Machine learning–driven identification of early-life air toxic combinations associated with childhood asthma outcomes
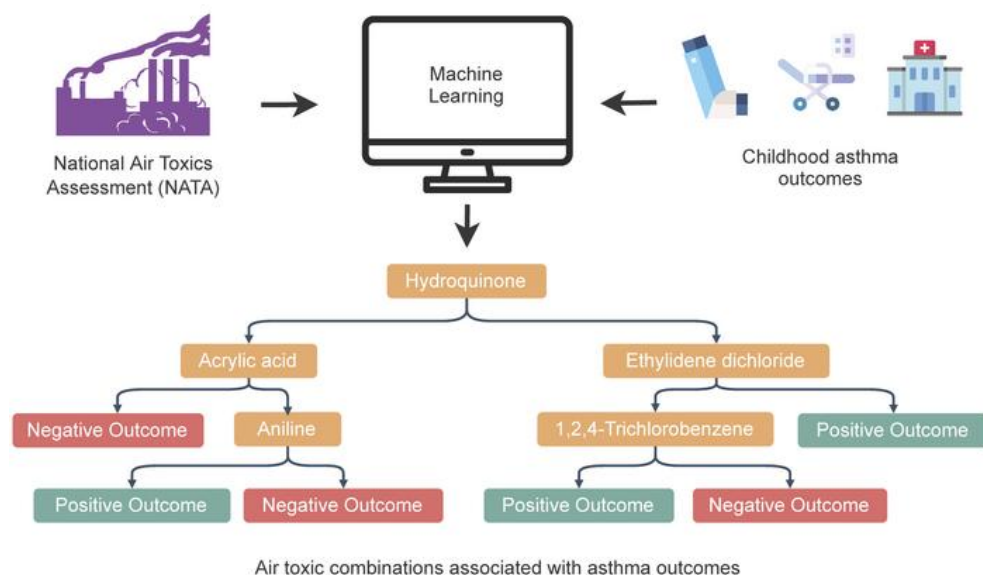
Yan-Chak Li, … , Gaurav Pandey, Supinda Bunyavanich

Research Article   **Pulmonology**

Graphical abstract

Air toxic combinations associated with asthma outcomes

Find the latest version:

https://jci.me/152088/pdf

Machine learning–driven identification of early-life air toxic combinations associated with childhood asthma outcomes

Yan-Chak Li,[1] Hsiao-Hsien Leon Hsu,[2,3] Yoojin Chun,[1] Po-Hsiang Chiu,[1] Zoe Arditi,[1,4] Luz Claudio,[2,3] Gaurav Pandey,[1,3] and Supinda Bunyavanich[1,4]

[1]Department of Genetics and Genomic Sciences, [2]Department of Environmental Medicine and Public Health, [3]Institute for Exposomic Research, and [4]Division of Allergy and Immunology, Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

Air pollution is a well-known contributor to asthma. Air toxics are hazardous air pollutants that cause or may cause serious health effects. Although individual air toxics have been associated with asthma, only a limited number of studies have specifically examined combinations of air toxics associated with the disease. We geocoded air toxic levels from the US National Air Toxics Assessment (NATA) to residential locations for participants of our AiRway in Asthma (ARIA) study. We then applied Data-driven ExposurE Profile extraction (DEEP), a machine learning–based method, to discover combinations of early-life air toxics associated with current use of daily asthma controller medication, lifetime emergency department visit for asthma, and lifetime overnight hospitalization for asthma. We discovered 20 multi–air toxic combinations and 18 single air toxics associated with at least 1 outcome. The multi–air toxic combinations included those containing acrylic acid, ethylidene dichloride, and hydroquinone, and they were significantly associated with asthma outcomes. Several air toxic members of the combinations would not have been identified by single air toxic analyses, supporting the use of machine learning–based methods designed to detect combinatorial effects. Our findings provide knowledge about air toxic combinations associated with childhood asthma.

Introduction

Air toxics are hazardous air pollutants that cause or may cause serious health effects (1). They are well-established detriments to human respiratory health, especially for children (2–8). In particular, exposure to air toxics early in life predisposes children to asthma, one of the most prevalent diseases in this demographic group. Epidemiological studies have linked prenatal and early-life exposure to air toxics with childhood wheeze, asthma, and altered lung function (6–14).

Although air toxics are generally analyzed and regulated as individual chemicals (6), we are exposed to combinations of air toxics in ambient air. The specific combinations of individual air toxics that influence childhood asthma have not been studied adequately. Assessing the respiratory health effects of multiple air toxics is challenging for several reasons (7, 15). First, it is logistically difficult and expensive to collect detailed, individualized exposure data for multiple air toxics using personal or local monitoring. Additionally, there are limited statistical methods to parse the effects of mixtures where individual air toxics may contribute only slightly to an adverse outcome but have a different impact in combination with other air toxics (15). As a result, few studies have considered exposure to air toxic mixtures and their associations with children's health, including asthma (8, 15–17).

Several studies linking air toxic mixtures and health outcomes, as well as a prior review of 57 studies that examined air pollutants and their health effects, reached no consensus on the ideal methods for multi-pollutant analyses (6, 7, 15, 16, 18). A key limitation of the studies reviewed is that most metrics assume pure additivity of the effects of multiple air toxics, without consideration of synergistic and/or antagonistic interactions. Because of these challenges, air toxic combinations that collectively influence childhood asthma remain suboptimally characterized. Furthermore, identifying air toxic combinations associated with health outcomes is also difficult because of the exponentially large number of combination subsets in a set of air toxics, i.e., $2^N - 1$ combinations in a set of N air toxics. Conventional statistical methods (19–24) and feature importance assessment using machine learning algorithms (16, 25–27) have not been effective for this task, since they generally assess the association of air toxics individually.

In this study, we hypothesized that exposure to combinations of air toxics during early life is associated with asthma outcomes in later childhood. These outcomes included current need for daily asthma controller medication, lifetime emergency room visit for asthma, and lifetime overnight hospitalization for asthma (Figure 1). Asthma-related medication use, emergency room visits, and hospitalizations are frequently studied asthma outcomes that reflect asthma severity, control, and health care usage (28–30). Although some studies have reported associations
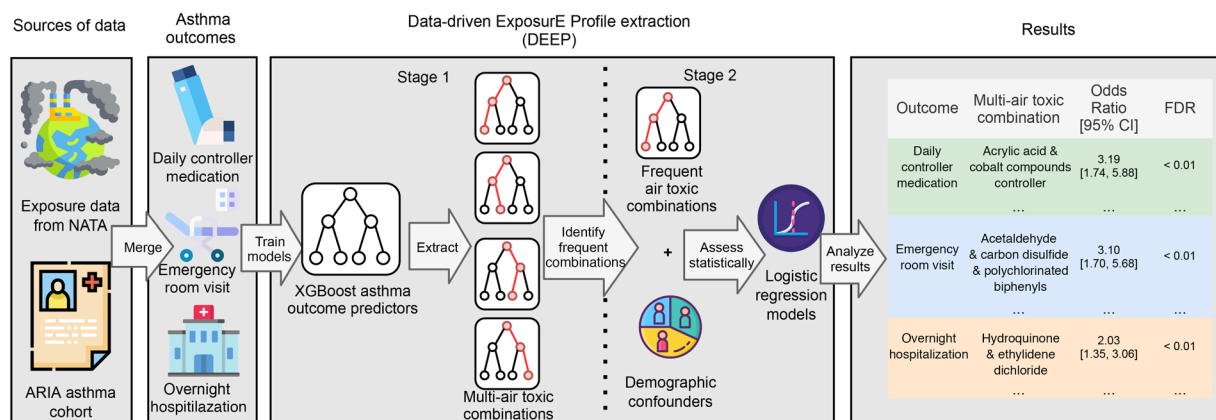
Figure 1. Study overview. Exposure data for over a hundred air toxics from the US Environmental Protection Agency's (EPA) National Air Toxic Assessment (NATA) database were geocoded to AiRway in Asthma (ARIA) cohort participants with mild to severe persistent asthma (*n* = 151), based on participants' residential zip code. The Data-driven ExposurE Profile extraction (DEEP) method developed in this study was then applied to the air toxic data to identify multi–air toxic combinations associated with 3 childhood asthma outcomes: use of prescribed daily asthma controller medication, lifetime emergency department visit for asthma, and lifetime overnight hospitalization for asthma. In the first stage of DEEP, multi–air toxic combinations were identified via eXtreme Gradient Boosting (XGBoost) models consisting of decision trees. In the second stage, multivariable logistic regression models were used to identify air toxic combinations significantly associated with childhood asthma outcomes after adjustment for age, sex, race and ethnicity, and family income. (Some images in this figure were obtained from the open-source collection at https://www.flaticon.com and were made by Wanicon, Freepik, and flaticon.)

between particular air toxics and these asthma subphenotypes (9, 11, 12), none addressed our goal to identify combinations of air toxics from a large national assessment of air toxics associated with these asthma outcomes. We tested our hypothesis by geocoding levels of 125 air toxics from the US EPA's NATA (31), one of the richest sources of multi–air toxic profiling across the United States, to the residential addresses of children with asthma from our ARIA study (32) to map each child's exposure to air toxics during the first years of life. We addressed the challenges of combinatorial air toxic analysis by applying a machine learning–based algorithm called DEEP, which, to the best of our knowledge, is a novel method for this problem. DEEP uses the high-performing XGBoost (33) algorithm to identify air toxic combinations associated with health outcomes. The combinations identified using XGBoost were then adjusted for potential confounders, including age, sex, race and ethnicity, and family income, to identify early-life multi–air toxic combinations, statistical interactions within combinations, and demographic profiles associated with adverse asthma outcomes in later childhood. Our approach identified several combinations of air toxics associated with asthma.

