International Journal of Epidemiology, 2022, 213–224
doi: 10.1093/ije/dyab187
Advance Access Publication Date: 19 October 2021
Original article

Air Pollution

Identifying impacts of air pollution on subacute asthma symptoms using digital medication sensors

Jason G Su,[1]* Meredith A Barrett,[2] Veronica Combs,[3] Kelly Henderson,[2] David Van Sickle,[4,5] Chris Hogg,[2] Grace Simrall,[6] Sarah S Moyer,[7] Paul Tarini,[8] Oktawia Wojcik,[8] James Sublett,[9] Ted Smith,[3,10] Andrew M Renda,[11] John Balmes,[1] Rahul Gondalia,[2] Leanne Kaye[2] and Michael Jerrett[12]

[1]Division of Environmental Health Sciences, School of Public Health, University of California at Berkeley, Berkeley, CA, USA, [2]Propeller Health, San Francisco, CA, USA, [3]Center for Healthy Air, Water and Soil, University of Louisville, Louisville, KY, USA, [4]Propeller Health, Madison, WI, USA, [5]Department of Population Health Sciences, School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA, [6]Louisville Metro, Office of Civic Innovation, Louisville, KY, USA, [7]Louisville Metro, Department of Public Health and Wellness, Louisville, KY, USA, [8]Robert Wood Johnson Foundation, Princeton, NJ, USA, [9]Family Allergy & Asthma, Louisville, KY, USA, [10]Envirome Institute, School of Medicine, University of Louisville, Louisville, KY, USA, [11]Humana, Louisville, KY, USA and [12]Fielding School of Public Health, University of California, Los Angeles, CA, USA

*Corresponding author. Division of Environmental Health Sciences, School of Public Health, 2121 Berkeley Way West, University of California at Berkeley, Berkeley, CA 94720-7360, USA. E-mail: jasonsu@berkeley.edu

Received 1 November 2020; editorial decision 29 July 2021

Abstract

Background: Objective tracking of asthma medication use and exposure in real-time and space has not been feasible previously. Exposure assessments have typically been tied to residential locations, which ignore exposure within patterns of daily activities.

Methods: We investigated the associations of exposure to multiple air pollutants, derived from nearest air quality monitors, with space-time asthma rescue inhaler use captured by digital sensors, in Jefferson County, Kentucky. A generalized linear mixed model, capable of accounting for repeated measures, over-dispersion and excessive zeros, was used in our analysis. A secondary analysis was done through the random forest machine learning technique.

Results: The 1039 participants enrolled were 63.4% female, 77.3% adult (>18) and 46.8% White. Digital sensors monitored the time and location of over 286 980 asthma rescue medication uses and associated air pollution exposures over 193 697 patient-days, creating a rich spatiotemporal dataset of over 10 905 240 data elements. In the generalized linear mixed model, an interquartile range (IQR) increase in pollutant exposure was associated with a mean rescue medication use increase per person per day of 0.201

[95% confidence interval (CI): 0.189-0.214], 0.153 (95% CI: 0.136-0.171), 0.131 (95% CI: 0.115-0.147) and 0.113 (95% CI: 0.097-0.129), for sulphur dioxide (SO_2), nitrogen dioxide (NO_2), fine particulate matter ($PM_{2.5}$) and ozone (O_3), respectively. Similar effect sizes were identified with the random forest model. Time-lagged exposure effects of 0–3 days were observed.

Conclusions: Daily exposure to multiple pollutants was associated with increases in daily asthma rescue medication use for same day and lagged exposures up to 3 days. Associations were consistent when evaluated with the random forest modelling approach.

Key words: Asthma, short-acting beta agonist, mobile health, digital sensor, environmental trigger

Key Messages

- Digital sensors monitored the time and location of over 286 980 asthma rescue medication uses and associated air pollution exposure from 1039 participants over 193 697 patient-days, creating a rich spatiotemporal dataset of over 10 905 240 data elements.
- Both linear mixed model and secondary analysis using a random forest model identified associations of multiple air pollutant exposures with rescue medication use.
- Pollutants NO_2, O_3, SO_2 and $PM_{2.5}$ were associated with increases in daily rescue medication use and their associations demonstrated non-linear relationships.
- The effect sizes identified in this study are 2-3-fold those identified using emergency department visits and hospitalizations as the health outcome.
- Positive associations of time-lagged exposures of up to 3 days were identified and were consistent with previous studies.

Introduction

Asthma affects 1 in 13 individuals in the USA,[1] accounts for $82 billion in direct and indirect cost annually,[2] impairs quality of life and leads to increased morbidity and mortality.[3–5] Air pollution may lead to increased asthma medication use[6] and has been shown to exacerbate asthma morbidity[7] and global mortality.[8–11] Previous studies have, however, relied on aggregated, infrequent or self-reported health outcome data. Furthermore, previous studies were limited to the residential locations of study participants in assessing daily air pollution exposure, which could mischaracterize variable exposure during daily activities outside the home.[7] Additionally, few studies have used a multipollutant approach.[12,13] Instead, many studies used single pollutant modelling approaches, despite the fact that people can be exposed to multiple pollutants simultaneously.[14]

Digital sensors may offer data to address these gaps.[15] Fitted onto inhalers, sensors can capture the date, time and location of medication use, thereby offering an objective signal of rescue inhaler use, a proxy for asthma symptoms in real-time and space. Although spatiotemporally rich, these data present some methodological challenges. The medication use outcome data tend to be over-dispersed and have frequent days without use (i.e. zero inflation), and the associations can be non-linear. To address these issues, a recent advance in biostatistics, the generalized linear mixed model using Template Model Builder (glmmTMB), can handle repeated measures from individuals and seasons, model both Poisson and negative binomial distributions and process data with excessive zeros with balanced speed and flexibility.[16–18]

In this analysis, we aimed to identify the associations of asthma rescue inhaler medication use with exposures to multiple air pollutants, derived from the nearest air quality monitoring stations to space-time rescue inhaler use, including nitrogen dioxide (NO_2), ozone (O_3), sulphur dioxide (SO_2) and particulate matter with aerodynamic diameter $\leq$2.5 microns ($PM_{2.5}$), for both same-day and lagged exposures up to 5 days. We collected data from a cohort in Louisville, Kentucky[19] (Figure 1), a city that ranks among the top 25 'most challenging places to live with asthma' in the USA,[20] as part of a public-private

