Modeling of Temporal Exposure to the Ambient Environment and Eczema Severity

Bjorn R. Thomas[1,2], Xiang L. Tan[1], Shagayegh Javadzadeh[1], Elizabeth J. Robinson[2], Bryan S. McDonald[1,2], Malvina A. Krupiczojc[1,2], Syedia R. Rahman[1,2], Samiha Rahman[1,2], Rehana A. Ahmed[1,2], Rubina Begum[1,2], Habiba Khanam[1,2], David P. Kelsell[1], Jonathan Grigg[1,2], Robert J. Knell[3] and Edel A. O'Toole[1,2]

Atopic eczema is a common and complex disease. Missing genetic hereditability and increasing prevalence in industrializing nations point toward an environmental driver. We investigated the temporal association of weather and pollution parameters with eczema severity. This cross-sectional clinical study was performed between May 2018 and March 2020 and is part of the Tower Hamlets Eczema Assessment. All participants had a diagnosis of eczema, lived in East London, were of Bangladeshi ethnicity, and were aged <31 years. The primary outcome was the probability of having an Eczema Area and Severity Index score > 10 after previous ambient exposure to commonly studied meteorological variables and pollutants. There were 430 participants in the groups with Eczema Area and Severity Index ≤ 10 and 149 in those with Eczema Area and Severity Index > 10. Using logistic generalized additive models and a model selection process, we found that tropospheric ozone averaged over the preceding 270 days was strongly associated with eczema severity alongside the exposure to fine particles with diameters of 2.5 μm or less (fine particulate matter) averaged over the preceding 120 days. In our models and analyses, fine particulate matter appeared to largely act in a supporting role to ozone. We show that long-term exposure to ground-level ozone at high levels has the strongest association with eczema severity.

INTRODUCTION

Eczema is the most common inflammatory skin disease (Karimkhani et al., 2017). Large population studies have shown that the prevalence of eczema is rising in developing countries as they industrialize, suggesting an environmental effect (Odhiambo et al., 2009; Williams et al., 2008).

Meteorological factors and ambient air pollution have been reported to be associated with the severity and prevalence of eczema (Ahn, 2014; Kathuria and Silverberg, 2016; Kim et al., 2017; Langan et al., 2006; Silverberg et al., 2013). Studies have shown that fine particles with

[1]Centre for Cell Biology and Cutaneous Research, Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom; [2]Department of Dermatology, The Royal London Hospital, Barts Health NHS Trust, London, United Kingdom; and [3]School of Biological and Behavioural Sciences, Queen Mary University of London, London, United Kingdom

Correspondence: Bjorn R. Thomas, Centre for Cell Biology and Cutaneous Research, Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, 4 Newark Street, London E1 2AT, United Kingdom. E-mail: b.r.thomas@qmul.ac.uk

Abbreviations: AIC, Akaike Information Criterion; EASI, Eczema Area and Severity Index; EseC, European Socio-Economic Classification; GAM, generalized additive model; IGA, Investigators Global Assessment; MAv, moving average; NO, nitric oxide; NO_2, nitrogen dioxide; NO_x, nitrogen oxide; O_3, ozone; PM, particulate matter; SCORAD, SCORing Atopic Dermatitis; SE, standard error; THEA, Tower Hamlets Eczema Assessment; VOC, volatile organic compound

diameters of 10 μm or less (particulate matter [PM]$_{10}$) are associated with a decrease in prevalence (Kathuria and Silverberg, 2016), whereas others have shown these to be associated with an increase in disease incidence (Belugina et al., 2018).

With respect to pollution, it has become clear that fine particles with diameters of 2.5 μm or less ($PM_{2.5}$) and tropospheric (ground-level) ozone (O_3) cause the majority of mortality and disease related to pollution (GBD 2013 Risk Factors Collaborators et al., 2015; Lim et al., 2012). $PM_{2.5}$ is not a single pollutant but a mixture of many chemical species, including black carbon emitted directly from fuel combustion. Combustion of fossil fuels contributes a large amount to $PM_{2.5}$ concentrations (Chow and Watson, 2002). A complicating feature when studying the effect of pollution on disease is that the composition of pollutants varies between countries and cities, particularly $PM_{2.5}$, making comparisons difficult (Li et al., 2019b).

Ground-level O_3 is created by photolytic reactions between volatile organic compounds (VOCs) and nitrogen oxides (NO_x), mainly nitric oxide (NO) and nitrogen dioxide (NO_2). In a system without VOCs, NO_2 and oxygen combine to create O_3 and NO. O_3 in turn reacts with NO to recreate NO_2 in a balanced cycle with no net increase in O_3 levels (Zhang et al., 2019). This balanced cycle depends on stable solar intensity, temperature, and a constant ratio of available NO_2 and NO. In reality, the situation is more complicated because VOCs continuously arise from things such as personal care products, pesticides (McDonald et al., 2018), and even isoprene released from trees (Fehsenfeld et al., 1992). These compounds can

form free radicals, which compete with O_3 in its reaction with NO, resulting in increasing O_3 levels. Increasing worldwide temperatures and escalating fuel combustion favor NO_x formation, which also results in surging O_3 production. Interestingly, in urban settings, falling NO_x levels lead to an elevation in O_3 owing to repartitioning of the $NO/NO_2/O_3$ balance (Lee et al., 2020). This was seen during the March 2020 COVID-19 lockdown in London, United Kingdom. The increase in O_3 levels seen during this period was compounded by higher UV levels, temperatures, and biogenic VOC production from plants in Southern England (Fitzky et al., 2019; Lee et al., 2020). As we move away from NO_x emissions, we can expect higher O_3 levels.

It is also known that $PM_{2.5}$ can act as a sink for the free radicals needed for O_3 formation. When PM concentrations fall, urban O_3 levels also increase (Li et al., 2019a; Ma et al., 2016), illustrated by events in China where active efforts were implemented to reduce annual $PM_{2.5}$ and NO_x levels. This has been associated with rising O_3 concentrations over the last few years (Li et al., 2019a; Ma et al., 2016). It is thought that controlling man-made VOCs will help combat this paradoxical O_3 increase (Le et al., 2020; Lee et al., 2020). Interestingly, wind also plays an important role in this system. Despite pollution control measures in one region, pollution, including O_3 from adjacent upwind areas, can still lead to increased mortality rates in the pollution-controlled downwind areas (Dedoussi et al., 2020).

Data exist to support the role of pollutants in the pathogenesis of eczema (Hendricks et al., 2020). In mouse models, diesel exhaust particles drive eczema by activating the aryl hydrocarbon receptor (Hidaka et al., 2017). The aryl hydrocarbon receptor also acts as an O_3 sensor in the skin (Afaq et al., 2009). Activation of the aryl hydrocarbon receptor can induce proinflammatory COX-2 (Lee et al., 2016) and IL-8 (Tsuji et al., 2011) and induce the expression of TRPA1 and TRPV1 in primary afferent nerve channels causing itch (Elitt et al., 2006; Hidaka et al., 2017). O_3 exposure led to the activation of the Jak2/signal transducer and activator of transcription 3 and Akt1/NF-κB pathways in mouse lungs (Mishra et al., 2016). In the skin, these pathways are relevant to eczema pathogenesis (Howell et al., 2019; Rogerson and O'Shaughnessy, 2018), with the Jak/signal transducer and activator of transcription pathway being an important therapeutic target.

The temporal association between individual pollutants, meteorological variables, and eczema is also not fully understood. In this cross-sectional clinical study, we investigated the association of these factors with eczema severity using a cohort of 579 children and young adults of Bangladeshi ethnicity with eczema (Table 1) in East London, United Kingdom. We have previously shown that this population has an increase in variation of the *FLG* gene (Pigors et al., 2018). We investigated the exposure period and value of a pollutant or meteorological variable that has the strongest association with eczema severity. Using logistic generalized additive models (GAMs) and a model selection approach (explained in Materials and Methods), we investigated which of the meteorological and pollutant variables were most influential in the models of eczema severity.

Table 1. Demographics and Clinical Characteristics of THEA Cohort

	EASI ≤ 10 Cohort (n = 430)	EASI > 10 Cohort (n = 149)
Characteristic		
Age		
Median (Q1−Q3)	9.25 (4.70−14.50)	11.60 (6.80−16.50)
Sex		
Female (%)	193 (44.88)	53 (35.57)
Male (%)	237 (55.12)	96 (64.42)
BMI		
Median (Q1−Q3)	17.79 (15.75−21.86)	18.63 (15.72−23.66)
ESEC		
Class 1-3: Higher (%)	141 (32.79)	31 (20.81)
Class 4-6: Middle (%)	71 (16.51)	45 (30.20)
Class 7-9: Working(%)	218 (50.70)	73 (48.99)
EASI		
Median (Q1−Q3)	2.80 (1.13−5.4)	17.95 (13.20−27.60)
TEWL		
Median (Q1−Q3)	10.82 (8.84−13.63)	14.50 (11.53−20.00)
SH		
Median (Q1−Q3)	26.53 (21.33−32.71)	21.33 (14.87−27.10)
Season		
Spring (%)	87 (20.23)	18 (12.08)
Summer (%)	138 (32.09)	46 (30.87)
Autumn (%)	128 (29.77)	46 (30.87)
Winter (%)	77 (17.91)	39 (26.17)
Investigator		
Research nurse (%)	252 (58.60)	36 (24.16)
Clinical fellow (%)	139 (32.33)	95 (63.76)
Consultant (%)	15 (3.49)	6 (4.03)
Other trained (%)	24 (5.58)	12 (8.05)

Abbreviations: BMI, body mass index; EASI, Eczema Area and Severity Index; EASI10, EASI score ≤ or > 10; ESEC, European Socio-economic Classification; n, number or participants; Q1−Q3, 25th and 75th percentiles; SH, skin hydration measured using a corneometer; TEWL, trans-epidermal water loss measured with a Tewameter.

RESULTS
Participants
A total of 579 participants were included in the analysis, with 430 participants in the group with Eczema Area and Severity Index (EASI) ≤ 10 and 149 in the group with EASI > 10 (Table 1).