Results

Characteristics of the study cohort. Table 1 shows the characteristics of the ARIA study (32) participants with asthma examined in this study. These 151 children with mild to severe persistent asthma were recruited from the Mount Sinai Health System, New York, New York, USA, with informed consent from their parents/guardians via an IRB-approved protocol. Participants had a mean age of 12 years (standard deviation 3.2 years) at the time of assessment and were of diverse self-identified racial and ethnic backgrounds (Table 1). Their asthma was generally not well controlled, with a mean score on the ACT (34) of 16.8 (maximum value 25 representing optimal control) and 96% of the cohort reporting regular use of a short-acting β-agonist rescue inhaler.

Children who used daily asthma controller medication (*n* = 84, 56%) were younger than those who did not (*n* = 65; *P* = 0.048). ICSs were used most frequently, both independently and in combination with LABAs. Children who had at least 1 lifetime emergency room visit for asthma (*n* = 103, 68%) were more likely to self-identify as Black or Latino, had lower (*P* = 0.03) ACT scores than their counterparts who had never required an emergency department visit for asthma (*P* = 9.54×10^{-3}), and were more likely to be taking combination ICS/LABA as their daily asthma controller medication (*P* = 5.97×10^{-3}). Children who had been hospitalized overnight for asthma in their lifetime (*n* = 51, 34%) had significantly lower FEV1% on spirometry (*P* = 0.04), and higher rates of ICS/LABA (*P* = 2.98×10^{-7}) and leukotriene receptor antagonist (*P* = 3.08×10^{-4}) use for daily asthma treatment, compared with the participants with asthma who had never been hospitalized overnight for asthma.

Air toxic characteristics. Ambient annual average concentrations for over a hundred air toxics based on emissions inventories and computer simulation models are publicly available for each US census tract in the EPA's NATA database (31). We mapped the available toxic levels to the residential zip code for each child in our cohort. Ninety-four zip codes spanning 443 square miles across New York, New Jersey, and Connecticut were represented in this cohort. We used the closest calendar year of NATA data available subsequent to a child's birth date. We retained only the air toxics whose levels were available for all the participants in the mapped data sets, yielding 125 air toxics for analysis.

DEEP-enabled identification of combinations of air toxics associated with childhood asthma. We then applied DEEP to identify air toxic combinations associated with each of the 3 childhood asthma outcomes, namely the need for daily asthma controller medication, lifetime emergency room visit for asthma, and lifetime overnight hospitalization for asthma. In the first analytical stage of DEEP (detailed in Methods), for each outcome, the full data

Table 1. Characteristics of the ARIA cohort participants included in this study

	All	Daily controller medication[A]			Emergency room visit			Overnight hospitalization		
		Yes	No			Yes	No		Yes	No	
	(*n* = 151)	(*n* = 84)	(*n* = 65)	*P* value	(*n* = 103)	(*n* = 48)	*P* value	(*n* = 51)	(*n* = 100)	*P* value
Age, years	12.0 (3.2)	11.5 (3.2)	12.6 (3.1)	0.048	11.7 (3.3)	12.4 (2.9)	0.23	12.1 (3.2)	11.9 (3.2)	0.69
Sex, female	62 (41.1%)	33 (39.3%)	29 (44.6%)	0.63	41 (39.8%)	21 (43.8%)	0.78	19 (37.3%)	43 (43%)	0.61
Race and ethnicity				0.50			0.03			0.07
Asian	6 (4.0%)	4 (4.8%)	2 (3.1%)		3 (2.9%)	3 (6.3%)		1 (2.0%)	5 (5.0%)	
Black	26 (17.2%)	13 (15.5%)	13 (20%)		20 (19.4%)	6 (12.5%)		12 (23.5%)	14 (14.0%)	
White	57 (37.7%)	33 (39.3%)	24 (36.9%)		32 (31.1%)	25 (52.1%)		12 (23.5%)	45 (45.0%)	
Latino	51 (33.8%)	25 (29.8%)	24 (36.9%)		42 (40.8%)	9 (18.8%)		23 (45.1%)	28 (28.0%)	
Mixed	10 (6.6%)	8 (9.5)	2 (3.1%)		5 (4.9%)	5 (10.4%)		3 (5.9%)	7 (7.0%)	
Not reported	1 (0.7%)	1 (1.2%)	0 (0%)		1 (1.0%)	0 (0.0%)		0 (0%)	1 (1.0%)	
Income, US$	49,349.3 (18,725.8)	48,954.8 (18,701.7)	49,961.7 (19,017.3)	0.75	48,870.9 (19,414.5)	50,375.8 (17,109.8)	0.63	45,576.7 (14,822.9)	51,273.3 (20,161.2)	0.053
Asthma Control Test (ACT) score	16.8 (3.9)	16.4 (3.9)	17.3 (4.0)	0.16	16.2 (4.0)	18.0 (3.5)	9.54×10^{-3}	15.9 (4.3)	17.3 (3.7)	0.05
Forced expiratory volume in 1 second percent predicted (FEV1%)	87.1 (17.5)	85.5 (17.7)	88.6 (17.3)	0.32	86.2 (18.0)	88.6 (16.6)	0.45	82.1 (19.7)	89.4 (15.9)	0.04
FEV1/forced vital capacity (FVC)	79.6 (10.4)	78.3 (10.5)	80.8 (10.2)	0.17	79.4 (10.9)	80.0 (9.5)	0.77	77.7 (10.5)	80.4 (10.3)	0.17
Regular use of asthma medicine										
β-Agonist	145 (96.0%)	83 (98.8%)	60 (92.3%)	0.11	102 (99.0%)	43 (89.6%)	0.02	51 (100%)	94 (94.0%)	0.18
Inhaled corticosteroid (ICS)	39 (25.8%)	35 (41.7%)	3 (4.6%)	7.19×10^{-7}	27 (26.2%)	12 (25.0%)	1	10 (19.6%)	29 (29.0%)	0.29
Combined ICS/long-acting β-agonist (LABA)	31 (20.5%)	30 (35.7%)	0 (0%)	2.16×10^{-7}	28 (27.2%)	3 (6.3%)	5.97×10^{-3}	23 (45.1%)	8 (8.0%)	2.98×10^{-7}
Leukotriene receptor antagonist	30 (19.9%)	29 (34.5%)	1 (1.5%)	1.81×10^{-6}	23 (22.3%)	7 (14.6%)	0.37	19 (37.3%)	11 (11.0%)	3.08×10^{-4}
Omalizumab	2 (1.3%)	1 (1.2%)	1 (1.5%)	1	2 (1.9%)	0 (0.0%)	0.84	0 (0.0%)	2 (2.0%)	0.79

Mean (standard deviation) and number (%) are shown for continuous and categorical variables, respectively. Characteristics stratified by asthma outcome are also shown. *P* values comparing patients with and without the outcome are shown from 2-sided Student's *t* test for continuous variables and a χ^2 test for categorical variables. [A]Two participants did not provide information about their daily controller medication use.

set was randomly split 100 times into training and test sets in an 80:20 ratio. For each split, an XGBoost model consisting of 100 decision trees was learned from the training set and evaluated on the test set in terms of the area under the receiver operating characteristic (ROC) curve (AUC score; ref. 35).

In the second analytical stage of DEEP, we analyzed the combinations of toxics from the XGBoost models, identified as root-to-leaf paths in the constituent decision trees, for each outcome. Note that in some cases, a combination may consist of only 1 air toxic if it is sufficient to predict the outcome under consideration for a subset of the cohort, thus giving DEEP flexibility in discovery. Also, in cases of multiple air toxics in these combinations, their sequence of appearance on the path also indicates their relative order of relevance to the outcome being predicted. This is because variables closer to the root of a decision tree have higher predictive power than those closer to the leaves.

Next, the frequency of each combination was calculated as the number of models (out of 100) where the combination was included in at least 1 of the constituent trees. Candidate combinations were then identified as those with a frequency of at least 10. These combinations were then used in multivariable regression models to test

their association with the asthma outcome of interest, while adjusting for age, sex, race and ethnicity, and income.

After the first XGBoost stage of DEEP, 689 profiles of air toxics across all the asthma outcomes were discovered. These sets included both individual air toxics and their combinations. In the second stage of DEEP, 359 of these sets were then found to be significantly associated ($P \le 0.05$) with the respective outcome. After multiple-hypothesis correction by the Benjamini-Hochberg procedure (36), 273 air toxic profiles were found to be significantly associated (FDR ≤ 0.05) with at least 1 of the 3 outcomes. Our goal was to identify air toxic combinations whose increased levels are associated with adverse asthma outcomes. Therefore, among the significantly associated combinations, we focused on groups that included air toxics with levels higher than threshold. Among these final determined combinations, 18 had 1 air toxic each (Figure 2), and 20 were multi–air toxic combinations (Figure 3).