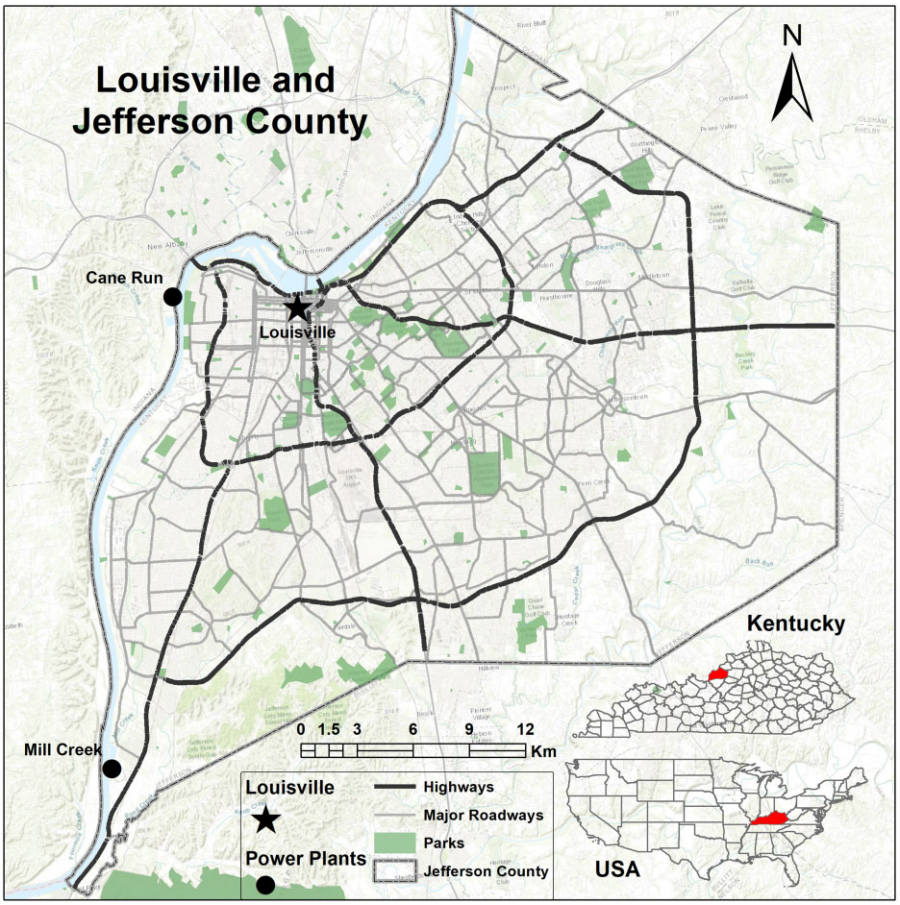

Figure 1 The City of Louisville in Jefferson County, Kentucky, USA

collaboration to address respiratory disease that was called AIR Louisville.[19]

Methods

Study population and medication use data collection

All participants signed the Propeller Health Terms of Service, which states that data collected by sensors may be used for public health analyses. The protocol was reviewed and approved by the Copernicus Independent Review Board (PRH1-17–508).

Eligible participants enrolled in AIR Louisville through a number of channels, including employer partner wellness programmes, clinics, community events and social media campaigns. The programme is described in detail elsewhere.[19] Inclusion criteria consisted of a self-reported diagnosis of asthma, a current prescription for a sensor-compatible inhaled medication for asthma and living or working in Jefferson County, Kentucky.

All participants received digital sensors to attach to their inhaled medications for asthma, including both daily controller medications and short-acting beta agonist (SABA) or rescue medications (Propeller Health, Madison, WI) (Figure 2). The sensors regularly transmit medication use data back to the server ('sync') through a smartphone or wireless hub, along with the 'heartbeat' signal. Rescue actuations occurring within 2 min were considered a single medication use 'event'. Events within a day were summed for each participant to achieve an outcome of rescue medication events per person per day. To limit the influence of outliers in our analysis, we capped the maximum number of rescue medication events per person per day at 20. Global Positioning System (GPS) locations of rescue medication use and heartbeats were assigned from paired smartphones or through a sequential back-filling process (Supplementary Data, available as Supplementary data at *IJE* online).

Environmental and adjustment variable assignments

Each rescue medication use or heartbeat event was assigned concentrations of the five critical pollutants. We included critical pollutants known to be relevant for

Figure 2 Digital medication sensors fit onto inhaled respiratory medications and record the date and time of use. Data are transmitted wirelessly via a paired smartphone

asthma, including: nitrogen dioxide (NO_2), ozone (O_3), sulphur dioxide (SO_2) and particulate matter of $\leq$2.5 microns ($PM_{2.5}$) and $\leq$10 microns (PM_{10}). Air pollutant data were acquired from the USA Environment Protection Agency (EPA)'s Air Quality System (AQS) for the entire Ohio River Valley Climate region, including the states of Illinois, Missouri, Ohio, West Virginia, Indiana, Kentucky and Tennessee. The concentrations of the criterical pollutants were measured to match the frequency of the EPA's monitoring, which includes an 8-h maximum for O_3 and 24-h mean for $PM_{2.5}$ and PM_{10}. Although concentration data are available on an hourly basis for NO_2 and SO_2, we averaged them to mean daily concentrations to keep the time scale consistent across all pollutants. The air pollutant data assignment process occurred in this order, depending on data availability: (i) assigned from the closest air quality monitoring station, bounded within Kentucky, within the same day; (ii) pollutant data assigned from the closest air quality monitoring station, bounded within the Ohio River Valley climate region, within the same day; (iii) pollutant data assigned from the closest air quality monitoring station, bounded within Kentucky, within 24 h of the event; (iv) pollutant data assigned from the closest air quality monitoring station, bounded within the Ohio River Valley Climate Region, within 24 h. If no data were matched after the fourth step, the events were removed from analysis (NO_2 and O_3: 0%; SO_2: 8%; $PM_{2.5}$ and PM_{10}: 0.2%). The climate region was used to bound assignments because climatic conditions and dispersion of air pollutants can differ significantly across different climate regions, as determined by National Oceanic and Atmospheric Administration (NOAA) longitudinal data.[21] The median matched distances from rescue or heartbeat event to the closest monitor for NO_2, O_3, SO_2, $PM_{2.5}$ and PM_{10} were, respectively, 11.91 km, 10.44 km, 12.50 km, 6.93 km and 13.07 km. These distances are similar to the 12-km downscaler resolution EPA uses for modelling O_3 concentrations.[22]

The daily mean (standard deviation) exposures to NO_2, O_3, SO_2 and $PM_{2.5}$ among the 1039 participants were, respectively, 11.4 (7.0) ppb, 27.4 (11.6) ppb, 0.8 (1.0) ppb and 9.0 (4.3) $\mu g\ m^{-3}$, and their 95th percentile values were, respectively, 24.7 ppb, 46.6 ppb, 2.5 ppb and 17.0 $\mu g\ m^{-3}$.

