

Clustering asthma symptoms and cleaning and disinfecting activities and evaluating their associations among healthcare workers

Feng-Chiao Su^a, Melissa C. Friesen^b, Michael Humann^a, Aleksandr B. Stefaniak^a, Marcia L. Stanton^a, Xiaoming Liang^a, Ryan F. LeBouf^a, Paul K. Henneberger^a, M. Abbas Virji^{a,*}

^a Respiratory Health Division, National Institute for Occupational Safety and Health (NIOSH), Centers for Disease Control and Prevention (CDC), Morgantown, WV, USA

^b Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA

ARTICLE INFO

Keywords:

Asthma symptoms
Cleaning and disinfecting activities
Hierarchical cluster analysis
Healthcare workers

ABSTRACT

Asthma is a heterogeneous disease with varying severity and subtypes. Recent reviews of epidemiologic studies have identified cleaning and disinfecting activities (CDAs) as important risk factors for asthma-related outcomes among healthcare workers. However, the complexity of CDAs in healthcare settings has rarely been examined. This study utilized a complex survey dataset and data reduction approaches to identify and group healthcare workers with similar patterns of asthma symptoms, and then explored their associations with groups of participants with similar patterns of CDAs. Self-reported information on asthma symptoms/care, CDAs, demographics, smoking status, allergic status, and other characteristics were collected from 2030 healthcare workers within nine selected occupations in New York City. Hierarchical clustering was conducted to systematically group participants based on similarity of patterns of the 27 asthma symptom/care variables, and 14 product applications during CDAs, separately. Word clouds were used to visualize the complex information on the resulting clusters. The associations of asthma health clusters (HCs) with exposure clusters (ECs) were evaluated using multinomial logistic regression. Five HCs were identified (HC-1 to HC-5), labelled based on predominant features as: “no symptoms”, “winter cough/phlegm”, “mild asthma symptoms”, “undiagnosed/untreated asthma”, and “asthma attacks/exacerbations”. For CDAs, five ECs were identified (EC-1 to EC-5), labelled as: “no products”, “housekeeping/chlorine”, “patient care”, “general cleaning/laboratory”, and “disinfection products”. Using HC-1 and EC-1 as the reference groups, EC-2 was associated with HC-4 (odds ratio (OR) = 3.11, 95% confidence interval (95% CI) = 1.46–6.63) and HC-5 (OR = 2.71, 95% CI = 1.25–5.86). EC-3 was associated with HC-5 (OR = 2.34, 95% CI = 1.16–4.72). EC-4 was associated with HC-5 (OR = 2.35, 95% CI = 1.07–5.13). EC-5 was associated with HC-3 (OR = 1.81, 95% CI = 1.09–2.99) and HC-4 (OR = 3.42, 95% CI = 1.24–9.39). Various combinations of product applications like using alcohols, bleach, high-level disinfectants, and enzymes to disinfect instruments and clean surfaces captured by the ECs were identified as risk factors for the different asthma symptoms clusters, indicating that prevention efforts may require targeting multiple products. The associations of HCs with EC can be used to better inform prevention strategies and treatment options to avoid disease progression. This study demonstrated hierarchical clustering and word clouds were useful techniques for analyzing and visualizing a complex dataset with a large number of potentially correlated variables to generate practical information that can inform prevention activities.

1. Introduction

Work-related asthma (WRA) subsumes occupational asthma and work-exacerbated asthma, and is a common but preventable respiratory disease occurring in the workplace (Friedman-Jimenez et al., 2015; Tarlo et al., 2008). In 2016, WRA affected as many as 2.7 million U.S. workers, especially healthcare workers who had among the highest prevalence of current and lifetime asthma (10.7–12.4%) (Dodd and

Mazurek, 2016; Mazurek and Weissman, 2016; Wiszniewska and Walusiak-Skorupa, 2014). Asthma is a heterogeneous disease characterized by chronic airway inflammation, with variable airflow limitation and symptoms such as wheeze, chest tightness, frequent cough, and shortness of breath (SOB) (Global Initiative for Asthma, 2018).