Air toxic combinations associated with asthma outcomes. Twenty multi–air toxic combinations and 18 individual air toxics were found to be significantly associated with at least 1 of the 3 asthma outcomes. The medians and interquartile ranges (IQRs) of the
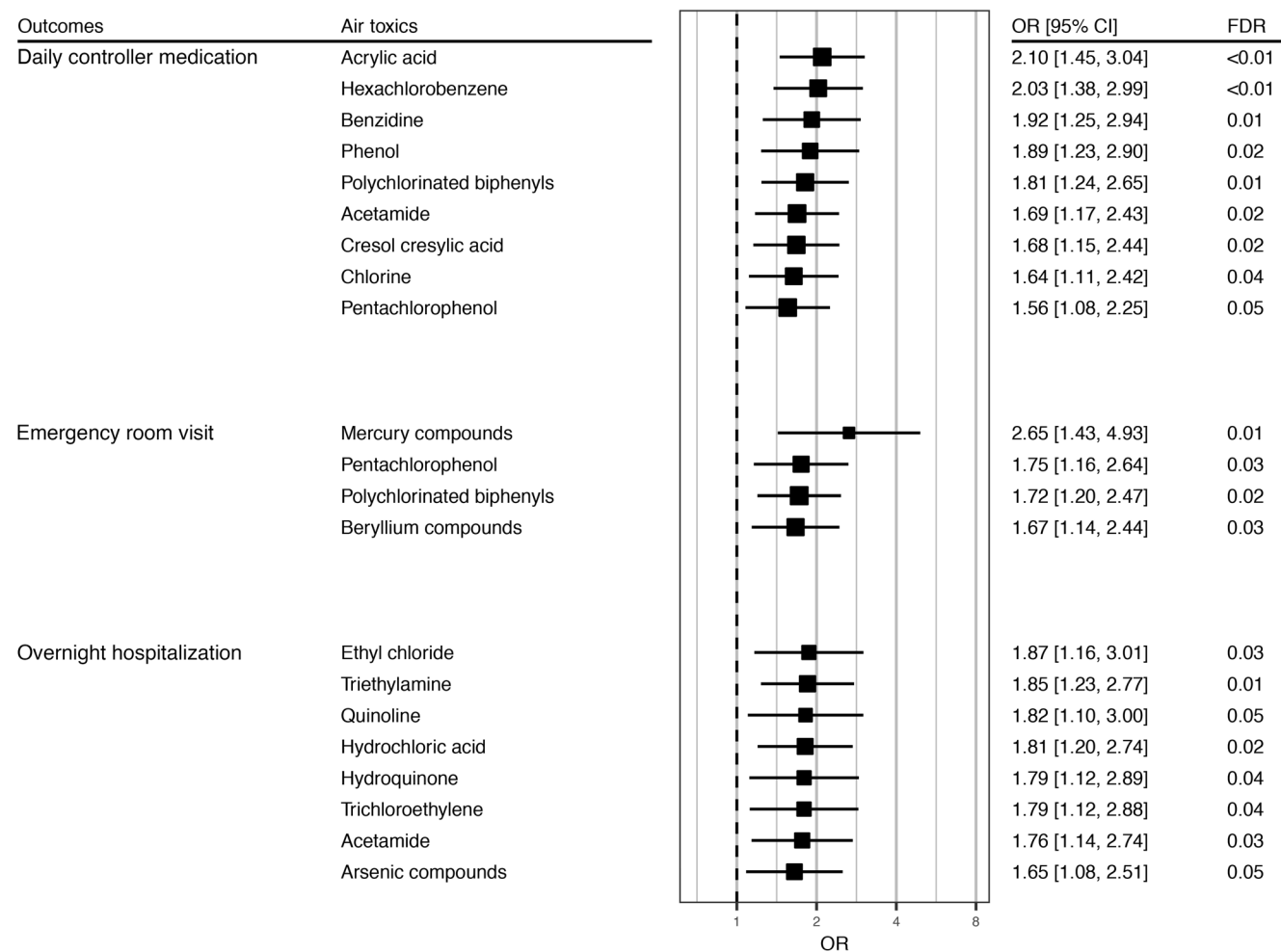
Figure 2. Air toxics individually associated with childhood asthma outcomes after adjustment for age, sex, race and ethnicity, and family income in ARIA cohort participants with persistent asthma (*n* = 151). For each outcome and air toxic, the strength of the association is shown in terms of its odds ratio (OR), 95% confidence interval (CI), and false discovery rate (FDR). *P* values for individual air toxics were obtained from multivariable logistic regression models and then adjusted for multiple hypothesis testing using the Benjamini-Hochberg procedure, yielding FDR values.

exposure levels of the 34 air toxics included in these associations are shown in Table 2.

Higher levels of 17 individual air toxics were significantly associated with worse asthma outcomes (Figure 2). ORs for these associations ranged from 1.56 to 2.65. Several of the identified toxics are established risk factors for childhood asthma, especially the chemicals previously categorized as halogenated, ketones, and ethers (8, 37–39). Among these, the air toxics most associated with the outcomes were acrylic acid (OR = 2.10), mercury compounds (OR = 2.65), and ethyl chloride (OR = 1.87), respectively. Acetamide, pentachlorophenol, and polychlorinated biphenyls were associated with more than 1 asthma outcome.

A major strength of DEEP is its ability to identify multi–air toxic combinations associated with health outcomes. Indeed, here DEEP revealed significant associations between higher exposure to 20 multi–air toxic combinations and the 3 asthma outcomes of interest (Figure 3). Among these, 19 combinations included 2 air toxics and 1 included 3. The associations of these combinations were generally stronger than those of the individual air toxics, with ORs ranging from 1.60 to 3.19 (Figure 3).

Notably, acrylic acid not only was the individual air toxic most strongly associated with daily controller medication (Figure 2) but also was the first (i.e., primary) member of 7 of the 9 multi–air toxic combinations associated with this outcome (Figure 3). Acrylic acid also appeared in 3 of the other 11 combinations associated with emergency room visit and overnight hospitalization for asthma (Figure 3), indicating that it is a major contributor to adverse asthma outcomes among children.

Three air toxic combinations were associated with lifetime emergency room visit for asthma, all with an OR of over 2 (Figure 3). Acetaldehyde, acrylamide, and acrylic acid were the primary exposures in these combinations, despite the fact they were not individually significantly associated with the outcome. Several other air toxics in these combinations, namely carbon disulfide and hydroquinone, were also not individually associated with this outcome. These findings highlight the main strength of DEEP, namely its ability to identify significant multi–air toxic combinations, whose constituent air toxics may not be individually associated with the health outcome of interest.

Among the 8 air toxic combinations associated with lifetime overnight hospitalization for asthma, 1,4-dioxane, carbonyl sul-
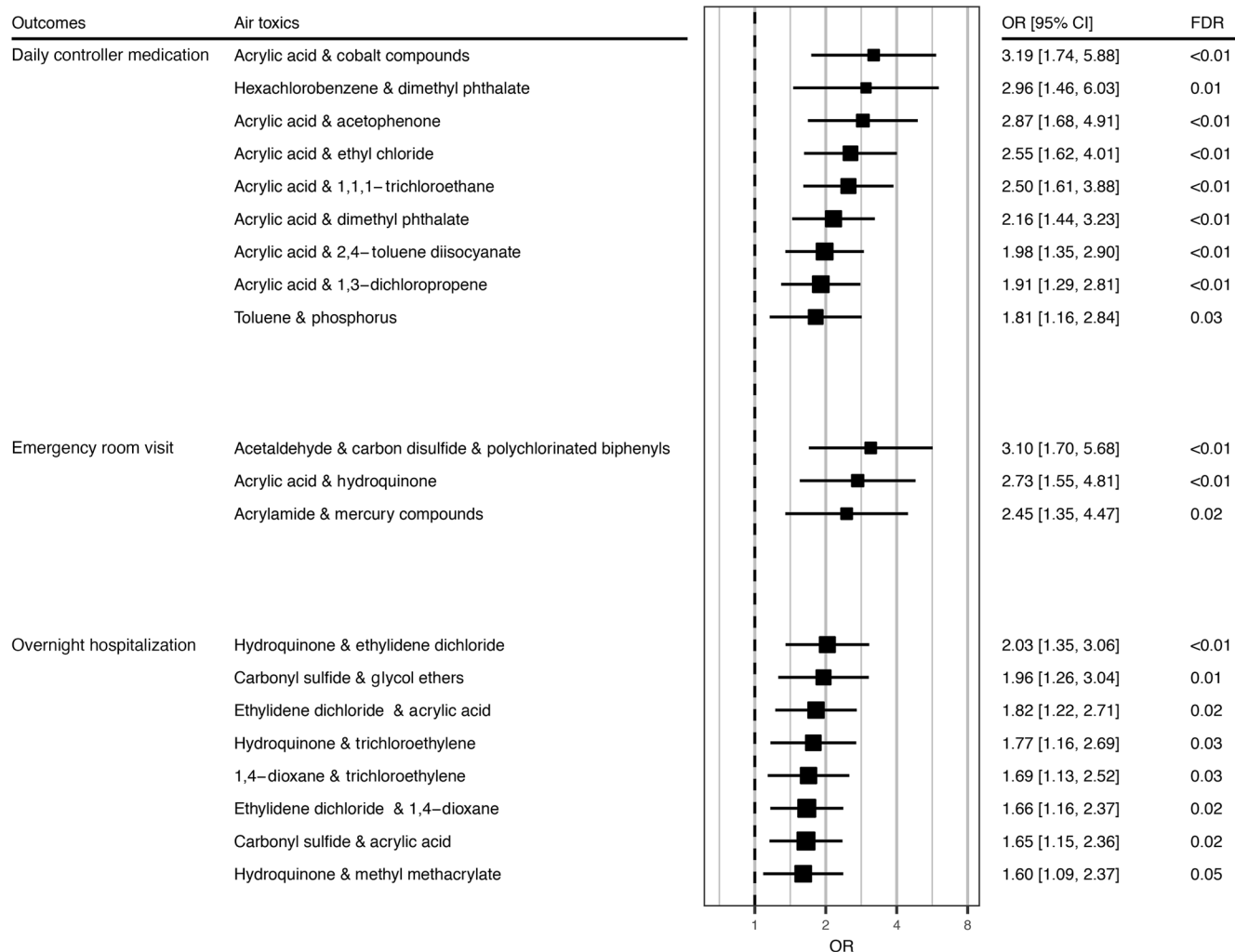
Figure 3. Multi–air toxic combinations associated with childhood asthma outcomes after adjustment for age, sex, race and ethnicity, and family income in ARIA cohort participants with persistent asthma (*n* = 151). For each outcome and combination, the strength of the association is shown in terms of its odds ratio (OR), 95% confidence interval (CI), and false discovery rate (FDR). *P* values for multi–air toxic combinations were obtained from multivariable logistic regression models and then adjusted for multiple hypothesis testing using the Benjamini-Hochberg procedure, yielding FDR values.

fide, ethylidene dichloride, hydrochloric acid, and hydroquinone were the primary exposures (Figure 3). Both ethylidene dichloride and hydroquinone appeared in 3 of these 8 combinations, indicating that these 2 chemicals may play a role in the development of poor asthma outcomes among children. Most other air toxics in these combinations (Figure 3) were largely not individually associated with this outcome (Figure 2), again supporting DEEP's ability to identify multi–air toxic combinations that may not be inferred from single air toxic associations.