Each rescue or heartbeat event was assigned concentrations of the five critical pollutants. We also obtained daily pollen and mould counts (by a specialty clinic, Family Allergy & Asthma, which is a National Allergy Bureau-certified pollen counting station in Louisville) and weather data (temperature, relative humidity, atmospheric pressure acquired from the National Oceanic and Atmospheric Administration Quality Controlled Local Climatological Data Repository) and assigned them to individual exposure space in a way similar to assigning air pollution data (see Supplementary File, available as Supplementary data at *IJE* online). To adjust for neighbourhood socioeconomic status, we assigned the Centers for Disease Control and Prevention (CDC) Social Vulnerability Index (SVI) value at travel level for each participant.[23] The assignment process is fully described in the Supplementary data.

Statistical analysis

Before modelling, we identified whether significant multicollinearity existed among the environmental factors through a variance inflation factor (VIF) analysis (Supplementary data).[24] Multicollinearity inflates the variance of predictors and leads to biased coefficient estimation and a loss of

power.[25,26] AVIF score is calculated for each predictor by doing a linear regression of that predictor on all other predictors, and then obtaining the variance being explained value (R^2) from that regression. If no significant multicollinearity existed, we then modelled the associations of simultaneous air pollution exposure with daily rescue medication using glmmTMB. Otherwise, the air pollutant that was collinear with other environmental variables was removed from analysis.[18] Our model outcome represents the number of rescue medication uses per person per day, which is a count. To identify whether our glmmTMB model would be more appropriate using a Poisson or negative binomial modelling approach,[27] we tested model dispersion with the R package blmeco (v 1.3).[28] We further tested whether the linear mixed model should account for excessive zeros[29] through a zero-inflation test using the R package vcdExtra (v0.7.1).[30] We ran the R package glmmTMB (v0.2.3)[31] using the appropriate modelling specification (i.e. Poisson or negative binomial, and with or without the logit model) identified from the above analysis and included a random intercept for both individual participant and season. We assumed outcome of zeros (days without rescue medication use) could be associated with low air pollution exposure and incorporated NO_2, O_3, SO_2 and $PM_{2.5}$ as parameters in the logit model when excessive zeros existed.

To demonstrate the marginal effects of individual pollutants, we generated penalized partial dependence plots with 95% confidence intervals[32] through a smoothing method called penalized regression splines generalized additive model (GAM) (mgcv v1.8-31).[33,34] Partial dependence plots are low-dimensional graphical renderings of the relationship between the predicted outcome and a predictor of interest, which accounts for the mean effect of the other predictors in the model.[35] Due to the potential complex relationship between air pollutant exposure and rescue medication use, their partial dependency and identified marginal effects might be non-linear. To avoid overly large confidence intervals due to insufficient data support, a partial plot for a pollutant of interest was truncated at the 95th percentile of the observed exposure level. To be able to compare the marginal effects of individual air pollutant exposures on rescue medication use, we also identified linear trends from the detected partial dependence plots.

Due to the complex nature of the glmmTMB modelling framework, a secondary analysis was also conducted through machine learning techniques that also offer solutions for analysing high-dimensional patient data.[36] Random forest is an ensemble learning method that builds a multitude of decision trees at training and predicts an outcome using out-of-bag data.[37] It modifies the bagging technique by ensuring that the individual trees are de-correlated by using a bootstrap sample for each tree and randomly selecting a subset of predictors for testing at each split point in each tree. As one of the most frequently used machine learning techniques in health-related research,[38] random forest has a potential advantage over traditional linear models for several reasons: (i) lack of underlying data structure assumptions; (ii) resistance to model overfitting; (iii) ability to handle multicollinearity among predictors; and (iv) increased accuracy in prediction.[37]

Using the random forest R package,[39] we modelled daily rescue medication use using the same parameters as those specified in the linear mixed glmmTMB models. Both glmmTMB and random forest models included variables to account for confounding from individual characteristics, including: age, gender, self-reported race-ethnicity, and initial disease status defined by the Asthma Control Test (ACT).[40] The ACT is a five-question, 4-week recall of asthma symptoms and use of rescue medications, and self-assessment of asthma control. The scores range from 5 (poorly controlled) to 25 (well-controlled), with higher scores reflecting greater asthma control. An ACT score indicates level of control, with <20 indicating not well-controlled and <15 indicating very poorly controlled. For each individual, we also included the number of days per week with rescue medication use as a proxy for ongoing disease status,[41] aiming to adjust for each patient's individual variability in disease severity to better assess the impact of air pollutant exposure alone. As mentioned previously, we also controlled for impacts from daily weather conditions, ambient pollen and mould counts and census-tract-level socioeconomic status.

Time-lagged exposure analysis

To explore whether time-lagged air pollution exposure was associated with rescue medication use,[42] we evaluated associations of the preceding 5 days of averaged air pollutant concentrations by taking the average of the index day (i.e. lag 0) and preceding 1–5 days, in the modelling process.

Results

Study population and rescue medication use data collection

We recruited 1039 participants with asthma across Jefferson County: 63.4% of the participants were female, 77.3% were adult (>18), mean age was 34.9 [standard deviation (SD) 15.6]and 46.8% were White (Table 1). Participants had a median baseline ACT score of 14.1 (SD 3.4), and 93.7% of participants were considered not controlled at baseline (i.e. <20 on initial ACT). Neighbourhood-level social vulnerability (SVI, mean 0.53 and SD 0.59) was slightly higher than national figures (mean 0.50 and SD 0.29) based on CDC statistics.[23]

Table 1. Demographic statistics of the participants ($n = 1039$)