Asthma comprises multiple subtypes characterized by distinct underlying pathophysiological mechanisms sharing common observable characteristics (phenotypes) (Bradding and Green, 2010). The type of

* Corresponding author. ScD, CIH, 1095 Willowdale Road, Morgantown, WV, 26505, USA.

E-mail address: MVirji@cdc.gov (M.A. Virji).

<https://doi.org/10.1016/j.ijheh.2019.04.001>

Received 7 December 2018; Received in revised form 28 March 2019; Accepted 11 April 2019

1438-4639/© 2019 Elsevier GmbH. All rights reserved.

data used for asthma phenotypical classification schemes can include respiratory symptoms, sensitizer vs. irritant causative agents, and eosinophilic vs. neutrophilic responses to causative agents (Bradding and Green, 2010). In addition, some schemes are based on data-driven approaches such as principal components analysis, clustering, and machine learning applied to health-related information (Deliu et al., 2016; Howard et al., 2015). In occupational epidemiologic studies, asthma is often defined as a dichotomous outcome based on a series of questions on symptoms, asthma medication use and physician diagnosis (Pekkanen et al., 2005). However, given the heterogeneity of asthma, asthma likely occurs on a continuous spectrum with varying degrees of severity, phenotypes, and subtypes (Deliu et al., 2016; Pekkanen et al., 2005). Approaches to better characterize this heterogeneity in asthma include using grouping methods, for example hierarchical clustering to identify asthma subtypes or combining multiple symptoms to generate an asthma severity score (Deliu et al., 2016; Sunyer et al., 2007). Numerous studies have identified asthma subtypes via grouping asthma symptoms, physiologic function and biomonitoring results, but occupational epidemiologic studies have generally not evaluated the associations between these asthma subtypes and workplace exposures.

Several recent reviews of epidemiologic studies have identified cleaning and disinfecting activities (CDAs) as important risk factors for asthma-related outcomes among healthcare workers (Folletti et al., 2017; Mazurek and Weissman, 2016; Wiszniewska and Walusiak-Skorupa, 2014; Zock et al., 2010). However, the complexity of CDAs and the exposures they generate in healthcare settings has rarely been examined. A recent population-based study in France used clustering methods to characterize the pattern of cleaning product use in domestic settings (Marbac et al., 2018). However, CDAs and the resulting exposures are typically evaluated in epidemiologic models as single exposure variables, with the exception of one study (Arif et al., 2009) that evaluated the effects of grouped cleaning and disinfecting products on WRA. To better represent the reality of mixed exposures in the workplace, the current study utilizes a complex survey dataset and systematic data reduction approaches to characterize the patterns of CDAs among healthcare workers. The objective of the study was to identify clusters of healthcare workers with similar asthma symptom patterns that may represent different underlying asthma subtypes or severity, and to explore their associations with groups of participants with similar patterns of CDAs and exposures. Understanding the associations of these clusters with workplace exposures can better inform prevention strategies and treatment options.

2. Materials and methods

2.1. Study population and data collection

This study was reviewed and approved by the NIOSH Institutional Review Board and informed consent was obtained before participation. During 2014, healthcare workers who were members of the Service Employees International Union Local 1199 who worked in hospitals or nursing homes in New York City received a letter of invitation, information about the study and consent form, followed by telephone calls to administer the interviewer-led questionnaire and several subsequent reminders. A total of 2030 participants from nine selected occupations completed a telephone or web-based questionnaire, and represented 13.3% of the effective sample size (i.e., invitees known or presumed to be eligible), and 22.5% of the 9009 with whom we had at least some telephone contact; a majority of those who did not complete the questionnaire were never reached by telephone. The nine selected occupations were central supply workers, dental assistants,