Effect sizes of multi–air toxic combinations may not be evident from the individual associations of their members. Some air toxics had relatively low effect sizes when assessed individually (Figure 2) compared with the larger ORs from combination analyses (Figure 3). For example, acrylic acid was associated with daily controller medication, with an OR of 2.10 as an individual air toxic (Figure 2), but the ORs of its combinations with dimethyl phthalate, 1,1,1-trichloroethane, ethyl chloride, acetophenone, and cobalt were higher (OR 2.16 to 3.19; Figure 3). Also, none of these 5 air toxics was individually associated with the out-

come. Similarly, hexachlorobenzene was associated with daily controller medication with an OR of 2.03 (Figure 2), while simultaneous exposure to the combination of hexachlorobenzene and dimethyl phthalate identified by DEEP had an OR of 2.96 (Figure 3). This was despite the fact that there was no significant individual association between dimethyl phthalate and the outcome. For the pair of toluene and phosphorus, neither air toxic was individually associated with daily controller medication (Figure 2), but their combination was associated with the outcome with an OR of 1.81 (Figure 3).

Similar cases of combinatorial effects were also seen for lifetime emergency room visit for asthma. For example, simultaneous exposure to polychlorinated biphenyl, acetaldehyde, and carbon disulfide had 3.10-fold higher odds of the outcome (Figure 3), while polychlorinated biphenyl's individual effect size was substantially lower (OR = 1.72; Figure 2). Similarly, the combination of acrylic acid and hydroquinone was significantly associated with emergency room visit with an OR of 2.73 (Figure 3), but neither was associated with the outcome individually (Figure 2).

Table 2. Air toxics identified by DEEP as significantly associated with at least 1 of the 3 asthma outcomes, either individually or in combination with other air toxics

Air toxic	Level at residential zip code Median (IQR) values (in ng/m³)	Individually significant	Significant in combination(s)
1,1,1-trichloroethane	667.7 (120.5, 1452.9)		D
1,3-dichloropropene	129 (1.75, 405.18)		D
1,4-dioxane	3.16×10^{-3} (1.3×10^{-3}, 0.18)		O
2,4-toluene diisocyanate	0.08 (7.2×10^{-3}, 0.15)		D
Acetaldehyde	2420.6 (1967.0, 2768.3)		E
Acetamide	2.61×10^{-4} (1.7×10^{-4}, 9.1×10^{-4})	D, O	
Acetophenone	0.1 (0.02, 0.17)		D
Acrylamide	4.24×10^{-4} (2.5×10^{-6}, 6.6×10^{-3})		E
Acrylic acid	1.27×10^{-3} (4.98×10^{-4}, 0.06)	D	D, E, O
Arsenic compounds	0.73 (0.49, 1.13)	O	
Benzidine	1.44×10^{-5} (1.2×10^{-5}, 2.5×10^{-5})	D	
Beryllium compounds	0.16 (0.09, 0.27)	E	
Carbon disulfide	5.66 (5.59, 9.57)		E
Carbonyl sulfide	0.03 (1.77×10^{-3}, 0.25)		O
Chlorine	0.91 (0.39, 1.28)	D	
Cobalt compounds	0.15 (0.10, 0.19)		D
Cresol cresylic acid	20.05 (15.1, 24.9)	D	
Dimethyl phthalate	0.07 (0.04, 0.08)		D
Ethyl chloride	8.21 (7.8×10^{-3}, 20.26)	O	D
Ethylidene dichloride	0.12 (0.02, 0.22)		O
Glycol ethers	12.67 (2.58, 117.36)		O
Hexachlorobenzene	1.2×10^{-6} (2.47×10^{-7}, 3.91×10^{-4})	D	D
Hydrochloric acid	208.8 (146.8, 301.7)	O	
Hydroquinone	7.3×10^{-3} (1.4×10^{-3}, 0.03)	O	E, O
Methyl methacrylate	1.62 (1.06, 2.79)		O
Mercury compounds	0.4 (0.22, 1.87)	E	E
Pentachlorophenol	4.6×10^{-7} (0, 1.4×10^{-6})		D, E
Phenol	3.34 (2.03, 22.35)	D	
Phosphorus	0.21 (0.11, 0.31)		D
Polychlorinated biphenyls	1.1×10^{-4} (1.2×10^{-5}, 8.8×10^{-3})	D, E	E
Quinoline	0.14 (5.23×10^{-4}, 0.17)	O	
Toluene	9392 (4984.58, 21,249.3)		D
Trichloroethylene	106.4 (18.1, 132.6)	O	O
Triethylamine	2.81 (1.82, 5.77)	O	

D, air toxic associated with daily asthma controller medication; E, air toxic associated with lifetime emergency room visit for asthma; O, air toxic associated with lifetime overnight hospitalization for asthma.

We observed similar results for multi–air toxic combinations and lifetime overnight hospitalization for asthma. Exposure to hydroquinone was individually associated with this outcome with an OR of 1.79 (Figure 2), but in combination with ethylidene dichloride, the association was stronger (OR = 2.03; Figure 3). Similarly, carbonyl sulfide was not individually associated with this outcome (Figure 2), but it was the primary member in 2 of the multi–air toxic combinations found to be associated with overnight hospitalization (Figure 3).

In summary, the above comparison of the effect sizes of the individual air toxic (Figure 2) and multi–air toxic (Figure 3) associations demonstrated that combinations of air toxics had effects that were not fully explained by simply adding together the individual

effects from their constituents. Overall, DEEP identified 34 air toxics associated with the asthma outcomes (Table 2), including 16 air toxics with significant effects only as members of combinations.

Statistical interactions among members of air toxic combinations. To assess potential synergy between members of air toxic combinations associated with asthma outcomes, we conducted statistical tests for interactions. Significant statistical interactions detected between air toxic members within the combinations are shown in Table 3. Acrylic acid was the primary air toxic (i.e., primary branch point in the decision tree) of all the combinations with significant statistical interactions. Although other combinations did not reveal significant interactions, such interactions remain possible given the limitations of statistical detection of interactions. Directed experimental work could be undertaken to test for additional interactions.

Representative air toxic combinations and demographic risk factors. Finally, one of the advantages of DEEP is that the trees constituting its underlying XGBoost models can be visualized and interpreted, which is difficult to do for several other machine learning methods. However, since it is difficult to simultaneously depict all the trees inferred by DEEP, we visualized sample trees that contained the most strongly associated multi–air toxic combination for each childhood asthma outcome. Sample decision trees inferred by DEEP for each of the outcomes are shown in Figures 4, 5, and 6, respectively. To provide an additional level of interpretation, we also compared the demographic characteristics (age, sex, race and ethnicity, and family income) of children exposed to each of these combinations with those of children who were not exposed (Tables 4, 5, and 6). Differences could suggest demographic risk factors that may increase a child's exposure to these multi–air toxic combinations.

Acrylic acid and cobalt compounds was the air toxic combination associated with daily controller medication use with the highest OR of 3.19 (Figures 3 and 4). Children exposed to this combination were older than those who were not exposed ($P = 0.02$; Table 4).

Acetaldehyde, carbon disulphide, and polychlorinated biphenyls was the air toxic combination most strongly associated with lifetime emergency room visit for asthma (OR = 3.10; Figure 3 and Figure 5). Children exposed to this combination were younger ($P = 5.34 \times 10^{-8}$; Table 5) and had lower family income than those who were not exposed ($P = 0.019$; Table 5). Exposed children were also less likely to be White ($P = 0.0046$; Table 5). These observations point to social disparities among these groups of children.

The most strongly associated combination for overnight hospitalization was hydroquinone and ethylidene dichloride (OR = 2.03; Figure 3 and Figure 6). Children exposed to this combination were younger ($P = 0.00218$; Table 6) and had lower family incomes ($P = 8.26 \times 10^{-5}$; Table 6) than those who were not exposed.

Table 3. Air toxic combinations associated with asthma outcomes with statistically significant interactions between combination members

Outcome	Air toxic combination	Interactions	P value
Daily controller medication	Acrylic acid & dimethyl phthalate	Acrylic acid & dimethyl phthalate	0.02
Daily controller medication	Acrylic acid & cobalt compounds	Acrylic acid & cobalt compounds	0.02
Emergency room visit	Acrylic acid & hydroquinone	Acrylic acid & hydroquinone	<0.004

The P values are for the interaction term of multivariable logistic regression models where asthma outcome was the dependent variable and independent variables included member air toxic levels, interaction term, and covariates (age, sex, race and ethnicity, family income).

Discussion

Our application of a machine learning–driven algorithm called DEEP to a cohort of children with mild to severe asthma identified several individual air toxics and combinations of air toxics, to which increased exposure during early life was associated with adverse asthma outcomes in later childhood. In particular, due to the unique ability of DEEP to examine air toxic combinations, we identified 16 air toxics that were found to be significantly associated with childhood asthma outcomes only in combination with other air toxics.

Many air toxics in the identified combinations, such as carbonyl sulfide, carbon disulfide, ethyl chloride, and ethylidene chloride, are similar in structure and have analogous formation, production, chemical fate, and chemical transport properties (40). Ten air toxics in the combinations contained chlorine, 3 included heavy metal compounds, and many were acidic chemicals. This aligns with prior literature implicating acidic chemicals, chlorinated chemicals, and heavy metal compounds as risk factors for asthma and asthma severity (8, 41–44). However, the biological mechanisms through which these combinations of air toxics can jointly affect respiratory health and asthma merit further study.

Among the air toxics individually associated with asthma outcomes (Figure 2), triethylamine was associated with increased overnight hospitalizations for asthma. Triethylamine is a clear, colorless liquid used in waterproofing and as a catalyst, corrosion inhibitor, and propellant (45). It is a respiratory irritant, to which chronic exposure even at low levels can inhibit the function of organic cationic transporters, thus preventing efficient uptake of inhaled bronchodilators used to control acute asthma symptoms (46, 47).