Models: individual environmental factors
A total of 10 models per environmental factor were run to examine the association of each variable with the severity of eczema. The 10 models include variables matched to the day of participant recruitment and nine increasing moving averages (MAvs) preceding recruitment (Figure 1). For each variable, the MAv threshold model with the lowest Akaike Information Criterion (AIC) was chosen as the top-performing model (Supplementary Tables S1 and 2) (Burnham and Anderson, 2004, 2002). The range in values of the chosen variables over the entirety of the study is shown in Figure 2 and Supplementary Table S3.
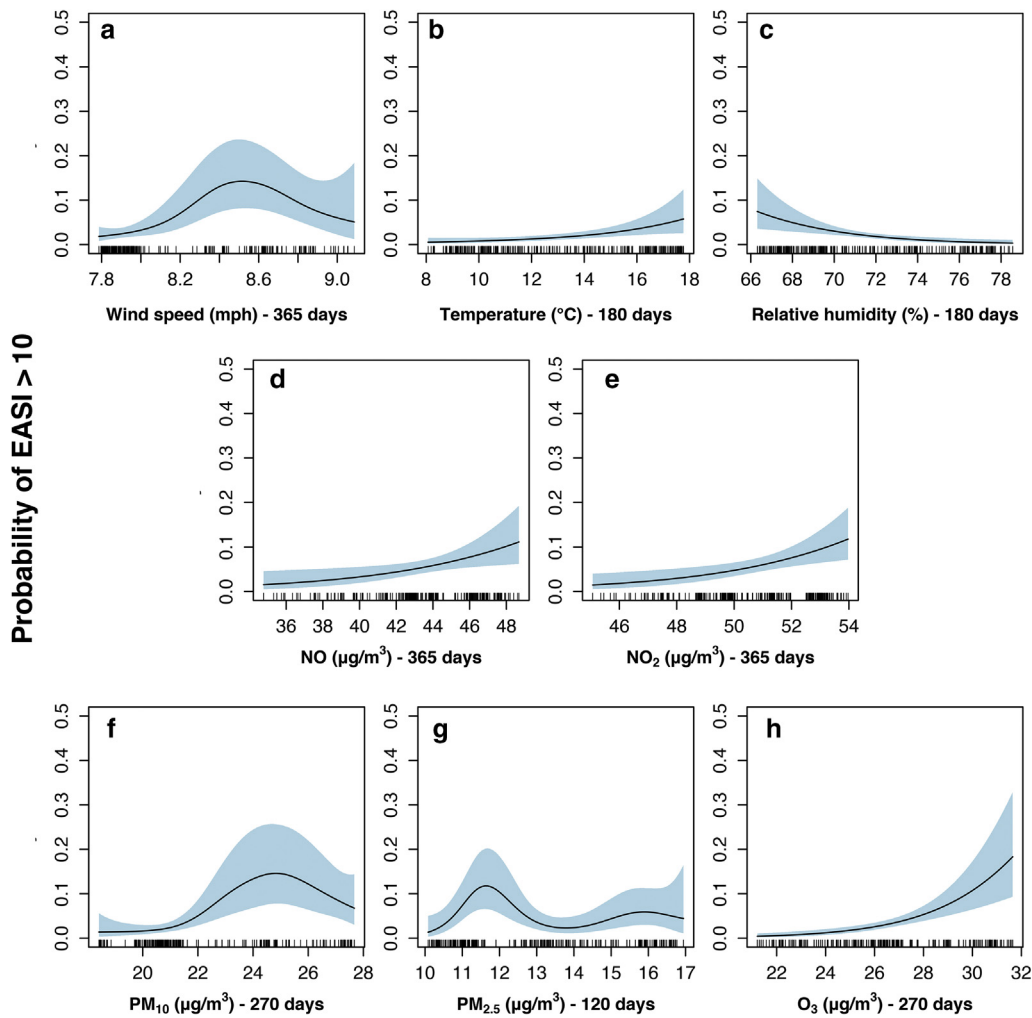
Figure 1. Target variable partial effect plots from the best supported models selected from the 10 model variable set using Δ6 AIC. Probability of having an EASI > 10 (y-axis) versus the mean ambient weather and pollution variables (x-axis): (**a**) Wind 365, Wind speed averaged over 365 days (**b**) Temperature 180, Temperature averaged over 180 days (**c**) Hum 180, Relative humidity averaged over 180 days (**d**) NO 365 , NO averaged over 365 days (**e**) NO_2 365, NO_2 averaged over 365 days (**f**) PM_{10} 270 , PM_{10} over 270 days (**g**) $PM_{2.5}$ 120, $PM_{2.5}$ averaged over 120 days (**h**) O_3 270, O_3 averaged over 270 days. The light blue shaded area represent the 95% confidence intervals. The small dashes overlying the x-axis are known as the rug and they represent all of the cases. Note: Inability to draw a horizontal line throughout the 95% confidence interval indicates the smooth is significant. AIC, Akaike Information Criterion; EASI, Eczema Area and Service Index; NO, nitric oxide; NO_2, nitrogen dioxide; O_3, ozone; $PM_{2.5}$, 2.5 µm particulate matter; PM_{10}, 10µm particulate matter.

The partial effects plots (Figure 1) show that increasing temperature and relative humidity averaged over 180 days were associated with a rise and fall in the probability of having an EASI > 10, respectively. In contrast, increasing wind speed averaged over 365 days and PM_{10} levels averaged over 270 days were associated with a higher probability of EASI > 10. It is interesting to note that the model that is best supported for $PM_{2.5}$ (averaged over 120 days) appears to be mostly associated with an increased probability of EASI >10 at the lower ranges. For O_3 levels (averaged over 270 days), there is an almost linear increase in the probability of EASI > 10 from 0 to about 20%.

Of these models, the O_3 270 (O_3 averaged over 270 days) model, according to the AIC, performed best. The PM_{10} 270 (PM_{10} levels averaged over 270 days) and wind 365 (wind speed averaged over 365 days) were the second and third best-performing models, respectively, with significantly less support (ΔAIC > 7 from the O_3 270 model). All other environmental parameters individually modeled have ΔAIC > 10, indicating very little support when compared with O_3 270.

Top model selection
All environmental variables were included together in a base model to understand which variables associated best with EASI 10 in a more comprehensive system. Backward

selection using ΔAIC of 6 and the nesting rule was used to remove variables with less support, leaving nine models in the top set (Table 2 and Supplementary Figure S5a–h and Supplementary Table S4a–h).

O_3 270 is included as a variable in eight of the nine models in the top set and is only missing from the one with the highest ΔAIC, so the inclusion of this variable is well-supported, in contrast with NO 365/NO_2 365, which is only in two of the top set models and therefore has less support. $PM_{2.5}$ 120 is also present in eight of nine models, but the partial effects were significant only in models 5 and 6. Model 9, with the highest ΔAIC (5.37), contains all environmental variables except O_3 270.

Model 1 performed best in the set, and the partial effects for this model can be seen in Figure 3 and model statistics in Table 3. This model had good discriminatory ability for EASI 10 (area under the receiver operating curve/C-Index: 0.87; 95% CI = 0.83–0.91; $P < 0.001$) (Supplementary Figure S6). In this model, as O_3 270 increases, the probability of EASI >10 increases to above 15%. The nonsignificant $PM_{2.5}$ partial effect in model 1 has a similar shape to the single $PM_{2.5}$ model (Figure 1g), with a bump at lower levels. When we look at the partial effects in the top set, O_3 is significant in all models in which it is contained, including model 1, with an almost linear increase across its range.
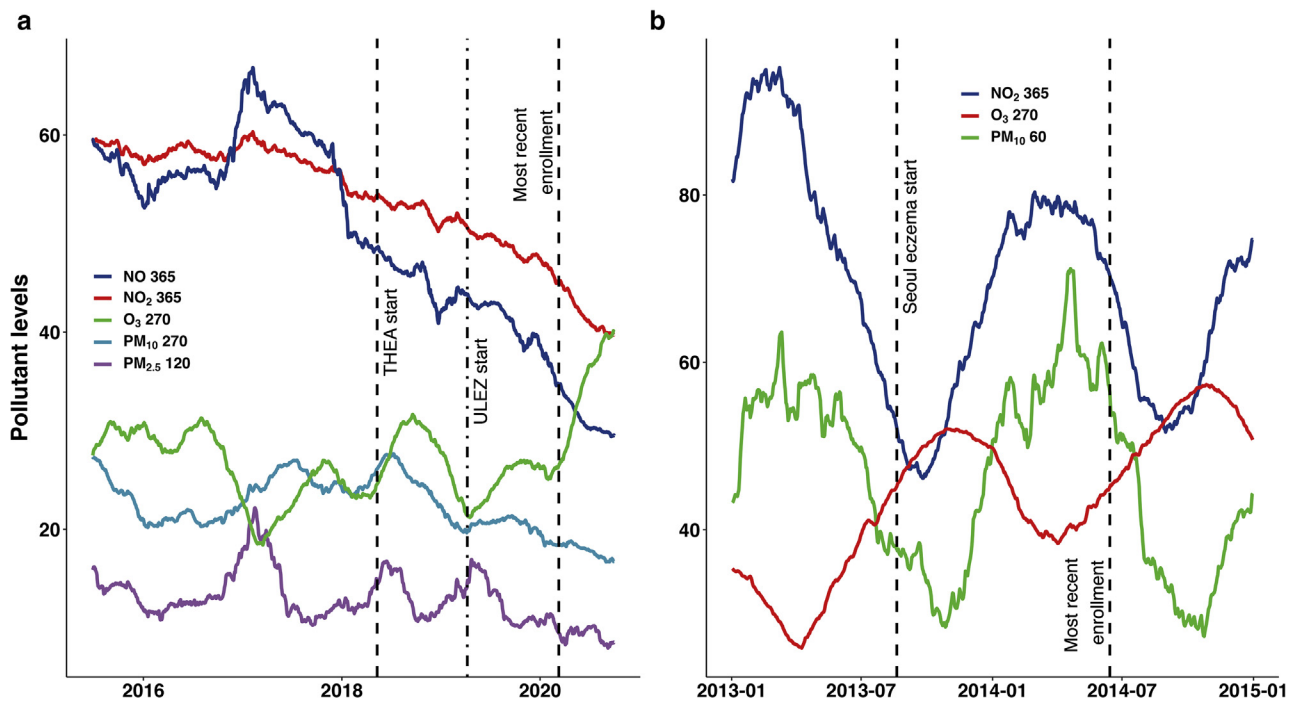
Figure 2. Main pollutant levels during the entirety of the THEA study (a) and Korean dataset (b). (**a**) Moving average line graphs for selected single pollutant models from 28 June 2015 to 30 September 2020. Dashed lines represents the recruitment start date and most recent enrolment. The dash−dot line represents the date in which the ULEZ was introduced in central London. NO 365 − NO levels over the last 365 days; NO_2 365, NO_2 over the last 365 days; O_3 270, Ground level O_3 over the last 270 days; PM_{10} 270, average level of PM_{10} over the last 270 days; $PM_{2.5}$, 2.5μm particulate matter; $PM_{2.5}$ 120, average level of $PM_{2.5}$ over the last 120 days. (**b**) Three available pollutants from Daechi-dong, Gangnam-gu, Seoul, Korea from January 2013 to January 2015. NO_2 365, NO_2 over the last 365 days; O_3 270, Ground level O_3 over the last 270 days; PM_{10} 60, average level of PM_{10} over the last 60 days. *All values in μg/m³. NO, nitric oxide; NO_2, nitrogen dioxide; O_3, ozone; $PM_{2.5}$, 2.5 μm particulate matter; PM_{10}, 10μm particulate matter; THEA, Tower Hamlets Eczema Assessment; ULEZ, Ultra Low Emission Zone.

