius E, Schmid S. The PASTURE project: EU support for the improvement of knowledge about risk factors and preventive factors for atopy in Europe. *Allergy* 2006; **61:** 407–13.

14 Baniecki H, Kretowicz W, Piatyszek P, Wisniewski J, Biecek P. dalex: responsible machine learning with interactive explainability and fairness in Python. *J Mach Learn Res* 2021; **22:** 1–7.

15 de Gouveia Belinelo P, Collison AM, Murphy VE, et al. Maternal asthma is associated with reduced lung function in male infants in a combined analysis of the BLT and BILD cohorts. *Thorax* 2021; **76:** 996–1001.

16 Decrue F, Gorlanova O, Salem Y, et al. Increased impact of air pollution on lung function in preterm versus term infants: the BILD study. *Am J Respir Crit Care Med* 2022; **205:** 99–107.

17 Gorlanova O, Thalmann S, Proietti E, et al. Effects of breastfeeding on respiratory symptoms in infancy. *J Pediatr* 2016; **174:** 111–17.e5.

18 Latzin P, Frey U, Armann J, et al. Exposure to moderate air pollution during late pregnancy and cord blood cytokine secretion in healthy neonates. *PLoS One* 2011; **6:** e23130.

19 Salem Y, Oestreich MA, Fuchs O, et al. Are children born by cesarean delivery at higher risk for respiratory sequelae? *Am J Obstet Gynecol* 2022; **226:** 257.e1–e11.

20 Soti AL, Usemann J, Schaub B, Frey U, Latzin P, Fuchs O. Can biomarkers in umbilical cord blood predict atopic disease at school age? *Pediatr Res* 2021; **89:** 389–92.

21 Usemann J, Decrue F, Korten I, et al. Exposure to moderate air pollution and associations with lung function at school-age: a birth cohort study. *Environ Int* 2019; **126:** 682–89.

22 Mallol J, Crane J, von Mutius E, Odhiambo J, Keil U, Stewart A. The International Study of Asthma and Allergies in Childhood (ISAAC) phase three: a global synthesis. *Allergol Immunopathol (Madr)* 2013; **41:** 73–85.

23 Bräm DS, Nahum U, Atkinson A, Koch G, Pfister M. Evaluation of machine learning methods for covariate data imputation in pharmacometrics. *CPT Pharmacometrics Syst Pharmacol* 2022; **11:** 1638–48.

24 Pujianto U, Wibawa AP, Akbar MI. K-nearest neighbor (k-NN) based missing data imputation. 2019 5th International Conference on Science in Information Technology (ICSITech); Oct 23–24, 2019, 83–88.

25 van Meel ER, Mensink-Bout SM, den Dekker HT, et al. Early-life respiratory tract infections and the risk of school-age lower lung function and asthma: a meta-analysis of 150 000 European children. *Eur Respir J* 2022; **60:** 2102395.

26 Allinson JP, Chaturvedi N, Wong A, et al. Early childhood lower respiratory tract infection and premature adult death from respiratory disease in Great Britain: a national birth cohort study. *Lancet* 2023; **401:** 1183–93.

27 Kothalawala DM, Kadalayil L, Weiss VBN, et al. Prediction models for childhood asthma: a systematic review. *Pediatr Allergy Immunol* 2020; **31:** 616–27.