Acrylic acid was individually associated with daily controller medication (Figure 2) and appeared as a member of at least 1 combination associated with all 3 outcomes (Figure 3). Furthermore, it was found to interact with other member air toxics of 3 combinations (Table 3). Acrylic acid is used in the manufacture of adhesives, elastomers, plastics, and coatings, as well as floor paints and polishers (48). Literature has suggested that the presence of water-soluble cobalt complexes increases the conversion of polyacrylic acid into acrylic acid, which is more biologically viable. Acrylic acid also reacts with cobalt complexes to produce organocobalt complexes (49). Additionally, hydroquinone acts as a stabilizer to prevent the polymerization of acrylic acid, which keeps the latter in a form with a lower molecular weight that is more biologically viable (50, 51). Our results, including evidence of statistical interactions between acrylic acid and other chemical compounds, suggest further investigation of mechanisms for acrylic acid's associations with adverse childhood asthma outcomes.

Ethyl chloride, also known as chloroethane (C_2H_5Cl), and ethylidene dichloride ($C_2H_4Cl_2$) are both chlorinated hydrocarbons. Ethyl chloride is used as a thickening agent and binder in paints and cosmetics, refrigerant, aerosol spray propellant, anesthetic, and blowing agent for foam packaging (52). We found ethyl chloride to be associated with asthma outcomes, both as an individual air toxic (overnight hospitalizations; Figure 2) and as a member of multi–air toxic combinations (daily controller medication; Figure 3). Ethylidene dichloride, which is used mainly as a solvent for plastics, oils, and fats, and as a degreaser and fumigant in insecticide sprays (53), appeared as a member of several combinations associated with lifetime overnight hospitalization for asthma (Figure 3). Both these compounds are well-known members of the chloroethane family, which comprises liposoluble chemicals that can be taken up by the lipoprotein within the alveolar film layer (AFL) (54). AFL disruption is observed in multiple pulmonary diseases, including acute respiratory distress syndrome, infant respiratory distress syndrome, emphysema, chronic obstructive pulmonary disease, asthma, chronic bronchitis, pneumonia, pulmonary infections, and idiopathic pulmonary fibrosis (55). Thus, chronic exposure to ethyl chloride and ethylidene dichloride may lead to dysfunction in the AFL, which may contribute to worse asthma control.

Hydroquinone, a commonly studied air toxic, was also identified in our analyses. While exposure to higher levels of hydroquinone alone was not associated with overnight hospitalization for asthma, DEEP identified it as a member of several multi–air toxic combinations associated with this outcome (Figure 3). Hydroquinone is commonly found in the indoor environment, and exposure to it has been associated with airway hypersensitivity (56–59). Hydroquinone is widely seen in cosmetic and health products, including skin creams (60). It is thought to prevent the polymerization of acrylic acid, methyl methacrylate, cyanoacrylate, and other monomers that are susceptible to radical-initiated polymerization, thus allowing them to persist in their original form (50, 51). This suggests a mechanism through which the identified synergistic combination of hydroquinone with acrylic acid (emergency room visit; Figure 3) is associated with adverse asthma outcomes. Although our analysis did not find a statistically significant interaction between hydroquinone and ethylidene dichloride, potentially due to lower exposure levels, hydroquinone is industrially added to shelf ethylidene dichloride as a stabilizer (61). This suggests that in the presence of hydroquinone, ethylidene dichloride
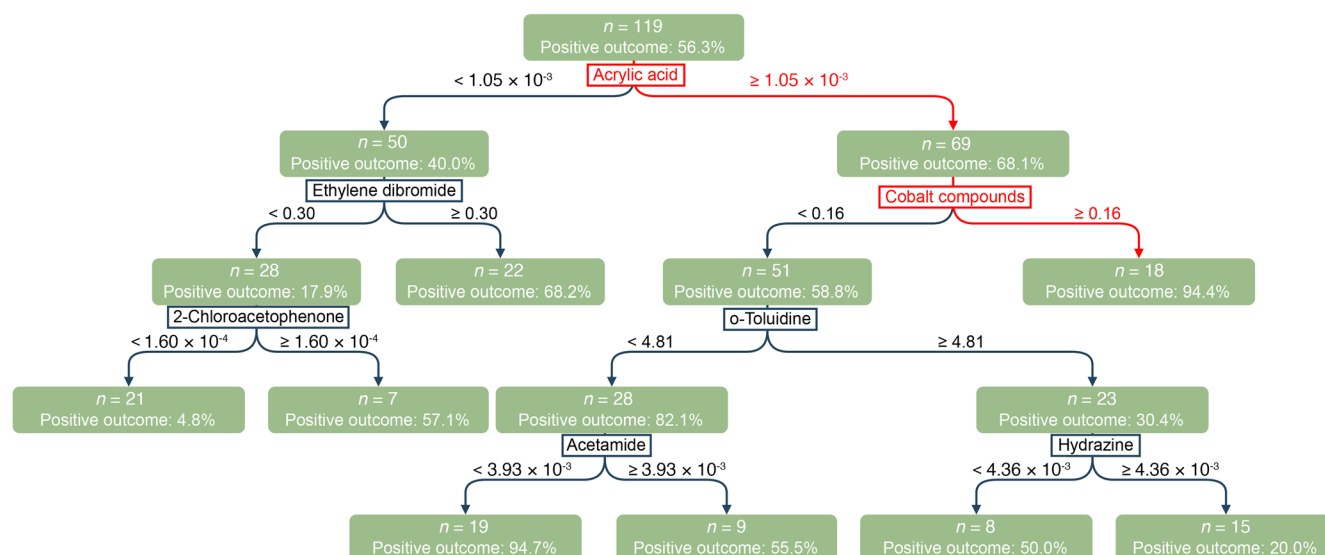
JCI

Figure 4. A sample decision tree learned by DEEP to predict daily asthma controller medication using NATA-derived air toxic data geocoded to patients (*n* = 149). Each node in the tree indicates the number of participants satisfying the air toxic decision path until that point and the percentage of participants with that outcome. The sample corresponding to each node is stratified into 2 subpopulations based on the air toxic and its threshold associated with the node. The multi–air toxic combination acrylic acid and cobalt compounds, which was most significantly associated with this outcome, is highlighted in red.

is less likely to react with other chemicals in the environment and thus retains its toxic form longer (similar to acrylic acid). Thus, it is still possible that hydroquinone and ethylidene dichloride may act synergistically, but this needs to be investigated in future studies.

Although our study has advanced the identification of air toxic combinations associated with childhood asthma outcomes, it also has limitations. We used the NATA model (31) to estimate exposures rather than personal sampling or local monitors. Collecting personal or locally monitored measures for 125 air toxics at each cohort participant's residence would be logistically and financially challenging. Given this, NATA is commonly used for estimating ambient exposures, since it is a well-validated deterministic dispersion chemical transportation model created by the EPA that accounts for sources included in the NATA emission inventory (31). NATA estimates of a given air toxic may underreport a personally or locally monitored value, since the latter may include emissions from indoor and undocumented sources not in the EPA's inventory. For instance, higher personally monitored benzene concentrations relative to NATA-predicted values are likely due, at least in part, to indoor sources not included in the EPA's inventory (62). Other studies also found discrepancies between NATA estimates and monitored chemical concentrations, due again, in part, to local or indoor sources (63, 64).

Despite the above limitations, using NATA as the primary source of exposure estimates has several strengths over locally monitored values. First, NATA has a finer geographical prediction resolution and spread than currently available monitoring sites (31). This enabled us to include participants in our study who may not have had a monitoring site close to their residence. Second, NATA data are generated from an advanced chemical transportation model that aggregates exposure over a long period and thus is able to capture transient exposures. Also, several factors potentially affecting air toxic estimates, such as seasonality, ambient

temperature, meteorology, precipitation, and solar radiation, have already been incorporated into NATA's model (31). This level of comprehensive modeling is typically not available from personal or local monitoring. Finally, local measurements may also have detection and quantification limits, while NATA is able to estimate air toxics even at lower levels and over a longer period. Because of these strengths, we chose to study NATA data in this work.

We recognize that our results do not provide evidence of a causal effect of any chemical on adverse childhood asthma outcomes, which will need further investigation. Additionally, our geocoding was based on a single zip code for each participant and so would not account for potentially dynamic exposures due to residential moves. Last, our study participants were all from the New York, New Jersey, and Connecticut tristate area. Thus, our findings may not generalize to other US regions or parts of the world. Future studies could examine combinations from other geographical regions and/or utilize direct air sampling to confirm the combinations identified in this study.

In conclusion, this study demonstrated innovative use of data science methods and data sources to identify specific combinations of early-life air toxic exposures associated with later childhood asthma outcomes. Our study suggests that chemical pollutants should be closely monitored together in combination, especially in locations with vulnerable populations.

Methods

An overview of the study approach is shown in Figure 1.