As mentioned earlier, the NO/NO_2 composite appears only twice in the nine model top set with insignificant partial effects. When modeled individually (Supplementary Table S1), they both have an ΔAIC > 10 when compared with the single O_3 270 model. This gives less support for NO_x playing a direct role in influencing EASI 10.

Of note, model 9 with the highest ΔAIC did not contain O_3 270 but included dimension 1 of principal component analysis between (i) NO 365 and NO_2 365, (ii) PM_{10} 270 and wind 365, (iii) temperature 180 and humidity 180, and (iv) $PM_{2.5}$ 120 alone. It is known that NO, wind speed, humidity, and $PM_{2.5}$ can all affect O_3 levels (Gorai et al., 2015; Kavassalis and Murphy, 2017; Li et al., 2019a). We therefore looked at the raw values for NO 365, wind 365, humidity 180, $PM_{2.5}$ 120, and O_3 270 from June 28, 2015 to September 30, 2020. All variables appeared to be at least partially correlated—mostly inverse (Figure 4a). We trained GAM using default setting and no transformations on environmental data from 28 June 2015 to 1 October 2019 (training set) to see whether changes in NO 365, wind 365, humidity 180, and $PM_{2.5}$ 120 account for changes in the levels of O_3 270 seen over that period. We then used the trained model (adjusted R_2 of 92.10) to predict O_3 270 from 2 October 2019 to 30 September 2020 (test set). Despite the fact that the test data contained values of NO and $PM_{2.5}$ not seen before in the training data, we were able to reasonably predict the dramatic O_3 270 spike during the Spring 2020 United Kingdom COVID-19 lockdown (Figure 4a). This

experiment suggests that in general, changes in NO 365, wind 365, humidity 170, and $PM_{2.5}$ 120 account for O_3 270 levels, therefore explaining the presence of model 9 in the top set. Rather than having individual direct effects on eczema severity (which they still may have), the environmental variables in model 9 may be acting as a surrogate for O_3 270.

In summary, a total of eight models directly include O_3 270 and $PM_{2.5}$, with the ninth (model 9) possibly acting as a surrogate for O_3 270, as described earlier. One of these models only includes O_3 270 as a solitary environmental variable. Finding $PM_{2.5}$ and O_3 in the majority of the top set lend strong support to their role in determining eczema severity. Finally, when looking at the double penalty approach for model selection, this resulted in one model that includes only O_3 270, lending more support for the role of O_3 as a factor in determining eczema severity.

Reproducibility: Korean cohort
A previous paper from South Korea examined the short-term effects of air pollution and meteorological variable on eczema symptoms. Using the AIC, MAv model selection was performed for available pollutants (Figure 2b): PM_{10}, NO_2, and O_3. PM_{10} 60, O_3 270, and NO_2 365 performed best in the MAv model. According to the AIC, for single pollutants, PM_{10} performed the best, followed by NO_2, then O_3 (Supplementary Table S5).

Using the same rules as with the East London dataset, we performed full model selection (Table 4). These models

Table 2. Top Model Set

Model Number	Equation	R^2_{Adj}	AIC		C-Index[1] (95% CI)
Base Model Prior to Backward Selection and Double Penalty Selection					
Base	EASI10 ~ NO 365/NO$_2$ 365 Dim1 + PM$_{10}$ 270/Wind 365 Dim 1 + Temp 180/Hum 180 Dim 1 + PM$_{2.5}$ 120 + O$_3$ 270 + SH + TEWL + Age + Lon:Lat + BMI + Season + Sex + Investigator + EseC class	0.365	485.16		0.87 (0.83−0.91)
Nine Best Supported Models Backward Selection (Δ6 AIC and Nesting Rule)				**ΔAIC**	
1 (4a)	EASI10 ~ PM$_{2.5}$ 120 + O$_3$ 270 + SH + TEWL + Age + BMI + Season + Sex + Investigator + EseC class	0.368	478.17	0	0.87 (0.83−0.91)
2 (5f)	EASI10 ~ PM$_{2.5}$ 120 + O$_3$ 270 + SH + TEWL + Age + Season + Sex + Investigator + EseC class	0.365	478.73	0.56	0.86 (0.82−0.90)
3 (5h)	EASI10 ~ PM$_{2.5}$ 120 + O$_3$ 270 + SH + TEWL + Age + BMI + Season + EseC class	0.360	479.56	1.39	0.87 (0.83−0.97)
4 (5a)	EASI10 ~ O$_3$ 270 + SH + TEWL + Age + BMI + Season + Sex + Investigator + EseC class	0.351	480.39	2.22	0.86 (0.82−0.90)
5 (5g)	EASI10 ~ PM$_{2.5}$ 120 + O$_3$ 270 + SH + TEWL + Age + BMI + Sex + Investigator + EseC class	0.353	480.93	2.76	0.86 (0.82−0.90)
6 (4j)	EASI10 ~ NO 365/NO$_2$ 365 Dim1 + PM$_{2.5}$ 120 + O$_3$ 270 + SH + TEWL + Age + BMI + Season + Sex + Investigator + EseC class	0.363	482.53	4.36	0.87 (0.83−0.90)
7 (5i)	EASI10 ~ PM$_{2.5}$ 120 + O$_3$ 270 + SH + TEWL + Age + BMI + Season + Sex + Investigator + EseC class	0.363	482.55	4.38	0.86 (0.82−0.90)
8 (5j)	EASI10 ~ PM$_{2.5}$ 120 + O$_3$ 270 + SH + TEWL + Age + BMI + Season + Sex + Investigator	0.365	483.16	4.99	0.87 (0.83−0.91)
9 (2e)	EASI10 ~ NO 365/NO$_2$ 365 Dim1 + PM$_{10}$ 270/Wind 365 Dim 1 + Temp 180/Hum 180 Dim 1 + PM$_{2.5}$ 120 + SH + TEWL + Age + BMI + Season + Sex + Investigator + EseC class	0.360	483.54	5.37	0.87 (0.83−0.91)

Abbreviations: AIC, Akaike Information Criterion; BMI, body mass index; CI, confidence interval; ESeC, European socio-econmic classification; Hum 180, average relative humidity over the last 180 days; Investigator, researcher who perform the EASI score per participant; Lon:Lat, longitude and latitude; NO$_2$ 365, average nitrogen dioxide level over last 365 days; NO 365, average level nitric oxide over last 365 days; NO 365/NO$_2$ 365 Dim 1 , Dimension 1 of PCA between NO 365 and NO$_2$ 365; O$_3$ 270, average level of ozone over the last 270 days; PM$_{2.5}$ 120, average level of PM$_{2.5}$ over the last 120 days; PM$_{10}$ 270, average level of PM$_{10}$ over the last 270 days; PM$_{10}$ 270/Wind 365 Dim 1 , Dimension 1 of PCA between PM$_{10}$ 270 and Wind 365; R^2_{Adj}, Adjusted R-squared; SH, skin hydration; Temp 180, average temperature over the last 180 days; TEWL, transepidermal water loss; Wind 365, average wind speed over last 365 days.
[1]C-Index; Concordance statistic, measure of goodness of fit, equal to the area under the Receiver Operating Characteristic curve.

included only the NO$_2$/PM$_{10}$ composite and again O$_3$ 270. Both variables in this model are significant and can be seen in Supplementary Figures S7 and S8. As values of the PM$_{10}$/NO$_2$ composite rise and O$_3$ levels increase, the probability of having a SCORing Atopic Dermatitis (SCORAD) > 30 increases. For O$_3$, at levels >46 µg/m^3, the probability falls. When the O$_3$ partial effect plots from our Tower Hamlets Eczema Assessment (THEA) cohort and the Korean groups were compared, they were almost extensions of each other (Figure 5).

DISCUSSION
Biologic rationale for findings
In this modeling study, we show that prolonged exposure to O$_3$ rather than shorter time periods associate best with the severity of atopic eczema in the Bangladeshi population in East London. Full model selection left us with a top set of nine models, eight of which included O$_3$ as the only environmental variable (plus a ninth, model 9, a possible O$_3$ surrogate). An explanation for the selection of long-term O$_3$ (O$_3$ 270) exposure levels may be due to the protective nature of the lipid-rich layers of corneocytes (stratum corneum), which would include squalene, the single most abundant O$_3$-reactive skin lipid (Wisthaler and Weschler, 2010). Long-term O$_3$ exposure at increased levels may be required to overwhelm the protective mechanisms (Chen et al., 2007; Valacchi et al., 2005). This is in contrast to the lungs, which are known to

react to short exposures to O$_3$ and have simple columnar/cuboidal and thin squamous layers in the bronchioles and alveoli, respectively.

PM$_{2.5}$ appears in eight models, although the raw data do not show a convincing direct relationship (Supplementary Figure S9). The eczema phenotype that develops in mice after exposure to diesel exhaust particles (Hidaka et al., 2017) supports its inclusion in the top set. We do see from our models that PM$_{2.5}$ likely acts as an important O$_3$ regulator. When at low levels, as seen in the PM$_{2.5}$ partial effects plot, there is an increased probability of EASI >10. This is probably due to a reduction in consumption of O$_3$ precursors by available PM$_{2.5}$, with a resultant increase in O$_3$ (Figure 2a). There is an increased probability of EASI > 10 with higher levels of PM$_{2.5}$, particularly noticeable in model 6. It has been shown in mice that lung exposure to ultrafine PM (<0.1 µm) leads to little change, but the addition of O$_3$ results in greater damage, possibly through the degradation of the PM to more volatile compounds (Wong et al., 2018).