Demographics	*n*	Value
Gender		
Male	380	36.6%
Female	659	63.4%
Age		
3–17	236	22.7%
18–25	102	9.8%
26–45	380	36.6%
46–65	276	26.6%
66 and over	45	4.3%
Race-ethnicity		
White/Caucasian	486	46.8%
African American	228	21.9%
Hispanic	5	0.5%
Asian	9	0.9%
Native American	9	0.9%
Pacific Islander	2	0.2%
Other	2	0.2%
Not reported	298	28.7%
Baseline ACT Score	Mean (SD)	Score, value
	14.1 (3.4)	5, 0.87% [6–9], 9.72% [10–14], 54.48% [15–19], 28.68% [20–25], 6.26%
% Pollen exposure	Mean (SD)	1st–3rd Q
from grass (counts 3.12 m^3 air)	2.58 (11.98)	0.00–2.16
from weed (counts 3.12 m^3 air)	6.94 (22.06)	0.00–2.16
from trees (counts 3.12 m^3 air)	53.62 (211.99)	0.00–6.48
from mould (spores) (3.12 m^3 air)	4194 (3386.07)	1463–6502
Neighbourhood social vulnerability	Mean (SD)	1st–3rd Q
	0.53 (0.59)	0.29–0.80

For the baseline Asthma Control Test (ACT) scores, we provided the mean and standard deviation values. We further categorized each ACT score into five classes and those with an ACT score below 20 are considered uncontrolled (93.74%). For pollen exposure and neighbourhood social vulnerability values, their mean (standard deviation), 1st and 3rd quarter (Q) values are presented. Pollen exposures were seasonal, with the greatest grass and tree pollen impacts in spring and the greatest weed pollen impact in later summer and earlier autumn, except for mould spores impacts which were year round.

Data were collected from 12 June 2012 to 29 November 2017, acquiring 286 980 unique rescue medication uses. Mean number of rescue events per person per day was 1.24 (SD 2.6) after capping the maximum rescue events at 20 per day, which occurred on 315 patient days or 0.4% of total rescue events.[43] The daily environmental exposure values and their standard deviations, including air pollution and weather conditions, are detailed in the Supplementary data.

Seasonal variations in air pollution exposure and rescue medication use

We aggregated daily air pollution exposure levels and daily rescue medication use to form seasonal air pollution exposure and rescue medication use across all the study participants. We then plotted seasonal air pollution exposures along with seasonal rescue medication use for the four critical pollutants (see Figure S1, available as Supplementary data at *IJE* online). The air pollution exposures were generally higher in autumn but lower in spring/summer for NO_2, $PM_{2.5}$ and SO_2, except for O_3 which showed an opposite trend. For rescue medication use, it was generally higher in autumn/winter, but lower in summer.

glmmTMB modelling results

The VIF analysis (Table S1, available as Supplementary data at *IJE* online) indicated a score of 2.4 for PM_{10}, which approaches the multicollinearity limit of 2.5.[44] After

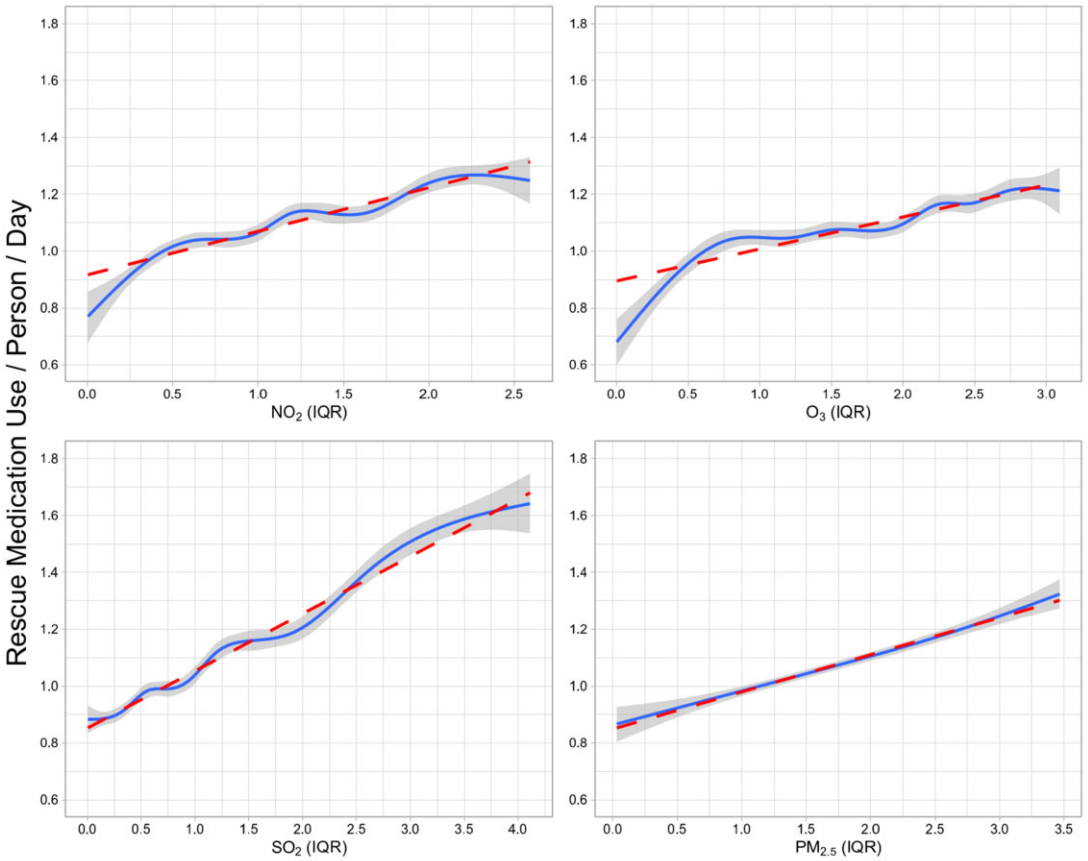

Figure 3 The penalized partial dependence plots of NO_2 (IQR = 9.44 ppb), O_3 (IQR = 15.65 ppb), SO_2 (IQR = 1.35 ppb) and $PM_{2.5}$ (IQR = 5.80 μg m^{-3}) with daily rescue medication use (y-axis), modelled through a glmmTMB (generalized linear mixed model using Template Model Builder) model and adjusted for potential impact from individual characteristics (including age, gender and self-reported race-ethnicity) and disease status [including initial asthma control as defined by the patient-reported Asthma Control Test (ACT) scores and number of days of rescue medication use per week], daily weather conditions, ambient pollen and mould counts and census-tract-level socioeconomic status in Social Vulnerability Index (SVI). The blue solid lines indicate the expected partial impact of each air pollutant on daily rescue medication use, holding all other variables in the model at their means, with a 95% confidence interval in grey. The red dotted lines represent expected rescue medication use per person per day as a linear function of respective penalized partial dependence plots

removing PM_{10}, all the other air pollutants had a VIF value much smaller than 2.5, indicating lack of significant multicollinearity. In the bivariate analysis between rescue medication use and PM_{10} associations, PM_{10} was not significantly associated with rescue medication use; therefore we did not analyse the effect of PM_{10} on rescue medication use. The variance inflation factor (VIF) analysis showed that NO_2, O_3, SO_2 and $PM_{2.5}$ did not have significant multicollinearity and could be included in the same model (Supplementary data).[24] The dispersion test showed the glmmTMB model had a dispersion value of 1.3 (Supplementary data), indicating lack of significant dispersion in the model,[45] and therefore a Poisson model was applied in the conditional model. The zero-inflation test indicated excessive zeros existed in the number of daily rescue medication uses ($p < 0.001$) (Supplementary data); therefore we included an inflation logit model with NO_2, O_3, SO_2 and $PM_{2.5}$ as parameters in both the Poisson conditional model and the inflation logit model.