Study population. The study population included children with asthma from the ARIA study, a cohort recruited from the Mount Sinai Health System, New York, New York, USA (32). The study was approved by the Mount Sinai IRB. Children with asthma had mild to severe persistent asthma according to National Asthma Education and Prevention Program/National Heart Lung Blood Institute Expert Panel Report
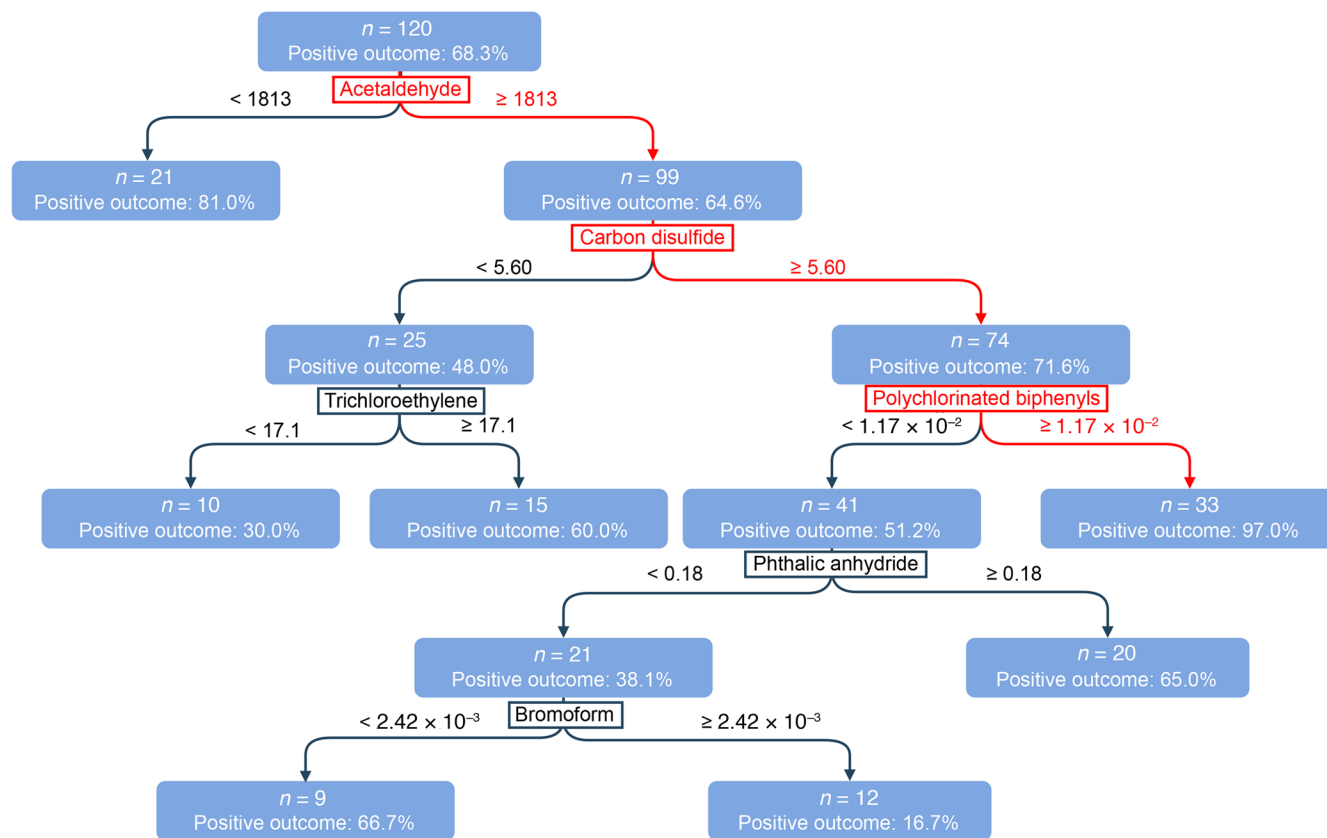
Figure 5. A sample decision tree learned by DEEP to predict lifetime emergency room visit for asthma from NATA-derived air toxic exposure data geocoded to each patient (*n* = 151). Each node in the tree indicates the number of participants satisfying the air toxic decision path until that point and the percentage of participants with that outcome. The sample corresponding to each node is stratified into 2 subpopulations based on the air toxic and its threshold associated with the node. The multi–air toxic combination acetaldehyde and carbon disulfide and polychlorinated biphenyls, which was most significantly associated with this outcome, is highlighted in red.

3 criteria (65) and positive bronchodilator response on spirometry or methacholine challenge, with provocative challenge causing a 20% fall in forced expiratory volume (PC20) < 12.5 mg/mL. Phenotyping for all participants included detailed questionnaires about asthma-related symptoms, medication and health care use, and pre- and post-bronchodilator spirometry following American Thoracic Society guidelines (65).

Asthma outcomes. We focused on the following 3 self-reported asthma-related outcomes: (a) current use of prescribed daily asthma controller medication (daily controller medication), (b) at least 1 emergency department visit for asthma during the patient's lifetime (emergency room visit), and (c) at least 1 overnight hospitalization for asthma during the patient's lifetime (overnight hospitalization).

Air toxic exposures. The air toxic exposure profile of each participant was derived from the EPA's NATA (31). NATA estimates the annual ambient concentrations of over a hundred air toxics at each census tract in the United States based on emissions inventories and advanced computer simulation models (31). Seasonality of air toxics, ambient temperature, meteorology, precipitation, and solar radiation are incorporated into NATA's model (31).

NATA data are available for 1996, 1999, 2002, 2005, 2011, and 2014 (27), and children in this study were born between 1997 and 2012. To assign the most representative ambient air toxic exposure levels to each participant, we mapped the residential zip code of each child in our cohort to the geometric centroid of the closest cen-

sus tract (31). We then used the annual exposure data for that tract from the NATA release closest in time following the child's birth year. This choice of year closest to birth was based on prior evidence that early-life exposure to air pollution is associated with childhood asthma outcomes (66–68). Finally, we retained the 125 air toxics that had data available, i.e., no missing data, for all participants in the final data set (15, 69).

Covariates. We included age, sex, race and ethnicity, and family income as covariates in multivariable regression models based on considerations that these variables could confound associations between air toxic levels and asthma outcomes. Since the questionnaire completed by ARIA participants did not include queries about family income, we used the average income of each participant's residential zip code obtained from US Census Business Patterns data (70) as a surrogate for this variable.

Data-driven ExposurE Profile extraction (DEEP). To identify multi–air toxic exposure profiles associated with asthma outcomes, we developed a data-driven method called DEEP (Figure 1). DEEP is inspired by a simpler method that we previously used to identify multi–air toxic combinations associated with children's cognitive skills (15).

In the first stage of DEEP, exposure combinations are identified using XGBoost (33), an algorithm that uses an ensemble method to iteratively learn decision trees. XGBoost generally yields strong predictive power (71, 72) due to its use of multiple optimization methods,
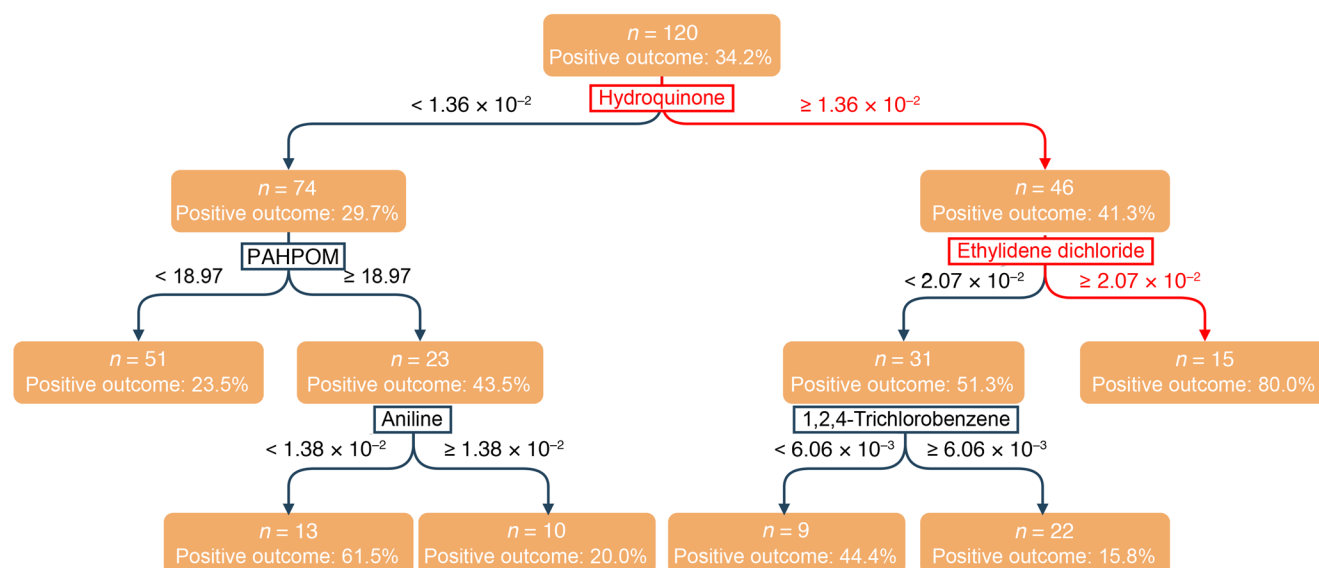
Figure 6. A sample decision tree learned by DEEP to predict lifetime overnight hospitalization for asthma from NATA-derived air toxic data geocoded to each participant (*n* = 151). Each node in the tree indicates the number of participants satisfying the air toxic decision path until that point and the percentage of patients with that outcome. The sample corresponding to each node is stratified into 2 subpopulations based on the air toxic and its threshold associated with the node. The multi–air toxic combination hydroquinone and ethylidene dichloride, which was most significantly associated with this outcome, is highlighted in red.

including regularization and gradient boosting, which reduces overfitting of models to training data. Specifically, the full exposure data set was randomly split 100 times into training and test sets in an 80:20 ratio. For each split, an XGBoost model consisting of 100 decision trees was trained from the training set to predict the outcome under consideration. This model was then applied to and evaluated on the corresponding test set in terms of the AUC score (35). The overall predictability of the target outcome was evaluated in terms of the average value of the AUC scores across the 100 training/test splits.