Interactions between ambient environmental variables
Model 9 is unusual because it does not contain O$_3$ but retains NO/NO$_2$, PM$_{10}$/wind, temperature/humidity, and PM$_{2.5}$. The presence of this combination of variables in one well-supported candidate model indicates a potential role for other atmospheric pollutants as well as O$_3$. This model support is considerably weaker than for O$_3$, with

Figure 3. Top performing model: Smooth terms (nonlinear parameters) partial effects plots: O_3 and $PM_{2.5}$ versus probability of EASI > 10. $PM_{2.5}$ and O_3 partial effects plot from top performing model - using $\Delta 6$ AIC and nesting rule. (**a**) $PM_{2.5}$ 120, $PM_{2.5}$ averaged over 120 days ($P = 0.126$) (**b**) O_3 270 ($P = <0.001$), O_3 averaged over 270 days. Other partial effects for this model and for other models in the top model set can be seen in the Supplementary materials. Probability of having a EASI > 10 (y-axis) versus the mean ambient weather and pollution variables (x-axis): O_3 270, O_3 over 270 days. The light blue shaded area represent the 95% confidence interval. The small dashes overlying the x-axis are known as the rug and they represent all of the cases. Note: Inability to draw a horizontal line throughout the 95% confidence interval indicates the smooth is significant. In this case $PM_{2.5}$ 120 is not significant. AIC, Akaike Information Criterion; EASI, Eczema Area and Service Index; O_3, ozone; $PM_{2.5}$, 2.5 µm particulate matter.
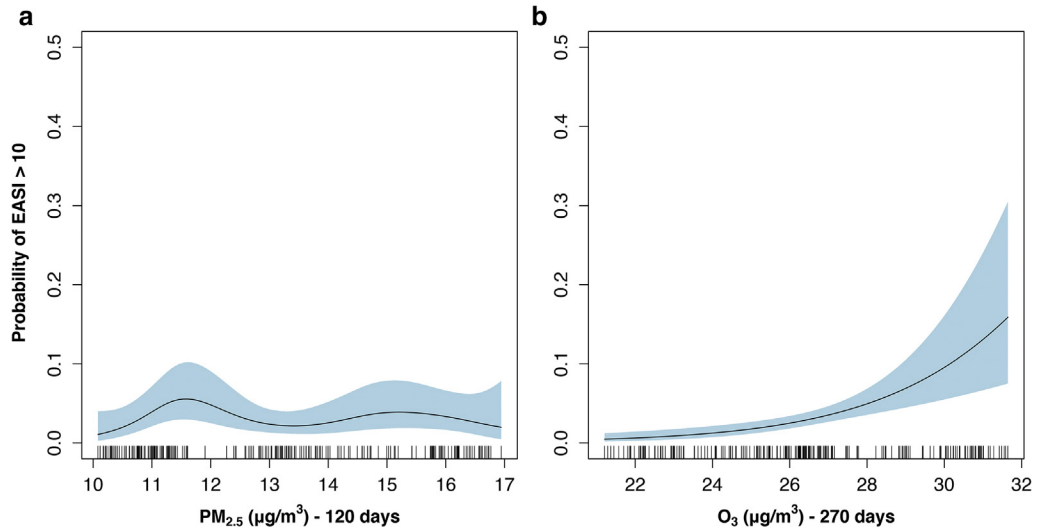
only the temperature/humidity composite showing a significant relationship. A review of the raw data and their relationship with change in EASI score suggests that these variables have less influence (Supplementary Figures S4 and S9 and Figure 2). O_3 levels clearly start to rise in line with the increase in EASI scores seen at the end of the recruitment timeline, unlike other variables.

Interestingly, when we use NO, wind speed, humidity, and $PM_{2.5}$ using the MAv selected by the initial EASI 10 models, we are able to predict trends in O_3 270. It may be that the association that these variables have with eczema severity may not be direct but by their ability to alter O_3 levels, for example, $PM_{2.5}$ reactions with O_3 precursors leading to lower production or NO reacting

Table 3. Top Performing Model: Linear Parameter ORs Versus EASI > 10

Linear Parameters:	OR	R^2_{Adj} / C-Index[1] 95% CI	AIC P-Value
Season: Autumn (Ref)	—	—	—
Season: Spring	4.32	0.98−19.08	0.053
Season: Summer	3.56	1.27−9.97	0.015
Season: Winter	4.17	1.54-11.31	0.005
Female (Ref)	—	—	—
Male	1.62	0.96−2.71	0.068
Investigator 1 (Ref)	—	—	—
Investigator 2	1.86	0.92−3.78	0.085
Investigator 3	0.28	0.06−1.28	0.100
Investigator 4	1.10	0.38−3.22	0.860
ESeC class 1−3 (Ref)	—	—	—
ESeC class 4−6	2.79	1.39−5.58	0.004
ESeC class 7−9	1.39	0.76−2.55	0.282

Abbreviations: 95% CI, 95% confidence interval; AIC, Akaike Information Criterion; ESeC, European Socio-economic Classification; R^2_{Adj}, Adjusted R^2; Ref, reference.
[1]Model performance metrics: R^2 Adj, 0.37; AIC, 478.17; C-Index, 0.871; C-Index; Concordance statistic, measure of goodness of fit, equal to the area under the Receiver Operating Characteristic curve.
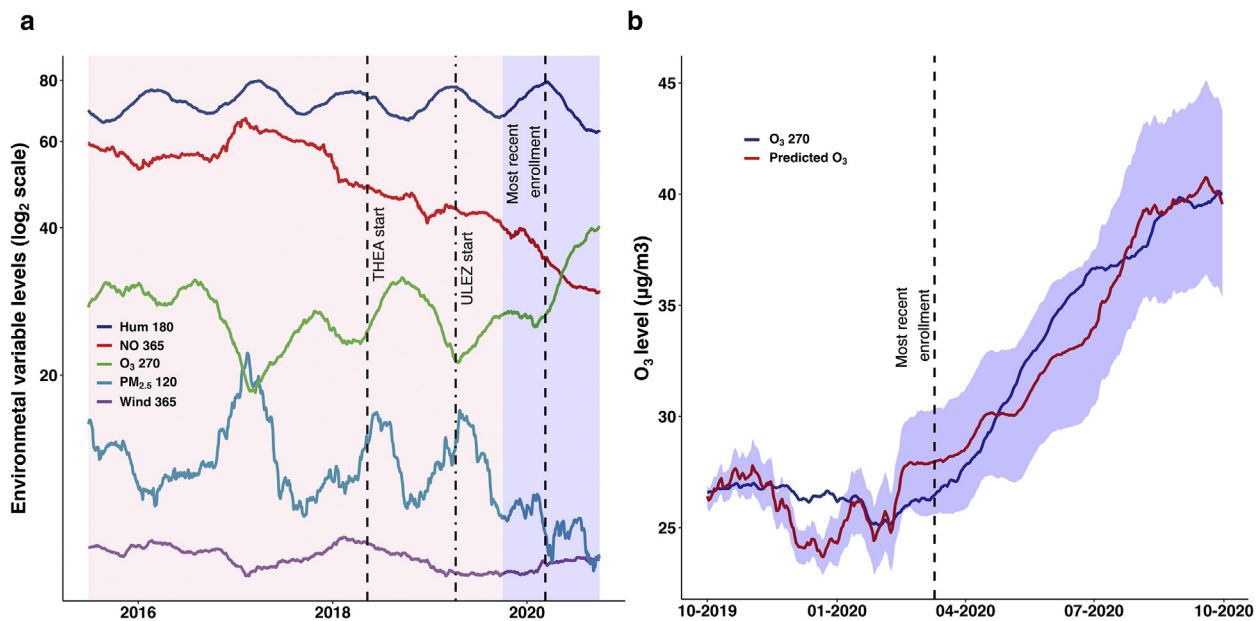
Figure 4. Meteorological and pollutant levels for O₃ 270 prediction. Raw values for GAM training and O_3 270 prediction. (**a**) Environmental values found in Model 9 of the top model set in addition to O_3 270 from 28 June 2015 to 30 September 2020. The pink area represents the data range for the training set (28 June 2015 to 1 October 2019) to predict O_3 270 using NO 365, Hum 180, Wind 365 and $PM_{2.5}$ 120 (adjusted R^2 of 92.10). The blue area represents the test set (2 October 2019 to 30 September 2020) for O_3 270. (**b**) Predicted O_3 270 (red line) with 95% confidence interval in light blue from 2 October 2019 to 30 September 2020. Actual O_3 270 (dark blue) during that period. GAM, generalized additive model; Hum, humidity; NO, nitric oxide; O_3, ozone; THEA, Tower Hamlets Eczema Assessment; ULEZ, Ultra Low Emission Zone.

with O_3 leading to increased consumption and subsequent lower O_3 levels (Li et al., 2019a; Ma et al., 2016).

Interpreting findings for Korean cohort
We also examined a dataset from Seoul, Korea that had eczema severity data for over 14 months using data from 76 monitoring stations in Seoul (Kim et al., 2017). The MAv model that best fits O_3 and NO_2 data was 270 and 365, respectively, the same MAv to those selected in the THEA cohort. Alongside being present in the top set from the Korean cohort, the fact that the O_3 MAv (270 days) was the same across both datasets (Korea and London) and that the O_3 partial effects plots aligned was very interesting, suggesting a relationship between eczema severity and O_3 over a wide range of values from 22 to 32 $\mu g/m^3$ in London and from 38 to 52 $\mu g/m^3$ in Seoul. The recapitulation of similar findings is reassuring for our single station London model.

A review of pollution levels over the Korean study period shows that there was only one peak and trough, unlike the

multiple cycles noted in the THEA cohort (Figure 2). The shorter time period in the Korean cohort is not long enough to reveal the underlying longer-term patterns. It is possible that if the recruitment process was continued through multiple cycles, O_3 may be found to be more influential in the single-pollutant models. We found it surprising that at O_3 270 levels >46 $\mu g/m^3$, the probability of having moderate or severe disease falls. An explanation could be cellular reduction–oxidation adaptions from repeated exposure to high levels of O_3 as shown in keratinocytes that were repeatedly exposed to cold plasma in vitro (the main constituent being O_3) (Schmidt et al., 2016). An adapted response with some tolerance has also been noted in lungs exposed to very high (200 ppb/399.14 $\mu g/m^3$) doses of O_3 (Jörres et al., 2000).