The model identified significant and positive associations among all four pollutants with daily rescue medication use. The penalized partial dependence plots (blue lines) indicated non-linear, consistently upward trends for all pollutants (Figure 3). When the linear trend was evaluated (red lines), it showed that per IQR increase of NO_2, O_3, SO_2 and $PM_{2.5}$ exposure, respective daily rescue medication use was estimated to increase by 0.153 (95% CI: 0.136-0.171), 0.113 (95% CI: 0.097-0.129), 0.201 (95% CI: 0.189-0.214) and 0.131 (95% CI: 0.115-0.147). For interpretability, we translated these increases into relative percent increase over the mean (1.24 rescue uses/person/day), which included: 12.34% (95% CI: 10.97-13.79%),

Table 2 Linear trend of per IQR increase in pollutant exposure associated with an absolute increase in daily rescue medication use (top) and percent increase relative to the mean (bottom) per person from glmmTMB and secondary analysis random forest modelling results

Rescue increase	Pollutant	glmmTMB		Random forest	
		Estimate	95% CI	Estimate	95% CI
Absolute values	NO_2	0.153	(0.136-0.171)	0.134	(0.109-0.160)
	O_3	0.113	(0.097-0.129)	0.114	(0.091-0.137)
	SO_2	0.201	(0.189-0.214)	0.179	(0.161-0.198)
	$PM_{2.5}$	0.131	(0.115-0.147)	0.124	(0.102-0.148)
Percent increase (%)	NO_2	12.34%	(10.97-13.79%)	10.81%	(8.79-12.90%)
	O_3	9.11%	(7.82-10.40%)	9.19%	(7.34-11.05%)
	SO_2	16.21%	(15.24-17.26%)	14.44%	(12.98-15.97%)
	$PM_{2.5}$	10.56%	(9.27-11.85%)	10.00%	(8.23-11.94%)

glmmTMB, the generalized linear mixed model using Template Model Builder.

9.11% (95% CI: 7.82-10.40%), 16.21% (95% CI: 15.24-17.26%) and 10.56% (9.27-11.85%) for NO_2, O_3, SO_2 and $PM_{2.5}$, respectively (Table 2).

Results from secondary analysis

Based on the penalized partial dependence plots from random forest modelling (Figure S1), the same-day exposure impact from a per IQR increase using the linear trend for a pollutant was similar to the glmmTMB modelling results. The linear trend indicated that per IQR increase of NO_2, O_3, SO_2 and $PM_{2.5}$ exposure, daily rescue medication use was estimated to increase by 0.134 (95% CI: 0.109-0.160), 0.114 (95% CI: 0.091-0.137), 0.179 (95% CI: 0.161-0.198) and 0.124 (95% CI: 0.102-0.148), respectively (Table 2).

Time-lagged exposure analysis

The time-lagged exposure analysis showed relatively varied effects of each pollutant when averaged up to 5 days of preceding exposures (Figure 4). For NO_2, we saw a 1-day lagged exposure effect: a steeper positive slope was observed from day 0 to day 1, with a gradual increase up to day 5. For SO_2, an increase in daily rescue medication use was seen over days 0–3, after which the rate showed a gradual decrease (i.e. a 3-day lagged exposure effect). For O_3, the effect decreased after the index day of exposure, which had the greatest impact on increase in rescue medication use (i.e. no time-lagged effect). For $PM_{2.5}$, a slight increase was seen from the index day to the 1-day lagged exposure, and then steady decrease was observed (i.e. a 1-day lagged exposure effect).

Discussion

This analysis demonstrated positive associations of exposures to multiple air pollutants with a spatially and temporally-resolved health outcome collected by digital sensors. The spatiotemporal exposures and locations of health outcome measure (here rescue medication use) were identified through medical sensors, rather than the traditional analysis that largely used home address as location of exposure and location of health outcome measure. The glmmTMB modelling results and the secondary analysis in random forest modelling demonstrated consistent, positive associations of NO_2, O_3, SO_2 and $PM_{2.5}$ with daily rescue medication use. Time-lagged exposure effects were also observed for NO_2 (1-day), SO_2 (3-day) and $PM_{2.5}$ (1-day), except for O_3.

The rich data collected by digital sensors, however, demand new methodological approaches. For example, traditional linear mixed models may not be sufficient in dealing with the frequent zeros in the health outcome data, as was seen here in rescue medication use per person per day measure. A linear mixed model capable of processing excessive zeros, such as the glmmTMB used in this analysis, addressed this issue by creating two models: one dealt with normal count data with a Poisson function and the second dealt with the excessive zeros through a logit function. The random forest model, like other complex nonparametric models (e.g. neural networks, support vector machines and super learners), is becoming more common in predictive analytics, especially when dealing with large observational datasets that do not adhere to the strict assumptions imposed by traditional statistical techniques (e.g. multiple linear regression assumes linearity, homoscedasticity and normality).[32]

Just as the distinctive dataset required a unique approach, analysing and interpreting the results requires a