The decision trees constituting each XGBoost model contain internal decision nodes, edges, and leaf nodes to represent how the value of an outcome could be predicted based on air toxic levels. Figures 4–6 show several trees derived in the current work. Each decision node in these trees contains an air toxic and a threshold value for its level. It is also connected by 2 edges representing the decisions made depending on whether an individual's exposure was higher or lower than the threshold. Each of these edges is connected to either the next decision node or a leaf node. A leaf node determines the value of the outcome for the individual with the exposure profile represented by the decision path taken to reach it. Each decision and leaf node also represents a subpopulation of the cohort exposed to the air toxics on the path taken to reach it. Candidate multi–air toxic combinations are then defined as the air toxics and thresholds in the decision nodes constituting the paths from the root of a tree to the

Table 4. Demographic characteristics of children exposed and not exposed to the acrylic acid and cobalt compounds combination, which was associated with daily asthma controller medication

	Multi–air toxic combination: acrylic acid & cobalt compounds			
Variable	Total (*n* = 149)	Children exposed to combination (*n* = 23)	Children not exposed to combination (*n* = 126)	*P* value
Age, years	12.0 (3.2)	13.7 (3.6)	11.7 (3.0)	0.02
Sex, female	87 (58.4)	14 (60.9)	73 (57.9)	0.97
Race and ethnicity				0.089
Asian	6 (4.0)	0 (0.0)	6 (4.8)	
Black	26 (17.4)	5 (21.7)	21 (16.7)	
White	57 (38.3)	6 (26.1)	51 (40.5)	
Latino	49 (32.9)	8 (34.8)	41 (32.5)	
Mixed	10 (6.7)	3 (13.0)	7 (5.6)	
Not reported	1 (0.7)	1 (4.3)	0 (0.0)	
Income, US$	49,394.0 (18,846.6)	46,339.2 (15,671.1)	49,951.7 (19,318.1)	0.34

Mean (standard deviation) and number (%) are shown for continuous and categorical variables, respectively. Also shown are *P* values for the differences between the 2 groups of children, calculated using a 2-sided Student's *t* test for continuous variables and a χ^2 test for categorical variables.

Table 5. Demographic characteristics of children exposed and not exposed to the acetaldehyde and carbon disulfide and polychlorinated biphenyls combination, which was associated with lifetime emergency room visit for asthma

	Multi–air toxic combination: acetaldehyde & carbon disulfide & polychlorinated biphenyls			
Variable	Total (n = 151)	Children exposed to combination (n = 48)	Children not exposed to combination (n = 103)	P value
Age, years	12.0 (3.2)	10.0 (2.6)	12.9 (3.0)	5.34×10^{-8}
Sex, female	89 (58.9)	28 (58.3)	61 (59.2)	0.94
Race/ethnicity				0.0046
Asian	6 (4.0)	3 (6.2)	3 (2.9)	
Black	26 (17.2)	14 (29.2)	12 (11.7)	
White	57 (37.7)	9 (18.8)	48 (46.6)	
Latino	51 (33.8)	19 (39.6)	32 (31.1)	
Mixed	10 (6.6)	2 (4.2)	8 (7.8)	
Not reported	1 (0.7)	1 (2.1)	0 (0.0)	
Income, US$	49,349.3 (18,725.8)	44,708.3 (14,015.1)	51,512.0 (20,194.5)	0.019

Mean (standard deviation) and number (%) are shown for continuous and categorical variables, respectively. Also shown are P values for the differences between the 2 groups of children, calculated using a 2-sided Student's t test for continuous variables and a χ^2 test for categorical variables.

leaf nodes. We calculated the frequency of each combination as the number of XGBoost models (out of 100) where it was included in at least 1 of the constituent trees and set a threshold of at least 10 XGBoost models to identify the most relevant combinations. Note that, if 2 or more variables are highly correlated, and thus similarly associated with the outcome, a key characteristic of the decision trees in the XGBoost model is that they will include only 1 of these variables as an internal decision node. Thus, unlike traditional regression models, XGBoost is not as adversely affected by collinearity among the input variables. Furthermore, DEEP executes XGBoost 100 times on randomly selected training sets, and different selections of these variables may be included in the different trees inferred, thus enhancing the coverage of the air toxic profiles.

In the second stage of DEEP, a multivariable linear regression model is built to assess the association of a candidate combination with the target outcome, adjusted for covariates. The asthma outcome is the dependent variable in this model, while the air toxic combina-

tion and covariates are its independent variables. The variable representing the combination takes a value of 1 for individuals exposed to it, determined using the threshold values of the constituent air toxics, and 0 otherwise. One model is built for each outcome and candidate combination, yielding the OR denoting the strength of the association between the two. The P values of all the associations are converted into FDRs after correcting for multiple-hypothesis testing using the Benjamini-Hochberg method (36). In this study, significant associations were identified as those with FDR ≤ 0.05.

Assessment of synergistic interactions in air toxic combinations. To assess potential synergy between members of air toxic combinations associated with asthma outcomes, we conducted statistical tests for interactions. Interactions between pairs of air toxics were assessed through additional multivariable regression models where the outcome was the dependent variable and predictors included the levels of the 2 toxics, their product as a representative of their interaction, and

Table 6. Demographic characteristics of children exposed and not exposed to the hydroquinone and ethylidene dichloride combination, which was associated with lifetime overnight hospitalization for asthma

	Multi–air toxic combination: hydroquinone & ethylidene dichloride			
Variable	Total (n = 151)	Children exposed to combination (n = 17)	Children not exposed to combination (n = 134)	P value
Age, years	12.0 (3.2)	9.9 (2.5)	12.2 (3.2)	0.0022
Sex, female	89 (58.9)	12 (70.6)	77 (57.5)	0.44
Race and ethnicity				0.32
Asian	6 (4.0)	0 (0.0)	6 (4.5)	
Black	26 (17.2)	2 (11.8)	24 (17.9)	
White	57 (37.7)	4 (23.5)	53 (39.6)	
Latino	51 (33.8)	10 (58.8)	41 (30.6)	
Mixed	10 (6.6)	1 (5.9)	9 (6.7)	
Not reported	1 (0.7)	0 (0.0)	1 (0.7)	
Income, US$	49,349.3 (18,725.8)	40,226.8 (6929.4)	50,506.6 (19,420.4)	8.26×10^{-5}

Mean (standard deviation) and number (%) are shown for the continuous and categorical variables, respectively. Also shown are P values for the differences between the 2 groups of children, calculated using a 2-sided Student's t test for continuous variables and a χ^2 test for categorical variables.

covariates. For combinations with 2 air toxics, this regression model was inferred from the whole cohort, while for combinations with 3 air toxics, analyses were conducted for the last 2 toxics on the sample meeting the threshold for the first toxic level in the combination. A significant interaction was identified if the P value of the interaction term in the model was lower than 0.05.

The DEEP framework is implemented in the Python programming language (71). The XGBoost, model evaluation (AUC score calculation), and regression components are implemented using the xgboost (33), scikit-learn (72), and statsmodels (73) Python packages, respectively.

Statistics. Statistical comparisons of subjects stratified by outcome or exposure were performed using 2-sided Student's t tests for continuous variables and χ^2 tests for categorical variables. Statistics employed within DEEP and for air toxic interaction analyses were performed as described above.

Study approval. The study was approved by the Mount Sinai IRB. Parents of participants provided written informed consent.

Author contributions

HHLH, PHC, GP, and SB conceived the study. GP and SB supervised the work. YCL, HHLH, PHC, ZA, and YC managed and analyzed the data. YCL and HHLH drafted the manuscript. YCL, HHLH, YC, PHC, ZA, LC, GP, and SB reviewed, edited, and approved the manuscript. Order among co-first authors was determined based on contribution to results generation.

Acknowledgments

Address correspondence to: Supinda Bunyavanich or Gaurav Pandey, One Gustave Levy Place, Box #1498, New York, New York 10029, USA. Phone: 1.212.241.5548; Email: supinda@post.harvard.edu (SB). Phone: 1.212.659.8535; Email: gaurav.pandey@mssm.edu (GP).