Clinical relevance
Our findings suggest that exposure to low levels of tropospheric O_3 may be the most influential pollutant associated with the severity of eczema. In vitro evidence exists that O_3-

Table 4. Top Model Set: Korea

Model Number	Equation	R^2_{Adj}	AIC		C-Index[1] (95% CI)
Base Model Before Backward Selection and Double Penalty Selection					
Base	SCORAD30 ~ NO₂ 365/PM₁₀ 60 Dim 1 + O₃ 270 + Age + Sex + Season	0.122	228.53		0.76 (0.69−0.83)
Two Best Supported Models Backward Selection (Δ6 AIC and Nesting Rule)				ΔAIC	
3c	SCORAD30 ~ NO₂ 365/PM₁₀ 60 Dim 1 + O₃ 270	0.368	221.66	0	0.74 (0.67−0.83)
1d	SCORAD30 ~ NO₂ 365/PM₁₀ 60 Dim 1 + O₃ 270 + Age + Season	0.365	227.13	5.47	0.76 (0.69−0.83)

Abbreviations: 95% CI, 95% confidence interval; AIC, Akaike Information Criterion; NO₂ 365, average level of nitrogen dioxide over the last 365 days; NO₂ 365/PM₁₀ 60 Dim 1, Dimension 1 of PCA between NO₂ 365 and PM₁₀ 60; O₃ 270, average level of Ozone over the last 270 days; PM₁₀ 60, average level of PM₁₀ over the last 60 days.
[1]C-Index; Concordance statistic, measure of goodness of fit, equal to the area under the receiver operating characteristic curve.
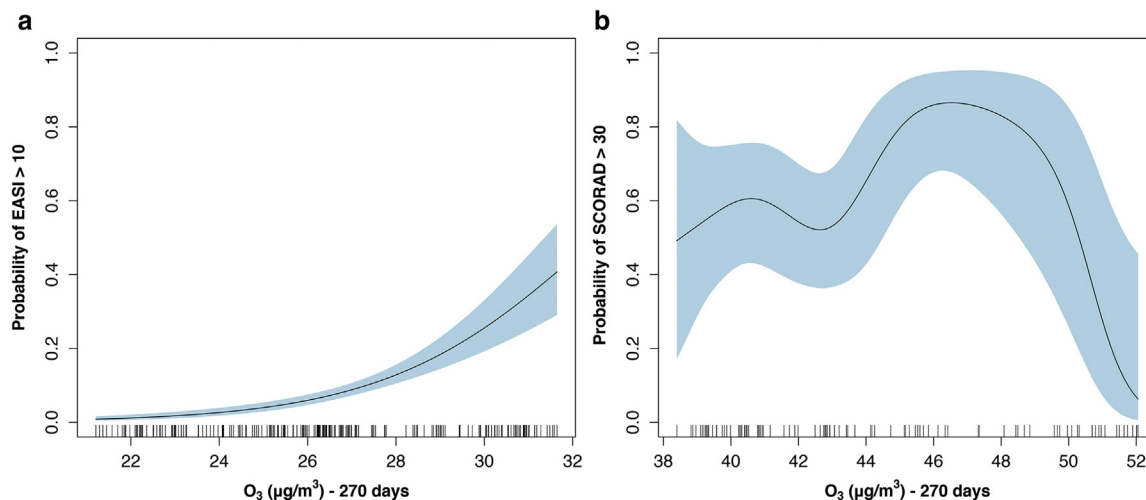
Figure 5. Ground level O₃ average over 270 days − THEA cohort and Kim et al. (2017) Korean Eczema Panel study. O_3 partial effects plot from (**a**) THEA cohort, n = 579 from 11 May 2018 to 10 March 2020 and the (**b**) Kim et al. (2017) Korean Panel study, n = 177 from 20 August 2013 to 14 June 2014. As age, sex and season were the only main features in common between cohorts, adjustments were only made to these features in both groups. Having these plots side by side demonstrate that they are extensions of each other with (**a**) showing lower average ozone levels and (**b**) showing the higher end of the spectrum. Wider 95% confidence interval for (**b**) as fewer cases studied. Probability of having an (**a**) EASI > 10 (y-axis) or a (**b**) SCORAD > 30 versus O_3 270 , O_3 over 270 days (x-axis). The light blue shaded area represent the 95% confidence interval. The small dashes overlying the x-axis are known as the rug and they represent all of the cases. Note: Inability to draw a horizontal line throughout the 95% confidence interval indicates the smooth is significant. EASI, Eczema Area and Severity Index; O_3, ozone; SCORAD, SCORing Atopic Dermatitis severity score; THEA, Tower Hamlets Eczema Assessment.

related oxidative and cytotoxic effects in keratinocytes can be reduced/repaired by the application of vitamin C compounds (Valacchi et al., 2015) and a proprietary snail mucus called HelixComplex (Gentili et al., 2020). This suggests that topical therapy as a preventative measure or treatment for O_3 related flaring could be a new therapeutic approach.

It has been previously shown that pollutants, including O_3, are associated with more visits to the emergency department (Wang et al., 2021) and an increase in eczema symptoms (Kim et al., 2017). If O_3 levels do indeed have the strongest association with eczema severity, it may be that O_3 is also driving flares and exacerbation in symptoms. Finding treatment options, such as those described earlier, to protect from the effects of O_3 should be explored.

Finally, further work is required because epidemiological data suggest that there may be an association between eczema prevalence and O_3 levels. We estimate the prevalence of eczema in Bangladeshi children in East London to be similar to that of the general population at around 16% (Ban et al., 2018; Williams et al., 1995). The rate seen in Bangladesh is 6–12% of children (Ahmed et al., 2010; Kabir et al., 2005). We reviewed published pollution data in Dhaka, Bangladesh from 2013 to 2017. We found that average yearly tropospheric O_3 levels were much lower in Dhaka than that seen in East London: 18.4 (standard error [SE]: ±0.68) µg/m³ versus 26.2 (SE: ±0.06) µg/m³. In addition, the PM_{10} and $PM_{2.5}$ were markedly higher in Dhaka than in London: 154.9 (SE: ±2.59) and 85.8 (SE: ±1.59) µg/m³ versus 22.8 (SE: ±0.09) and 13.5 (SE: ±0.02) µg/m³, respectively.

Furthermore, in Korea, the prevalence of itchy ever eczema in individuals aged 6–7 years has increased from 17 to 27% from the year 2000 to 2010 (Park et al., 2016).

From 1999 to 2016 in Seoul, Korea, there was a 17.5 µg/m³ decrease in PM_{10} and a 17.6 µg/m³ increase in O_3 (Kim and Lee, 2018; Seo et al., 2018). Over this period, solar irradiation, temperature, and humidity remained stable, and only wind speed increased (Seo et al., 2018). This is thought to be one explanation to account for some of the changes in PM_{10} and O_3 (Seo et al., 2018). It is also important to note that $PM_{2.5}$ fell significantly over this period (Kim and Lee, 2018). This reveals interesting information about the relationships between eczema prevalence and the ambient environment. If the relationship between O_3 and eczema is real, we would expect to see more (severe) disease in the future as current ($PM_{2.5}$ and NO_x) pollution reduction strategies are employed and worldwide temperatures increase.

Strengths and limitations
All data were carefully collected with detailed phenotyping, allowing us to accurately classify the disease and severity. The longer duration of the study period allowed the demonstration of longer trends in pollutant variables than single annual cycles. A final strength of the study was using the AIC with a ΔAIC of 6 to select a top model set rather than finding a minimal adequate model using null-hypothesis significance testing, which can lead to inflated type I errors (Harrison et al., 2018). One of the limitations of this study is that participants were only assessed during a single visit. As in any cohort study, some of the initial participants recruited may have had more severe eczema and may have been more likely to join at early stages, leading to a selection bias. This may account for the downward trend in EASI score over the study period—this would not however account for the uptick starting at the end of 2019. When we remove severe and clear patients from the dataset, the shape of severity remained

the same, and outcomes from modeling remained unchanged.

This study uses one roadside monitoring station for its pollution data, and we do not have measured hyperlocal levels. Because O_3 has become a more prominent pollutant, air quality monitoring networks with O_3 measurements have been established in several countries in Europe, North America, Australia, Japan, and South Korea, and efforts have been made to regulate O_3. There are 16 sites in London that record roadside pollution, with only one in East London. This is compared with 76 available stations in Seoul, Korea. We also do not account for indoor pollution, although there is some evidence that outdoor O_3 levels strongly influence indoor levels, with the potential of indoor O_3 levels exceeding those of outdoors (Huang et al., 2019).

In summary, these findings lend support to an important role of O_3 in determining eczema severity in our cohort. The role of $PM_{2.5}$ is important, although the strength of association is smaller and possibly secondary to ambient O_3 levels. Further research on the relationship of ground-level O_3, eczema, and its interplay with $PM_{2.5}$ should be a focus. Mapping biological pathways and working out how to protect the skin from the effects of air pollutants would be a sensible next step to give clarity on how much influence it has on this very common disease. Future work on an analysis of global trends in pollution parameters and trends in the prevalence of eczema should also be undertaken.

MATERIALS AND METHODS
Regulatory approvals
Recruitment for this cross-sectional study occurred between 11 May 2018 and 10 March 2020. Ethical approval was obtained from Health Research Authority, United Kingdom after review by Hampstead Regional Ethics Committee (reference: 18/LO/0018; Research Database Application (ReDA) Reference: 011978). Before recruitment and donation of samples, patients and/or parents gave their written informed consent. We followed the Strengthening the Reporting of Observational Studies in Epidemiology reporting guideline.

Study and subjects
The participants were recruited as part of the THEA project, a study of atopic eczema in the Bangladeshi population in East London. All have a diagnosis of eczema, confirmed by Consultant Dermatologists, and are seen at the Royal London Hospital (London, United kingdom) for their condition. They all live in East London (Supplementary Figure S1) and are aged <31 years. Patients with known congenital recessive and X-linked ichthyoses, equivocal diagnoses of eczema, or mixed ethnicities were excluded from the study.

Outcome
As a measure of severity, we used the EASI score (Schmitt et al., 2014). The primary outcome of this study was the probability of having an EASI score > 10 associated with previous exposure to pollutants and meteorological variables. To aid in the end-user interpretation of our models, we have dichotomized the EASI score into EASI 10 (EASI $\leq$ 10 and EASI > 10), which would translate into clear/mild versus moderate/severe.

We have chosen the EASI 10 cut-off using data previously generated from the cohort. For this, clustering was performed on binary eczema distribution data (e.g., eczema on left extensor elbow? Yes/No) in the first 409 participants in this cohort. K-means clustering using Manhattan distance was used. The number of clusters (k value) was suggested by using the elbow method heuristic using within-cluster sum of squares, and we settled on k = 4. Waikato Environment for Knowledge Analysis machine learning toolkit (version 3.9; University of Waikato, Hamilton, New Zealand) was used for clustering. Binary distribution data for each cluster group were converted into proportional means so that it gave the proportion of participants that had eczema at a particular site within each cluster. These data were then mapped onto body silhouettes to create a body heatmap for each cluster. We see that participants in cluster 1 are clear or almost clear, cluster 2 has the classic flexural disease, cluster 3 has extensor disease, and cluster 4 has extensive lesional coverage fitting with patterns that we see in the clinic (Supplementary Figure S2a).

When plotting EASI score (y-axis) by any variable (Supplementary Figure S2b), we use O_3 averages over 365 days and color participants by cluster membership, and we can fit parsimonious demarcation lines between cluster 1 (clear or almost clear) and clusters 2–4 at an EASI score of 10.