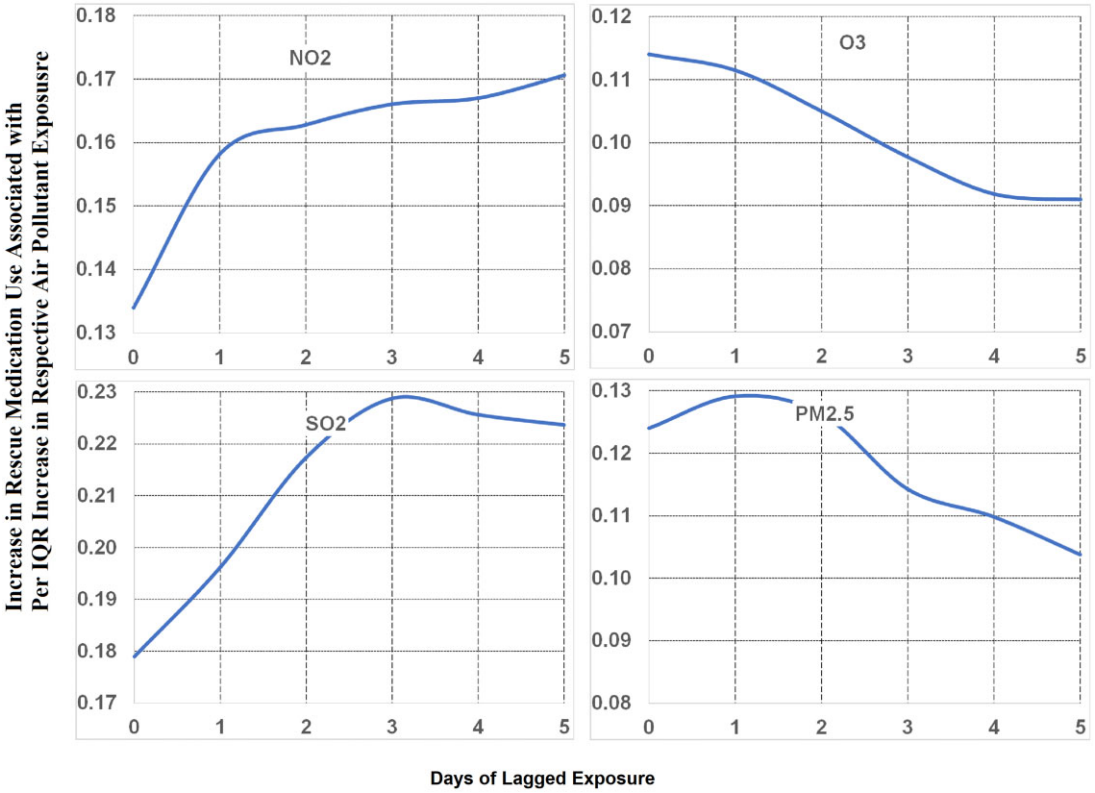

Figure 4 The impacts of duration of exposure to NO_2 (IQR = 9.44 ppb), O_3 (IQR = 15.65 ppb), SO_2 (IQR = 1.35 ppb) and $PM_{2.5}$ (IQR = 5.80 μg m^{-3}) on increases in daily rescue medication use through the respective marginal effects modelled through the random forest technique with comprehensive confounding control. The x-axes represent the number of days of averaged exposure to each pollutant as counted preceding the index day (0) and the y-axes represent the increase in the number of daily rescue medication uses associated with a per IQR increase in respective pollutant exposure; 0-day refers to the index day, 1-day refers to the average exposure from the index day plus 1 day preceding it, and so on. The 5-day refers to the average exposure from days 0–5

new approach as well. For example, glmmTMB models counts separately using a Poisson model and zeros using a logit model, which makes understanding the overall marginal effect of a predictor less straightforward. Random forest uses a multitude of decision trees for training and prediction, and typical packages (e.g. R randomForest) lack the ability to identify the overall marginal effect of a predictor. To address these limitations, we applied partial dependence plots which can display the expected marginal effects from complex or black box models.[32]

Overall, the percent increase in daily rescue medication ranged 9–14% per IQR increase in air pollution exposure, depending on the pollutant. These effect directions are consistent with previous literature; however, the magnitudes of these effects are approximately 2–3-fold those found for emergency department (ED) visits and hospitalizations in a recent meta-analysis.[46] The greater effects identified in this study suggest that rescue medication use may be an outcome more sensitive to air pollution than the more rare, acute ED visit or hospitalization event, which could help enhance our understanding of the lower=acuity impacts of air pollution.

We observed variable impacts of lagged exposures across the pollutants over their preceding days. This finding is consistent with controlled chamber studies, where time-lagged effects were demonstrated, over 24-48 h, for NO_2 and SO_2.[47,48] Other studies have demonstrated up to 5 days of lagged exposure impacts, including 1–2 days for NO_2,[42,49–51] 0–3 days for SO_2[52–54] and 3–5 days for O_3.[55,56] The difference for O_3 may be influenced in our study by a lower proportion of children, who are more sensitive to ozone.[57] Children are at greatest risk from exposure to ozone because their lungs are still developing and they are more likely to be active outdoors when ozone levels are high.

There were several limitations in this study. Participants self-selected to enroll in the programme from a variety of recruitment sites, which may introduce some bias. For air pollution data, we were limited by the distribution of the existing regulatory air quality monitoring network, which does not densely monitor all neighbourhoods in Jefferson County. Additionally, we focused on ambient air pollution and did not take indoor air quality into consideration, due to lack of data; however, we

believe that indoor air quality is extremely important and should be explored in future studies.

In summary, simultaneous exposures to multiple pollutants were associated with increases in daily asthma rescue medication use, for both current day exposures and time-lagged exposures of up to 3 days, although it varied by pollutant. The spatiotemporally rich data collected by sensors demanded novel methodological approaches, which performed comparably and provided consistent results.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

This work was supported by the Robert Wood Johnson Foundation (Grant # 71592), Foundation for a Healthy Kentucky, Norton Healthcare Foundation, Owsley Brown Charitable Foundation and the American Lung Association. The views expressed here do not necessarily reflect the views of these organizations.

Data availability

Daily air pollution data were acquired from the USA Environmental Protection Agency (EPA: https://www.epa.gov/outdoor-air-quality-data) and are publicly available. Daily weather data were acquired from the National Oceanic and Atmospheric Administration (NOAA) Quality Controlled Local Climatological Data Repository (QCLCD) and are publicly available. The daily pollen data were collected and provided to the research team by the Family Allergy & Asthma clinic (FAA), which is a National Allergy Bureau (NAB)-certified pollen counting station in Louisville. Contact with FAA or NAB is required for access to those pollen data. The Propeller Health data are owned by ResMed, and any access to the patients' rescue medication use and other related health outcome data is required to contact ResMed and seek corresponding IRB approval.