1. US EPA. About Urban Air Toxics. https://www.epa.gov/urban-air-toxics/about-urban-air-toxics. Accessed September 27, 2021.
2. Walters GI, et al. Occupational asthma caused by acrylic compounds from SHIELD surveillance (1989-2014). *Occup Med (Lond)*. 2017;67(4):282–289.
3. North ML, et al. Effects of phthalates on the development and expression of allergic disease and asthma. *Ann Allergy Asthma Immunol*. 2014;112(6):496–502.
4. Dumas O, et al. Respiratory effects of trichloroethylene. *Respir Med*. 2018;134:47–53.
5. Schenker MB, Jacobs JA. Respiratory effects of organic solvent exposure. *Tuber Lung Dis*. 1996;77(1):4–18.
6. Nurmatov UB, et al. Volatile organic compounds and risk of asthma and allergy: a systematic review. *Eur Respir Rev*. 2015;24(135):92–101.
7. Rumchev K, et al. Association of domestic exposure to volatile organic compounds with asthma in young children. *Thorax*. 2004;59(9):746–751.
8. Weisel CP. Assessing exposure to air toxics relative to asthma. *Environ Health Perspect*. 2002;110(suppl 4):527–537.
9. Ye D, et al. Estimating acute cardiorespiratory effects of ambient volatile organic compounds. *Epidemiology*. 2017;28(2):197–206.
10. Rosa MJ, et al. Prenatal exposure to polycyclic aromatic hydrocarbons, environmental tobacco smoke and asthma. *Respir Med*. 2011;105(6):869–876.
11. Ran J, et al. Source-specific volatile organic compounds and emergency hospital admissions for cardiorespiratory diseases. *Int J Environ Res Public Health*. 2020;17(17):E6210.
12. Delfino RJ. Epidemiologic evidence for asthma and exposure to air toxics: linkages between occupational, indoor, and community air pollution research. *Environ Health Perspect*. 2002;110(suppl 4):573–589.
13. Bably M, et al. Prenatal use of cleaning and scented products and its association with childhood asthma, asthma symptoms, and mental health and developmental comorbidities. *J Asthma*. 2021;58(1):46–51.
14. Adgent MA, et al. A combined cohort analysis of prenatal exposure to phthalate mixtures and childhood asthma. *Environ Int*. 2020;143:105970.
15. Stingone JA, et al. Using machine learning to identify air pollution exposure profiles associated with early cognitive skills among US children. *Environ Pollut*. 2017;230:730–740.
16. Patel CJ. Analytic complexity and challenges in identifying mixtures of exposures associated with phenotypes in the exposome era. *Curr Epidemiol Rep*. 2017;4(1):22–30.
17. Choi H, et al. Common household chemicals and the allergy risks in pre-school age children. *PLoS One*. 2010;5(10):e13423.
18. Oakes M, et al. Evaluating the application of multipollutant exposure metrics in air pollution health studies. *Environ Int*. 2014;69:90–99.
19. Khreis H, et al. Exposure to traffic-related air pollution and risk of development of childhood asthma: a systematic review and meta-analysis. *Environ Int*. 2017;100:1–31.
20. Esposito S, et al. Possible molecular mechanisms linking air pollution and asthma in children. *BMC Pulm Med*. 2014;14:31.
21. Orellano P, et al. Effect of outdoor air pollution on asthma exacerbations in children and adults: systematic review and multilevel meta-analysis. *PLoS One*. 2017;12(3):e0174050.
22. Edwards SC, et al. Prenatal exposure to airborne polycyclic aromatic hydrocarbons and children's intelligence at 5 years of age in a prospective cohort study in Poland. *Environ Health Perspect*. 2010;118(9):1326–1331.
23. Gehring U, et al. Exposure to air pollution and development of asthma and rhinoconjunctivitis throughout childhood and adolescence: a population-based birth cohort study. *Lancet Respir Med*. 2015;3(12):933–942.
24. Chang HH, et al. Time-series analysis of satellite-derived fine particulate matter pollution and asthma morbidity in Jackson, MS. *Environ Monit Assess*. 2019;191(suppl 2):280.
25. Brokamp C, et al. Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. *Atmos Environ (1994)*. 2017;151:1–11.
26. Coker E, et al. Multi-pollutant exposure profiles associated with term low birth weight in Los Angeles County. *Environ Int*. 2016;91:1–13.
27. Deng H, et al. Understanding the importance of key risk factors in predicting chronic bronchitic symptoms using a machine learning approach. *BMC Med Res Methodol*. 2019;19(1):70.
28. Centers for Disease Control and Prevention. Measures to Identify and Track Racial Disparities in Childhood Asthma: Prevalence and Outcome Measures. https://www.cdc.gov/asthma/asthma_disparities/outcome_measures.htm#anchor_1532268838154. Updated July 14, 2016. Accessed September 27, 2021.
29. Gliklich RE, et al. Harmonized outcome measures for use in asthma patient registries and clinical practice. *J Allergy Clin Immunol*. 2019;144(3):671–681.
30. Covar RA, et al. The Childhood Asthma Management Program (CAMP): contributions to the understanding of therapy and the natural history of childhood asthma. *Curr Respir Care Rep*. 2012;1(4):243–250.
31. United States Environmental Protection Agency. National-Scale Air Toxics Assessment. https://www.epa.gov/national-air-toxics-assessment. Accessed September 27, 2021.
32. Do AN, et al. Network study of nasal transcriptome profiles reveals master regulator genes of asthma. *J Allergy Clin Immunol*. 2021;147(3):879–893.

33. Tianqi Chen CG. XGBoost: a scalable tree boosting system. Paper presented at: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Association for Computing Machinery; August 13–17, 2016; New York, New York, USA. https://doi.org/10.1145/2939672.2939785. Accessed September 27, 2021.

34. Nathan RA, et al. Development of the asthma control test: a survey for assessing asthma control. *J Allergy Clin Immunol*. 2004;113(1):59–65.

35. Lever J, et al. Classification evaluation. *Nat Methods*. 2016;13(8):603–604.

36. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;57(1):289–300.

37. Chin JY, et al. Levels and sources of volatile organic compounds in homes of children with asthma. *Indoor Air*. 2014;24(4):403–415.

38. Gahleitner F, et al. Metabolomics pilot study to identify volatile organic compound markers of childhood asthma in exhaled breath. *Bioanalysis*. 2013;5(18):2239–2247.

39. Delfino RJ, et al. Respiratory symptoms and peak expiratory flow in children with asthma in relation to volatile organic compounds in exhaled breath and ambient air. *J Expo Anal Environ Epidemiol*. 2003;13(5):348–363.

40. Ferm RJ. The chemistry of carbonyl sulfide. *Chem Rev*. 1957;57(4):621–640.

41. Moghtaderi M, et al. Heavy metal concentration in classroom dust samples and its relationship with childhood asthma: a study from Islamic Republic of Iran. *East Mediterr Health J*. 2020;26(5):594–601.

42. Carder M, et al. Occupational and work-related respiratory disease attributed to cleaning products. *Occup Environ Med*. 2019;76(8):530–536.

43. Hsieh CY, et al. Combined exposure to heavy metals in PM2.5 and pediatric asthma. *J Allergy Clin Immunol*. 2021;147(6):2171–2180.

44. Andersson M, et al. Early life swimming pool exposure and asthma onset in children - a case-control study. *Environ Health*. 2018;17(1):34.

45. New Jersey Department of Health. Hazardous Substance Fact Sheet: Triethylamine. https://nj.gov/health/eoh/rtkweb/documents/fs/1907.pdf. Updated February 2010. Accessed September 27, 2021.

46. Mukherjee M, et al. Enhanced expression of Organic Cation Transporters in bronchial epithelial cell layers following insults associated with asthma - Impact on salbutamol transport. *Eur J Pharm Sci*. 2017;106:62–70.

47. Laborde-Casterot H, et al. Is the incidence of aliphatic amine-induced occupational rhinitis and asthma underestimated? *Am J Ind Med*. 2014;57(12):1303–1310.

48. Ohara T, et al. Acrylic acid derivatives. In: Ley C, et al., eds. *Ullmann's Encyclopedia of Industrial Chemistry*. Wiley-VCH; 2020:1–21.

49. Peng CH, et al. Organocobalt mediated radical polymerization of acrylic acid in water. *Macromolecules*. 2007;40(19):6814–6819.

50. Wang X, et al. Additive manufacturing of ceramics from preceramic polymers: a versatile stereolithographic approach assisted by thiol-ene click chemistry. *Addit Manuf*. 2019;27:80–90.

51. Pozdeeva NN, Denisov ET. Mechanism of hydroquinone-inhibited oxidation of acrylic acid and methyl methacrylate. *Kinet Catal*. 2011;52:506.

52. Rossberg M, et al. Chlorinated Hydrocarbons. In: Ley C, et al., eds. *Ullmann's Encyclopedia of Industrial Chemistry*. Wiley-VCH; 2006. Accessed October 6, 2021. https://doi.org/10.1002/14356007.a06_233.pub2.

53. U.S. Department of Health and Human Services Agency for Toxic Substances and Disease Registry. Toxicological Profile for 1,1-Dichloroethane. https://www.atsdr.cdc.gov/toxprofiles/tp133.pdf. Updated April 2013. Accessed October 6, 2021.

54. Fiserova-Bergerova V, et al. Effects of biosolubility on pulmonary uptake and disposition of gases and vapors of lipophilic chemicals. *Drug Metab Rev*. 1984;15(5–6):1033–1070.

55. Holder JW. Physical and physicochemical factors effecting transport of chlorohydrocarbon gases from lung alveolar air to blood as measured by the causation of narcosis. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev*. 2012;30(1):42–80.

56. Becher R, et al. Environmental chemicals relevant for respiratory hypersensitivity: the indoor environment. *Toxicol Lett*. 1996;86(2-3):155–162.

57. Sivertsen B, Clench-Aas J. Exposure to environmental chemicals relevant for respiratory hypersensitivity: European aspects. *Toxicol Lett*. 1996;86(2-3):143–153.

58. Schwela D. Exposure to environmental chemicals relevant for respiratory hypersensitivity: global aspects. *Toxicol Lett*. 1996;86(2-3):131–142.

59. Triggiani M, et al. Modulation of mast cell and basophil functions by benzene metabolites. *Curr Pharm Des*. 2011;17(34):3830–3835.

60. Olumide YM, et al. Complications of chronic use of skin lightening cosmetics. *Int J Dermatol*. 2008;47(4):344–353.

61. World Health Organization. *Hydroquinone: Health and Safety Guide*. World Health Organization; 1996.

62. George B, et al. An evaluation of EPA's National-Scale Air Toxics Assessment (NATA): comparison with benzene measurements in Detroit, Michigan. *Atmos Environ*. 2011;45(19):3301–3308.

63. Xue ZJia C. A model-to-monitor evaluation of 2011 National-Scale Air Toxics Assessment (NATA). *Toxics*. 2019;7(1):E13.

64. Garcia E, et al. Evaluation of the agreement between modeled and monitored ambient hazardous air pollutants in California. *Int J Environ Health Res*. 2014;24(4):363–377.

65. National Asthma Education and Prevention Program. Expert Panel Report 3 (EPR-3): guidelines for the diagnosis and management of asthma-Summary Report 2007. *J Allergy Clin Immunol*. 2007;120(suppl 5):S94–S138.

66. Bettiol A, et al. The first 1000 days of life: traffic-related air pollution and development of wheezing and asthma in childhood. A systematic review of birth cohort studies. *Environ Health*. 2021;20(1):46.

67. Gehring U, et al. Air pollution and the development of asthma from birth until young adulthood. *Eur Respir J*. 2020;56(1):2000147.

68. Hsu HH, et al. Prenatal particulate air pollution and asthma onset in urban children. Identifying sensitive windows and sex differences. *Am J Respir Crit Care Med*. 2015;192(9):1052–1059.

69. Weitekamp Chelsea A, et al. An examination of national cancer risk based on monitored hazardous air pollutants. *Environ Health Perspect*. 2021;129(3):037008.

70. United States Census Bureau. United States Census 2016. https://www2.census.gov/programs-surveys/cbp/datasets/2016/. Updated April 19, 2018. Accessed September 27, 2021.

71. Oliphant TE. Python for scientific computing. *Comput Sci Eng*. 2007;9(3):10–20.

72. Pedregosa F, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–2830.

73. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. Paper presented at: 9th Python in Science Conference; June 28–July 3, 2010; Austin, Texas, USA. https://doi.org/10.25080/Majora-92bf1922-011. Accessed September 27, 2021.