Previous studies have created severity strata (clear, mild, moderate, and severe) for the EASI score (Chopra et al., 2017a; Leshem et al., 2015). The kappa coefficient for gestalt subjective Investigators Global Assessment (IGA) and EASI in the Chopra et al. (2017a) study was 0.69, meaning that 31% of data were not concordant. The EASI range for moderate eczema was classified as 6.0–22.9. When the kappa is manually calculated, using supplementary raw data (Chopra et al., 2017a) for agreement between moderate IGA and EASI between 6.0 and 22.9 versus other EASI/IGA scores, the kappa is 0.57. As the authors state, their severity strata are possible potential thresholds and by no means final. An EASI of 10 falls at the lower end of the moderate strata from the proposed strata in the Chopra et al. (2017a) paper. In a similar paper by Leshem et al. (2015), EASI and IGA have a kappa of 0.75, with the range of EASI of 7.1–21.1 for moderate disease. In the Lesham paper, there is significant overlap between EASI and mild and moderate IGA, with an approximate minimum, maximum, and median (25th and 75th percentiles) of 2, 10, and 5 (3–6), respectively, for EASI in the mild IGA group. For moderate IGA, the minimum, maximum, and median (25th and 75th percentiles) EASI scores are approximately 3, 24, 10 (7–14), respectively. This is probably where their 18% misclassification was seen (raw data not available). With this in mind, the EASI severity strata, particularly for moderate disease, although accepted, is not yet clear.

We therefore classify that an EASI $\leq$ 10 is clear/mild and that an EASI > 10 is moderate or severe without the need for using the nonstandardized IGA as an anchor (Futamura et al., 2016).

We do not have a concern about dichotomizing the EASI score into clear/mild and moderate or severe because it is standard practice in dermatology to separate these groups apart for access to clinical trials and treatment. Intermediate inter-rater reliability in EASI scores (Schmitt et al., 2007), as seen in the seminal Hanifin et al. (2001) paper where one rater was excluded from the analysis to improve the results, still displays large differences in EASI score. This finding makes using EASI as a continuous variable unreliable. In addition, power issues seen with dichotomizing variables are largely a problem associated

with dichotomizing explanatory variables (Altman and Royston, 2006).

Statistical analysis/model selection

The probability of having an EASI > 10 depending on environmental variables was fitted using multivariable logistic GAMs while adjusting for potential confounders in the EASI > 10 and EASI ≤ 10 cohorts.

Details of covariates used in models

Age and sex were recorded, and weight and height from the recruitment date were converted into body mass index. Participants' postcodes were collected and converted into longitude and latitude. Longitude and latitude were used to adjust for any differences in the hyperlocal environment. The enrolment date was used to calculate the season at recruitment—the Solstices and Equinoxes being used as cut-offs. Social class was calculated using the European Socio-Economic Classification (ESeC) (Rose and Harrison, 2007). The ESeC score was collapsed into three groups: working class (ESeC 6–9), middle class (ESeC 4–6), and higher class (ESeC 1–3) as per Rose and Harrison (2007).

Transepidermal water loss, a surrogate for skin barrier function, was measured using a Tewameter TM300 probe from Courage + Khazaka Electronics (Köln, Germany). Skin hydration of the epidermis was measured with the Corneometer CM825 probe, also from Courage + Khazaka Electronics. Both these measurements came from volar forearm nonlesional skin and an average of triplicate measurements was used. All measurements were performed on skin that was exposed to a climatically controlled environment for at least 20 minutes. Differences in the values of transepidermal water loss and skin hydration may represent underlying genetic variation resulting in defective skin barrier or could represent the effect of environmental factors directly impacting the barrier.

The investigator who performed the EASI assessment was recorded to allow adjustment for expected inter-rater variability in EASI scores.

Weather/pollution stations and data quality control

Locally measured levels of wind speed (mph), relative humidity (%), and temperature (°C) and hourly concentrations of PM_{10} and $PM_{2.5}$, NO, NO_2, and tropospheric O_3 were used in the analysis. Hourly measurements of wind speed (mph), relative humidity (%), and temperature (°C) were obtained from the Centre of Environmental Data Analysis web processing service (http://wps-web1.ceda.ac.uk/ui/home) using station source identification 18929 (London City Airport, United Kingdom), our nearest active station. All data used passed the Centre of Environmental Data Analysis's quality control check. Values that did not pass were removed and imputed (see the section below). Precipitation was not recorded because the nearest active station with historic precipitation data was Reading (44.5 miles to the West of our study area). UV index data were also not collected for analysis because they were also not locally available. Hourly concentrations of PM_{10} and $PM_{2.5}$, measured by Filter Dynamics Measurement System, and NO, NO_2, and tropospheric O_3 were retrieved from Air Quality England (https://www.airqualityengland.co.uk/) using the Tower Hamlets—Blackwall (TH004) roadside monitoring dataset. This site was chosen because it was the nearest site that recorded O_3 (Supplementary Figure S2). All pollutants were measured in μg/m³. All data used underwent quality assurance/control and was ratified (verified) by the Department of Environment, Food and Rural Affairs (London, United Kingdom). Data were collected from 1 October 2014 to 30 September 2020. We collected data from 4 years before the first participant recruitment to help to see the long-term trends and the relationships between the variables. All hourly data were converted into 24-hour average values.

GAMs and top model set selection process

Because generalized linear models relax the assumption of normality required for ordinary least squares regression, GAMs take this further and can handle nonlinear relationships between the outcome and explanatory variables. GAMs are useful for modeling nonlinear data such as pollution and meteorological parameters and have been used extensively to investigate the association between health, (skin) disease, and the ambient environment (Kim et al., 2017; Ravindra et al., 2019; Wang et al., 2021).

In this paper, we use information theory rather than traditional null hypothesis significance testing. We have not used null hypothesis significance testing because a comparison of models using the likelihood-ratio test requires that the compared models are nested and all from the same parent model. Our approach uses the AIC, a metric that uses the maximum likelihood penalized by the number of model parameters, to assess model performance. The AIC itself is meaningless, but differences in AIC can be used to rank competing models, which can be non-nested, as many are in this paper. Multiple models are ranked, and a top set of possible explanatory models is selected using an AIC threshold to increase the probability of including the best-expected model (95% probability with an ΔAIC of 6) (Harrison et al., 2018). This approach assumes that multiple models (rather than only one in null hypothesis significance testing) may have the ability to explain a system equally well (Burnham et al., 2011). The presence of a variable in multiple models in the top set supports their inclusion in the system/model.

Logistic GAMs were used to analyze the nonlinear relationships between weather and pollution variables. Using R (version 3.5.3), we called the gam() function from the mgcv package (Wood, 2017; Wood, 2011). Because the outcome variable was binary, we used the binomial family argument. The smoothing parameter estimation method of restricted maximum likelihood and the AIC was used for model selection. Restricted maximum likelihood was chosen over generalized cross-validation owing to its superior performance and the ability to penalize overfitted models (Reiss and Todd Ogden, 2009; Wood, 2011). The mgcv package has been optimized for the AIC to function correctly with restricted maximum likelihood using the appropriate degrees of freedom/accounting for penalization (Wood et al., 2016).

Currently, there is no clear evidence of which exposure period and which level of a pollutant or meteorological variable has the strongest association with eczema severity. To understand this, we created nine MAvs for each meteorological and pollutant variable (average over the 7, 15, 30, 60, 90, 120, 180, 270, or 365 days preceding recruitment), which were also matched to the participants' recruitment date to assess longer-term trends.

Per exploratory variable, the one model of 10 (value on the day of recruitment and nine MAv variables) with the lowest AIC was chosen to move forward to top model set selection. Simply put, the AIC is a number that can be used to compare models. It puts a value on how well models fit the data while penalizing complexity; lower values indicate improved support for the model (Burnham and Anderson, 2004, 2002).

The base model for meteorological and pollutant MAvs selection was as follows:

$$\text{Logit}(E[y_i]) = f_1(X_{1-10}) + f_2(age) + f_3 \text{ (body mass index)}$$
$$+ f_4 \text{ (longitude, latitude)}$$
$$+ f_5 \text{ (transepidermal water loss)}$$
$$+ f_6 \text{ (skin hydration)} + \text{factor (sex)}$$
$$+ \text{factor (season)} + \text{factor(investigator)}$$
$$+ \text{factor (ESeC)}, \ y_i \sim \text{binary}$$

where $E(y_i)$ is EASI score >10 or ≤10, and X_{1-10} are individual meteorological or pollutant variables and their nine associated MAvs. Per exploratory variable, the one model of 10 (value on the day of recruitment and nine MAv variables) with the lowest AIC was chosen to move forward to top model set selection.

It is well-established that climactic and pollutant data are often colinear (Supplementary Figure S3a), which can lead to problems with model fitting (Zuur et al., 2010). To avoid these problems, we generated composite variables in cases where variable pairs have correlation coefficients >0.8 (Dormann et al., 2013; Mateo et al., 2013). The variable pairs with the highest correlation undergo principal component analysis first, and dimension 1 is used as a composite. At this stage, the correlation analysis is repeated with the new composites and remaining variables until correlations between all variables are reduced below 0.8: in this case, NO 365 and NO$_2$ 365, PM$_{10}$ 270 and wind 365, and finally temperature 180 and humidity 180 are combined into three separate composites (Supplementary Figure S3b).

The base model for the top model set selection is as follows:

$$\text{Logit}(E[y_i]) = f_1(\text{NO/NO}_2 \text{ dimension 1})$$
$$+ f_2(\text{PM}_{10}/\text{wind dimension 1}) + f_3(\text{temperature /humidity })$$
$$+ f_4(\text{PM}_{2.5} \ 90) + f_5(\text{O}_3 \ 270) + f_6(\text{skin hydration})$$
$$+ f_7(\text{transepidermal water loss}) + f_8(age) + f_9(\text{body mass index})$$
$$+ f_{10}(\text{longitude, latitude}) + \text{factor (season)} + \text{factor (sex)}$$
$$+ \text{factor (investigator)} + \text{factor (ESeC)}, \ y_i \sim \text{binary}$$

In this model, f_{1-5} represents the selected meteorological/pollutant variable with its associated MAv.

Two different approaches were used for model selection (Marra and Wood, 2011). The first approach was to use backward selection using the AIC. The top model set is usually a set based on the change in AIC from the top-performing AIC model (ΔAIC) (Burnham and Anderson, 2002). Traditionally, a ΔAIC of 2 was used for a top model set.