Acknowledgements

We would like to acknowledge the network of local partners that made this programme possible, including the Institute for Healthy Air, Water and Soil, Louisville Metro, the Community Foundation of Louisville and all the AIR Louisville participants. Partners within the Louisville Metro Government include: Mayor Greg Fischer, the Office of Civic Innovation, Louisville Metro Department of Public Health and Wellness, the Air Pollution Control District, the Office of Sustainability, the Office of Advanced Planning, the Louisville Jefferson County Information Consortium and Louisville Forward. Employer and health plan partners include: Brown Forman Corporation, Centerstone, Humana Inc., Kindred Healthcare, Passport Health Plan, Papa John's Pizza and WHAS 11. Clinical partners included Family Allergy and Asthma, JenCare Senior Medical Center and the University of Louisville Division of General Pediatrics. Advisers included the Robert Wood Johnson Foundation, the Nature Conservancy, the West Jefferson County Community Task Force and the Park DuValle Community Health Center. We also thank Melissa Williams and Jennifer Morgan for their dedicated clinical outreach among AIR Louisville participants.

Author Contributions

J.G.S., M.A.B., V.C., K.H., C.H., D.V.S., G.S., S.M, P.T., O.W., J.L.S., L.K., A.R and T.S. conceptualized the work and contributed to its design. J.G.S., M.A.B., V.C., K.H., C.H., D.V.S., G.S., S.M, J.L.S., A.R. and T.S were involved in the collection of the data. J.G.S. and M.A.B. planned the methodological approach, J.G.S. modelled and assigned spatiotemporal exposures to all events. J.G.S., M.A.B., M.J. and J.B. modelled, analysed and interpreted the air pollutant impacts on rescue use. R.G. provided detailed statistical consultation. J.G.S. and M.A.B. wrote the manuscript with input from all authors, who then reviewed the manuscript critically.

Conflict of interest

At the start of manuscript preparation, M A.B, K.H., C.H., D.V.S., R.G. and L.K. were Propeller Health employees with salary support. At time of submission, M A.B, R.G. and L.K. are ResMed employees and receive salary and stock. ResMed acquired Propeller in 2019. J.G.S., MA.B., K.H., C.H. and D.VS. have a patent pending relevant to Propeller products. J.G.S. received support from the Robert Wood Johnson Foundation for this work, and some additional funding from Propeller Health through a UC Berkeley and Propeller Health collaborative agreement. The other authors (V.C., G.S., S.S. M.P.T., O.W., J.S., T.S., A.MR., J.B. and M.J.) declare they have no actual or potential competing financial interests.

References

1. Centers for Disease Control and Prevention. *500 Cities: Local Data for Better Health*. 2016. https://www.cdc.gov/500cities/ (12 January 2019, date last accessed).
2. Nurmagambetov T, Kuwahara R, Garbe P. The economic burden of asthma in the United States, 2008-2013. *Ann Am Thorac Soc* 2018;**15**:348–56.
3. Akinbami LJ, Moorman JE, Liu X. Asthma prevalence, health care use, and mortality: United States, 2005-2009. *Natl Health Stat Report* 2011;**32**:1–14.
4. Peters SP, Ferguson G, Deniz Y, Reisner C. Uncontrolled asthma: a review of the prevalence, disease burden and options for treatment. *Respir Med* 2006;**100**:1139–51.
5. Zeiger RS, Schatz M, Dalal AA *et al.* Utilization and costs of severe uncontrolled asthma in a managed-care setting. *J Allergy Clin Immunol Pract* 2016;**4**:120–29, e3.
6. Menichini F, Mudu P. Drug consumption and air pollution: an overview. *Pharmacoepidemiol Drug Saf* 2010;**19**:1300–15.
7. Guarnieri M, Balmes JR. Outdoor air pollution and asthma. *Lancet* 2014;**383**:1581–92.
8. Landrigan PJ, Fuller R, Acosta NJR *et al.* The Lancet Commission on pollution and health. *Lancet* 2018;**391**: 462–512.

9. Di Q, Wang Y, Zanobetti A *et al.* Air pollution and mortality in the Medicare population. *N Engl J Med* 2017;**376**:2513–22.
10. Ebmeier S, Thayabaran D, Braithwaite I, Bénamara C, Weatherall M, Beasley R. Trends in international asthma mortality: analysis of data from the WHO Mortality Database from 46 countries (1993-2012). *Lancet* 2017;**390**:935–45.
11. Lelieveld J, Evans JS, Fnais M, Giannadaki D, Pozzer A. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* 2015;**525**:367–71.
12. Pearce JL, Waller LA, Mulholland JA *et al.* Exploring associations between multipollutant day types and asthma morbidity: epidemiologic applications of self-organizing map ambient air quality classifications. *Environ Health* 2015;**14**:55.
13. Winquist A, Kirrane E, Klein M *et al.* Joint effects of ambient air pollutants on pediatric asthma emergency department visits in Atlanta, 1998–2004. *Epidemiology* 2014;**25**:666.
14. Levy I, Mihele C, Lu G, Narayan J, Brook JR. Evaluating multipollutant exposure and urban air quality: pollutant interrelationships, neighborhood variability, and nitrogen dioxide as a proxy pollutant. *Environ Health Perspect* 2014;**122**:65–72.
15. Su JG, Barrett MA, Henderson K *et al.* Feasibility of deploying inhaler sensors to identify the impacts of environmental triggers and built environment factors on asthma short-acting bronchodilator use. *Environ Health Perspect* 2017;**125**:254–61.
16. Brooks ME, Kristensen K, van Benthem KJ *et al.* glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R Journal* 2017;**9**:378–400.
17. Brooks ME, Kristensen K, van Benthem KJ *et al.* Modeling zero-inflated count data with glmmTMB. *BioRxiv* 2017;132753.
18. Magnusson A, Skaug H, Nielsen A *et al.* glmmTMB: generalized linear mixed models using template model builder. *R Package Version 01* 2017;**3**: 1–32.
19. Barrett M, Combs V, Su JG, Henderson K, Tuffli M; AIR Louisville Collaborative. AIR Louisville: addressing asthma with technology, crowdsourcing, cross-sector collaboration, and policy. *Health Aff (Millwood)* 2018;**37**:525–34.
20. Asthma and Allergy Foundation of America. *AAFA. Asthma Capitals: The Most Challenging Places to Live with Asthma*. Landover, MD: AAFA, 2013.
21. Karl TR, Koscielny AJ. Drought in the United-States - 1895-1981. *J Climatol* 1982;**2**:313–29.
22. Berrocal VJ, Gelfand AE, Holland DM. A spatio-temporal downscaler for output from numerical models. *J Agric Biol Environ Stat* 2010;**15**:176–97.
23. Centers for Disease Control and Prevention. The Social Vulnerability Index (SVI*)*. 2013. https://svi.cdc.gov/ (23 August 2017, date last accessed).
24. Thompson CG, Kim RS, Aloe AM, Becker BJ. Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. *Basic Appl Soc Psychol* 2017;**39**: 81–90.
25. Yoo W, Mayberry R, Bae S, Singh K, He QP, Lillard JW Jr. A study of effects of multicollinearity in the multivariable analysis. *Int J Appl Sci Technol* 2014;**4**:9.
26. Dormann CF, Elith J, Bacher S *et al.* Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 2013;**36**:27–46.
27. Gardner W, Mulvey EP, Shaw EC. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychol Bull* 1995;**118**:392–404.
28. Korner-Nievergelt F, Roth T, Felten SV, Guelat J, Almasi B, Korner-Nievergelt P. *blmeco: Data Files and Functions Accompanying the Book "Bayesian Data Analysis in Ecology Using R, BUGS and Stan"*, 2015.
29. Van den Broek J. A score test for zero inflation in a Poisson distribution. *Biometrics* 1995;**51**:738–43.
30. Friendly M, Turner H, Zeileis A. *vcdExtra: vcd Extensions and Additions*. R Package Version 06–0.. Vienna: R Foundation for Statistical Computing, 2014.
31. Magnusson A, Skaug H, Nielsen A *et al.* Package 'glmmTMB'. *R Package Version 02 0*. Vienna: R Foundation for Statistical Computing, 2017.
32. Greenwell BM. pdp: an R Package for constructing partial dependence plots. *R Journal* 2017;**9**:421–36.
33. Wood SN. mgcv: GAMs and generalized ridge regression for R. *R News* 2001;**1**:20–25.
34. Wood SN. Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J Am Stat Assoc* 2004; **99**:673–86.
35. Breiman L, Cutler A, Liaw A, Wiener M. Random Forest: Breiman and Cutler's random forests for classification and regression, 2018. https://CRAN.R-project.org/package=randomForest (14 November 2019, date last accessed).
36. Zhou CY, Jia Y, Motani M, Chew JW. Learning deep representations from heterogeneous patient data for predictive diagnosis. In: *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, 20-23 August 2017*. Boston, MA, 2017, 115–23.
37. Breiman L. Random forests. *Mach Learn* 2001;**45**:5–32.
38. Beeler C, Dbeibo L, Kelley K *et al.* Assessing patient risk of central line-associated bacteremia via machine learning. *Am J Infect Control* 2018;**46**:986–91.
39. Liaw A, Wiener M. Classification and regression by random forest. *R News* 2002;**2**:18–22.
40. Nathan RA, Sorkness CA, Kosinski M *et al.* Development of the asthma control test: a survey for assessing asthma control. *J Allergy Clin Immunol* 2004;**113**:59–65.
41. National AE, Prevention P. Expert Panel Report 3 (EPR-3): guidelines for the diagnosis and management of asthma-summary report 2007. *J Allergy Clin Immunol* 2007;**120**:S94.
42. Lin M, Chen Y, Burnett RT, Villeneuve PJ, Krewski D. Effect of short-term exposure to gaseous pollution on asthma hospitalisation in children: a bi-directional case-crossover analysis. *J Epidemiol Community Health* 2003;**57**:50–55.
43. National Asthma Education and Prevention Program. *Guidelines for the Diagnosis and Management of Asthma (EPR-3)*. 2007. https://www.nhlbi.nih.gov/health-topics/guidelines-for-diagnosis-management-of-asthma (16 January 2019, date last accessed).
44. Ekerstad N, Swahn E, Janzon M *et al.* Frailty is independently associated with short-term outcomes for elderly patients with non–ST-segment elevation myocardial infarction. *Circulation* 2011;**124**:2397–404.
45. Barry S, Brooks SP, Catchpole EA, Morgan BJ. The analysis of ring-recovery data using random effects. *Biometrics* 2003;**59**:54–65.

46. Orellano P, Quaranta N, Reynoso J, Balbi B, Vasquez J. Effect of outdoor air pollution on asthma exacerbations in children and adults: systematic review and multilevel meta-analysis. *PLoS One* 2017;**12**:e0174050.
47. Jenkins HS, Devalia JL, Mister RL, Bevan AM, Rusznak C, Davies RJ. The effect of exposure to ozone and nitrogen dioxide on the airway response of atopic asthmatics to inhaled allergen: dose-and time-dependent effects. *Am J Respir Crit Care Med* 1999;**160**:33–39.
48. Rusznak C, Devalia J, Davies R. Airway response of asthmatic subjects to inhaled allergen after exposure to pollutants. *Thorax* 1996;**51**:1105–08.
49. Raun LH, Ensor KB, Persse D. Using community level strategies to reduce asthma attacks triggered by outdoor air pollution: a case crossover analysis. *Environ Health* 2014;**13**:58.
50. Smith BJ, Nitschke M, Pilotto LS, Ruffin RE, Pisaniello DL, Willson KJ. Health effects of daily indoor nitrogen dioxide exposure in people with asthma. *Eur Respir J* 2000;**16**:879–85.
51. Schildcrout JS, Sheppard L, Lumley T, Slaughter JC, Koenig JQ, Shapiro GG. Ambient air pollution and asthma exacerbations in children: an eight-city analysis. *Am J Epidemiol* 2006;**164**:505–17.
52. P Boogaard P, Banton M, Deferme L *et al.* Review of recent health effect studies with sulphur dioxide. https://www.concawe.eu/wp-content/uploads/2017/01/rpt-16-4.pdf 2016 (25 September 2018, date last accessed).
53. Szyszkowicz M, Porada E, Tremblay N, Grafstein E. Sulfur dioxide and emergency department visits for stroke and seizure. *Stroke Res Treat* 2012;**2012**:824724.
54. Smargiassi A, Kosatsky T, Hicks J *et al.* Risk of asthmatic episodes in children exposed to sulfur dioxide stack emissions from a refinery point source in Montreal, Canada. *Environ Health Perspect* 2009;**117**:653–59.
55. Sheffield PE, Zhou J, Shmool JLC, Clougherty JE. Ambient ozone exposure and children's acute asthma in New York City: a case-crossover analysis. *Environ Health* 2015;**14**:25.
56. Mortimer KM, Tager IB, Dockery DW, Neas LM, Redline S. The effect of ozone on inner-city children with asthma: identification of susceptible subgroups. *Am J Respir Crit Care Med* 2000;**162**:1838–45.
57. Environmental Protection Agency U. Health Effects of Ozone Pollution. 2019. https://www.epa.gov/ground-level-ozone-pollution/health-effects-ozone-pollution (20 June 2021, date last accessed)