More recently, it has been shown that to ensure a 95% probability of having the best model included in the top model set, one should use a ΔAIC of 6 (Richards, 2008, 2005). When using the ΔAIC of 6, we used the nesting rule to help remove more complex but equally well-supported models within a ~ Δ2 AIC (Arnold, 2010; Richards, 2008, 2005; Richards et al., 2011). Marra and Wood (2011) show that backward selection has low false-positive rates but can remove influential covariates, thereby increasing the false-negative rate, and so models were also selected using a second method, the double penalty approach. This approach adds an additional penalty to the smooth function and penalizes the functions that are only in the null space of the original penalty. The process essentially selects of the model variables that are not as influential. This approach can perform better than backward selection in terms of mean squared error, particularly when the data do not have high information content (Marra and Wood, 2011).

Handling of missing data

In the weather dataset, 0.73% of the temperature and humidity data were missing, and 0.74% of the wind data were missing. In the Air Quality England pollution dataset, 25.80% of the PM$_{2.5}$, 13.51% of the PM$_{10}$, 6.44% of the O$_3$, and 3.69% of the NO and NO$_2$ data were absent. Multivariable imputation was performed for missing data. The multivariable approach using the mtsdi package was used (Junger and Ponce de Leon, 2015). The Expectation–Maximization methods used have been shown to perform well on data sets with up to 40% missing values and even perform reasonably well with missing-not-at-random data. Performance for gaps in data of up to 7 days was also good. Of the methods available in the mtsdi package, the Expectation–Maximization spline method was used because this has the greatest precision and accuracy. If missing values were present in other clinical variables, the case was removed from the analysis. A total of 25 cases were removed, leaving a total of 579 cases for analysis. Four cases lived completely out of the area. One case did not have recorded sex, one case did not have trans-epidermal water loss/skin hydration measurements, three cases did not have social class status, and 16 participants did not have height and weight recorded.

Multiple testing

We have not corrected for multiple testing for two reasons. As an exploratory study, we did not want to increase our type II error and miss true positives in an attempt to improve type I error rates. The second reason is that the AIC was used for model selection and not P-values. We feel that this approach is free from the multiple testing problem (Burnham and Anderson, 2002).

Sensitivity analysis

To ensure that our data were not reliant on underlying recruitment bias or sensitive to changes in data input, we checked for robustness by modifying the input dataset. We created three further truncated datasets (Supplementary Figure S4). We performed the same analysis for the main dataset and used the double penalty approach for top model selection—the resulting model outputs consistent with our main findings.

Reproducibility

We used a dataset from a study published by the Allergy Department at the Samsung Medical Centre in Seoul, South Korea (Kim et al., 2017). This study looked at eczema symptoms in children aged 5 years or younger from August 2013 to December 2014 (<1 year) and the effects of meteorological conditions and pollutants on the presence of eczema symptoms. Data were available for 177 participants, including the patients' date of enrolment, age, sex, season, and eczema severity as measured by another scoring criterion, SCORAD (Schmitt et al., 2014). SCORAD was only recorded on the day of enrolment; therefore, data from only this day were used.

We were able to access historic pollution data with O$_3$, NO$_2$, and PM$_{10}$ levels from Air Korea (https://www.airkorea.or.kr/). We collected data from the nearest monitoring station to the Samsung Medical Centre (Daechi-dong, Gangnam-gu, Seoul, Korea). The same data preparation and analysis described earlier for selecting single and full pollutant MAv models were used. Rather than EASI scoring, the SCORAD was used in this study. We dichotomized the

SCORAD at 30 with one group >30 and one ≤30. SCORAD > 30 is comparable with an EASI > 10 when comparing the THEA and Korean EASI and SCORAD histograms and consulting the literature (Chopra et al., 2017a, 2017b). In a paper that reviews the severity stratification of eczema scoring methods, a SCORAD of 30, such as an EASI of 10, sits at the lower end of the moderate classification (Chopra et al., 2017a). PM_{10} 60 and NO_2 365 had a correlation coefficient of −0.85. PM_{10} 60 and NO_2 365 have correlation coefficients of −0.73 and −0.52 with O_3 270, respectively. Principal component analysis was performed to create a PM_{10}/NO_2 composite. The composite variable had a correlation coefficient of −0.65 with O_3 270.

The base model for the top model set selection in the Korean cohort was as follows:

$$\text{logit}\left(E\left[y_i\right]\right) = f_1\left(PM_{10}/NO_2\right) + f_2(O_3\ 270) + f_3(\text{age}) + \text{factor (sex)} + \text{factor (season)}, \ y_i \sim \text{binary}$$

All data analyses were performed using R (version 3.5.3) and the mgcv for modeling and mtsdi for imputation.

Data availability statement
Datasets related to this article can be found at https://figshare.com/ DOI: 10.6084/m9.figshare.12547400 (Thelwall and Kousha, 2016).

ORCIDs
Bjorn R. Thomas: http://orcid.org/0000-0002-1815-3502
Xiang L. Tan: http://orcid.org/0000-0001-7685-7033
Shagayegh Javadzadeh: http://orcid.org/0000-0003-2415-607X
Elizabeth J. Robinson: http://orcid.org/0000-0002-2494-7421
Bryan S. McDonald: http://orcid.org/0000-0003-2590-0809
Malvina A. Krupiczojc: http://orcid.org/0000-0002-6909-2033
Syedia R. Rahman: http://orcid.org/0000-0002-8694-5173
Samiha Rahman: http://orcid.org/0000-0002-9265-5391
Rehana A. Ahmed: http://orcid.org/0000-0002-6843-8290
Rubina Begum: http://orcid.org/0000-0002-1086-8642
Habiba Khanam: http://orcid.org/0000-0002-1277-3559
David P. Kelsell: http://orcid.org/0000-0002-9910-7144
Jonathan Grigg: http://orcid.org/0000-0003-3109-6028
Robert J. Knell: http://orcid.org/0000-0002-3446-8715
Edel A. O'Toole: http://orcid.org/0000-0002-4084-4836

REFERENCES

Afaq F, Zaid MA, Pelle E, Khan N, Syed DN, Matsui MS, et al. Aryl hydrocarbon receptor is an ozone sensor in human skin. J Invest Dermatol 2009;129:2396−403.

Ahmed A, Islam A, Sarkar S, Mia A. Pattern of different eczema among outdoor patients attended Faridpur Medical College Hospital, Faridpur, Bangladesh. Faridpur Med Coll J 2010;5:53−5.

Ahn K. The role of air pollutants in atopic dermatitis. J Allergy Clin Immunol 2014;134:993−1000.

Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ 2006;332:1080.

Arnold TW. Uninformative parameters and model selection using Akaike's information criterion. J Wildl Manag 2010;74:1175−8.

Ban L, Langan SM, Abuabara K, Thomas KS, Abdul Sultan A, Sach T, et al. Incidence and sociodemographic characteristics of eczema diagnosis in children: a cohort study. J Allergy Clin Immunol 2018;141:1927−9.e8.

Belugina IN, Yagovdik NZ, Belugina OS, Belugin SN. Outdoor environment, ozone, radionuclide-associated aerosols and incidences of infantile eczema in Minsk, Belarus. J Eur Acad Dermatol Venereol 2018;32:1977−85.

Burnham KP, Anderson DR, editors. Model selection and multimodel inference: a practical information-theoretic approach. 2nd Ed. New York: Springer-Verlag; 2002.

Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res 2004;33:261−304.

Burnham KP, Anderson DR, Huyvaert KP. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav Ecol Sociobiol 2011;65:23−35.

Chen C, Arjomandi M, Balmes J, Tager I, Holland N. Effects of chronic and acute ozone exposure on lipid peroxidation and antioxidant capacity in healthy young adults. Environ Health Perspect 2007;115:1732−7.

Chopra R, Vakharia PP, Sacotte R, Patel N, Immaneni S, White T, et al. Severity strata for Eczema Area and Severity Index (EASI), modified EASI, Scoring Atopic Dermatitis (SCORAD), objective SCORAD, Atopic Dermatitis Severity Index and body surface area in adolescents and adults with atopic dermatitis. Br J Dermatol 2017a;177:1316−21.

Chopra R, Vakharia PP, Sacotte R, Patel N, Immaneni S, White T, et al. Relationship between EASI and SCORAD severity assessments for atopic dermatitis. J Allergy Clin Immunol 2017b;140:1708−10.e1.

Chow JC, Watson JG. Review of $PM_{2.5}$ and PM_{10} apportionment for fossil fuel combustion and other sources by the Chemical Mass Balance receptor model. Energy Fuels 2002;16:222−60.

Dedoussi IC, Eastham SD, Monier E, Barrett SRH. Premature mortality related to United States cross-state air pollution. Nature 2020;578:261−5.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 2013;36:27−46.

Elitt CM, McIlwrath SL, Lawson JJ, Malin SA, Molliver DC, Cornuet PK, et al. Artemin overexpression in skin enhances expression of TRPV1 and TRPA1 in cutaneous sensory neurons and leads to behavioral sensitivity to heat and cold. J Neurosci 2006;26:8578−87.

Fehsenfeld F, Calvert J, Fall R, Goldan P, Guenther AB, Hewitt CN, et al. Emissions of volatile organic compounds from vegetation and the implications for atmospheric chemistry. Global Biogeochem Cycles 1992;6:389−430.

Fitzky AC, Sandén H, Karl T, Fares S, Calfapietra C, Grote R, et al. The interplay between ozone and urban vegetation—BVOC emissions, ozone deposition, and tree ecophysiology. Front For Glob Chang 2019;6:50.

Futamura M, Leshem YA, Thomas KS, Nankervis H, Williams HC, Simpson EL. A systematic review of Investigator Global Assessment (IGA) in atopic dermatitis (AD) trials: many options, no standards. J Am Acad Dermatol 2016;74:288−94.

GBD 2013 Risk Factors Collaborators, Forouzanfar MH, Alexander L, Anderson HR, Bachman VF, Biryukov S, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries,

1990−2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015;386:2287−323.

Gentili V, Bortolotti D, Benedusi M, Alogna A, Fantinati A, Guiotto A, et al. HelixComplex snail mucus as a potential technology against O3 induced skin damage. PLoS One 2020;15:e0229613.

Gorai AK, Tuluri F, Tchounwou PB, Ambinakudige S. Influence of local meteorology and NO_2 conditions on ground-level ozone concentrations in the eastern part of Texas, USA. Air Qual Atmos Health 2015;8:81−96.

Hanifin JM, Thurston M, Omoto M, Cherill R, Tofte SJ, Graeber M. The eczema area and severity index (EASI): assessment of reliability in atopic dermatitis. EASI Evaluator Group. Exp Dermatol 2001;10:11−8.

Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher DN, Goodwin CED, et al. A brief introduction to mixed effects modelling and multi-model inference in ecology. PeerJ 2018;6:e4794.

Hendricks AJ, Eichenfield LF, Shi VY. The impact of airborne pollution on atopic dermatitis: a literature review. Br J Dermatol 2020;183:16−23.

Hidaka T, Ogawa E, Kobayashi EH, Suzuki T, Funayama R, Nagashima T, et al. The aryl hydrocarbon receptor AhR links atopic dermatitis and air pollution via induction of the neurotrophic factor artemin. Nat Immunol 2017;18:64−73.

Howell MD, Kuo FI, Smith PA. Targeting the Janus kinase family in autoimmune skin diseases. Front Immunol 2019;10:2342.

Huang Y, Yang Z, Gao Z. Contributions of indoor and outdoor sources to ozone in residential buildings in Nanjing. Int J Environ Res Public Health 2019;16:2587.

Jörres RA, Holz O, Zachgo W, Timm P, Koschyk S, Müller B, et al. The effect of repeated ozone exposures on inflammatory markers in bronchoalveolar lavage fluid and mucosal biopsies. Am J Respir Crit Care Med 2000;161:1855−61.

Junger WL, Ponce de Leon A. Imputation of missing data in time series for air pollutants. Atmos Environ 2015;102:96−104.

Kabir ML, Rahman F, Hassan MQ, Ahamed F, Mridha MA. Asthma, atopic eczema and allergic rhino-conjunctivitis in school children. Mymensingh Med J 2005;14:41−5.

Karimkhani C, Dellavalle RP, Coffeng LE, Flohr C, Hay RJ, Langan SM, et al. Global skin disease morbidity and mortality: an update from the global burden of disease study 2013. JAMA Dermatol 2017;153:406−12.

Kathuria P, Silverberg JI. Association of pollution and climate with atopic eczema in US children. Pediatr Allergy Immunol 2016;27:478−85.

Kavassalis SC, Murphy JG. Understanding ozone-meteorology correlations: a role for dry deposition. Geophys Res Lett 2017;44:2922−31.

Kim YM, Kim J, Han Y, Jeon BH, Cheong HK, Ahn K. Short-term effects of weather and air pollution on atopic dermatitis symptoms in children: a panel study in Korea. PLoS One 2017;12:e0175229.

Kim YP, Lee G. Trend of air quality in Seoul: policy and science. Aerosol Air Qual Res 2018;18:2141−56.

Langan SM, Bourke JF, Silcocks P, Williams HC. An exploratory prospective observational study of environmental factors exacerbating atopic eczema in children. Br J Dermatol 2006;154:979−80.

Le T, Wang Y, Liu L, Yang J, Yung YL, Li G, et al. Unexpected air pollution with marked emission reductions during the COVID-19 outbreak in China. Science 2020;369:702−6.

Lee CW, Lin ZC, Hu SC, Chiang YC, Hsu LF, Lin YC, et al. Urban particulate matter down-regulates filaggrin via COX2 expression/PGE2 production leading to skin barrier dysfunction. Sci Rep 2016;6:27995.

Lee JD, Drysdale WS, Finch DP, Wilde SE, Palmer PI. UK surface NO_2 levels dropped by 42% during the COVID-19 lockdown: impact on surface O_3. Atmos Chem Phys 2020;20:15743−59.

Leshem YA, Hajar T, Hanifin JM, Simpson EL. What the Eczema Area and Severity Index score tells us about the severity of atopic dermatitis: an interpretability study. Br J Dermatol 2015;172:1353−7.

Li K, Jacob DJ, Liao H, Shen L, Zhang Q, Bates KH. Anthropogenic drivers of 2013−2017 trends in summer surface ozone in China. Proc Natl Acad Sci USA 2019a;116:422−7.

Li X, Jin L, Kan H. Air pollution: a global problem needs local fixes. Nature 2019b;570:437−9.

Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990−2010: a systematic analysis for the Global Burden of Disease Study 2010 [published correction appears in Lancet 2013;381:1276] [published correction appears in Lancet 2013;381:628]. Lancet 2012;380:2224−60.

Ma Z, Hu X, Sayer AM, Levy R, Zhang Q, Xue Y, et al. Satellite-based spatiotemporal trends in PM2.5 concentrations: China, 2004−2013. Environ Health Perspect 2016;124:184−92.

Marra G, Wood SN. Practical variable selection for generalized additive models. Comput Stat Data Anal 2011;55:2372−87.

Mateo RG, Vanderpoorten A, Muñoz J, Laenen B, Désamoré A. Modeling species distributions from heterogeneous data for the biogeographic regionalization of the European bryophyte flora. PLoS One 2013;8:e55648.

McDonald BC, de Gouw JA, Gilman JB, Jathar SH, Akherati A, Cappa CD, et al. Volatile chemical products emerging as largest petrochemical source of urban organic emissions. Science 2018;359:760−4.

Mishra V, DiAngelo SL, Silveyra P. Sex-specific IL-6-associated signaling activation in ozone-induced lung inflammation. Biol Sex Differ 2016;7:16.

Odhiambo JA, Williams HC, Clayton TO, Robertson CF, Asher MI. ISAAC Phase Three Study Group. Global variations in prevalence of eczema symptoms in children from ISAAC phase three. J Allergy Clin Immunol 2009;124:1251−8.e23.

Park YM, Lee SY, Kim WK, Han MY, Kim J, Chae Y, et al. Risk factors of atopic dermatitis in Korean schoolchildren: 2010 international study of asthma and allergies in childhood. Asian Pac J Allergy Immunol 2016;34:65−72.

Pigors M, Common JEA, Wong XFCC, Malik S, Scott CA, Tabarra N, et al. Exome sequencing and rare variant analysis reveals multiple filaggrin mutations in Bangladeshi families with atopic eczema and additional risk genes. J Invest Dermatol 2018;138:2674−7.

Ravindra K, Rattan P, Mor S, Aggarwal AN. Generalized additive models: building evidence of air pollution, climate change and human health. Environ Int 2019;132:104987.

Reiss PT, Todd Ogden R. Smoothing parameter selection for a class of semi-parametric linear models. J R Stat Soc Series B Stat Methodol 2009;71:505−23.

Richards SA. Testing ecological theory using the information-theoretic approach: examples and cautionary results. Ecology 2005;86:2805−14.

Richards SA. Dealing with overdispersed count data in applied ecology. J Appl Ecol 2008;45:218−27.

Richards SA, Whittingham MJ, Stephens PA. Model selection and model averaging in behavioural ecology: the utility of the IT-AIC framework. Behav Ecol Sociobiol 2011;65:77−89.

Rogerson C, O'Shaughnessy RFL. Protein kinases involved in epidermal barrier formation: the AKT family and other animals. Exp Dermatol 2018;27:892−900.

Rose D, Harrison E. The European socio-economic classification: a new social class schema for comparative European research. Eur Soc 2007;9:459−90.

Schmidt A, Von Woedtke T, Bekeschus S. Periodic exposure of keratinocytes to cold physical plasma: an in vitro model for redox-related diseases of the skin. Oxid Med Cell Longev 2016;2016:9816072.

Schmitt J, Langan S, Williams HC, European Dermato-Epidemiology Network. What are the best outcome measurements for atopic eczema? A systematic review. J Allergy Clin Immunol 2007;120:1389−98.

Schmitt J, Spuls PI, Thomas KS, Simpson E, Furue M, Deckert S, et al. The Harmonising Outcome Measures for Eczema (HOME) statement to assess clinical signs of atopic eczema in trials. J Allergy Clin Immunol 2014;134:800−7.

Seo J, Park DR, Kim JY, Youn D, Lim YB, Kim Y. Effects of meteorology and emissions on urban air quality: a quantitative statistical approach to long-term records (1999−2016) in Seoul, South Korea. Atmos Chem Phys 2018;18:16121−37.

Silverberg JI, Hanifin J, Simpson EL. Climatic factors are associated with childhood eczema prevalence in the United States. J Invest Dermatol 2013;133:1752−9.

Thelwall M, Kousha K. Figshare: a universal repository for academic resource sharing? Online Inf Rev 2016;40:333−46.

Tsuji G, Takahara M, Uchi H, Takeuchi S, Mitoma C, Moroi Y, et al. An environmental contaminant, benzo(a)pyrene, induces oxidative

stress-mediated interleukin-8 production in human keratinocytes via the aryl hydrocarbon receptor signaling pathway. J Dermatol Sci 2011;62:42−9.

Valacchi G, Fortino V, Bocci V. The dual action of ozone on the skin. Br J Dermatol 2005;153:1096−100.

Valacchi G, Sticozzi C, Belmonte G, Cervellati F, Demaude J, Chen N, et al. Vitamin C compound mixtures prevent ozone-induced oxidative damage in human keratinocytes as initial assessment of pollution protection. PLoS One 2015;10:e0131097.

Wang HL, Sun J, Qian ZM, Gong YQ, Zhong JB, Yang RD, et al. Association between air pollution and atopic dermatitis in Guangzhou, China: modification by age and season. Br J Dermatol 2021;184:1068−76.

Williams H, Stewart A, von Mutius E, Cookson W, Anderson HR. International Study of Asthma and Allergies in Childhood (ISAAC) Phase One and Three Study Groups. Is eczema really on the increase worldwide? J Allergy Clin Immunol 2008;121:947−54.e15.

Williams HC, Pembroke AC, Forsdyke H, Boodoo G, Hay RJ, Burney PG. London-born black Caribbean children are at increased risk of atopic dermatitis. J Am Acad Dermatol 1995;32:212−7.

Wisthaler A, Weschler CJ. Reactions of ozone with human skin lipids: sources of carbonyls, dicarbonyls, and hydroxycarbonyls in indoor air. Proc Natl Acad Sci USA 2010;107:6568−75.

Wong EM, Walby WF, Wilson DW, Tablin F, Schelegle ES. Ultrafine particulate matter combined with ozone exacerbates lung injury in mature adult rats with cardiovascular disease. Toxicol Sci 2018;163:140−51.

Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J R Stat Soc Series B Stat Methodol 2011;73:3−36.

Wood SN. Generalized additive models: an introduction with R. 2nd ed. London, United Kingdom: Champion and Hall; 2017.

Wood SN, Pya N, Säfken B. Smoothing parameter and model selection for general smooth models. J Am Stat Assoc 2016;111:1548−63.

Zhang JJ, Wei Y, Fang Z. Ozone pollution: a major health hazard worldwide. Front Immunol 2019;10:2518.

Zuur AF, Ieno EN, Elphick CS. A protocol for data exploration to avoid common statistical problems. Methods Ecol Evol 2010;1:3−14.