environmental service workers or housekeepers, medical or clinical laboratory technicians, licensed practical nurses, nursing assistants, operating room technicians, registered nurses, and respiratory therapists or technicians. Inverse probability weights were calculated to address any potential selection or non-participation bias using models that addressed differences between participants vs non-participants and the sample frame, in terms of the distribution of age, gender, occupation, smoking status, and wheeze in the last 12 months, as described in detail by Caridi et al. (2019). The survey questionnaire was developed based on standardized instruments (Burney et al., 1994; Ferris, 1978; Jarvis, 2002; Sunyer et al., 2007) and a previous healthcare study (Delclos et al., 2006, 2007). The survey was designed to collect data on demographics (e.g., age, gender, race, education), smoking status (i.e., never, former, or current smoker), physician-diagnosed and last 12-months asthma symptoms (e.g., wheeze, SOB, coughing, phlegm or mucous production, asthma attack, treatment and medication use for asthma), CDAs (e.g., frequency and duration of tasks and product-use, tools, controls), medical history (e.g., allergic status), last 5-year employment history (e.g., industry, job title), and home environment (e.g., water damage, mold growth, renovations, cleaning activities). The survey questionnaire was presented in English and Spanish by trained interviewers on the telephone and in English online.

2.2. Data analyses

2.2.1. Respiratory health variables

The main health outcomes in the current study were 27 variables for asthma symptoms/care occurring in the last 12-months (Table 1). Majority of asthma symptoms/care variables were indicator variables (present/absent), except for number of asthma attacks, number of missed workdays because of asthma, and number of urgent treatments for asthma, which had three categories. Four composite health outcomes were created using variables collected in the survey questionnaire, including asthma score, bronchial hyper-responsiveness (BHR)-related symptoms, allergic status, and physician-diagnosed asthma. An asthma score was calculated as the sum of the positive responses to the following questions on symptoms in the last 12 months (yes = 1, no = 0; scores ranged from 0 to 5): breathless when wheezing, woken up with tightness in chest, daytime SOB at rest, SOB following strenuous activity, and woken by SOB. The asthma score was developed and validated by the European Community Respiratory Health Survey researchers as a measure of asthma severity (Sunyer et al., 2007). BHR-related symptoms were assessed via an algorithm that yielded a dichotomous outcome based on eight symptoms: trouble with breathing; attack or episode of SOB in the last 12 months; wheezing or whistling in the chest in the last 12 months; woken by an attack of cough in the last 12 months; woken by an attack of chest tightness in the last 12 months; itchy or watery eyes when near animals, feathers, or in a dusty part of the house; feeling of chest tightness when near animals, feathers, or in a dusty part of the house; and itchy or watery eyes when near flowers or pollen (Delclos et al., 2006, 2007). Allergic status was defined as a positive response to any of the following criteria: ever had nasal or sinus allergies (e.g., hay fever), ever had eczema or any kind of skin allergy, ever had allergies to animals, ever had allergies to dust or dust mites, or ever had allergies to latex. Physician-diagnosed asthma was defined as a positive response to all of the following criteria: ever had asthma, ever had an episode of asthma symptoms, and asthma confirmed by a physician (Caridi et al., 2019). Information on other respiratory diseases, such as previous diagnoses of chronic obstructive pulmonary disease (COPD) and emphysema, was also drawn from participants' self-reported medical histories.

Table 1

Fractions (shown in %) of asthma symptoms/care, composite health outcomes, and other variables overall and by resulting asthma health clusters.

Variable	Overall (n = 2030)	HC-1: No symptoms (n = 885)	HC-2: Winter cough/ phlegm (n = 640)	HC-3: Mild asthma symptoms (n = 357)	HC-4: Undiagnosed/untreated asthma (n = 63)	HC-5: Asthma attacks/ exacerbations (n = 85)
<u>Last 12-mo Asthma Symptoms/Care (included in clustering)</u>						
Wheezing or whistling (any)	14.3	0	6.88	30.5	100	88.2
Breathless when wheezing	7.04	0	2.03	7.56	87.3	56.5
Wheezing or whistling without a cold	8.67	0	1.56	17.4	73.0	68.2
Woken up with tightness in chest (any)	10.0	0	7.19	18.8	61.9	61.2
SOB (any)	14.5	0	1.09	44.8	95.2	78.8
SOB during the day at rest	6.90	0	0.47	17.4	58.7	44.7
SOB following strenuous activity	17.8	0	3.13	66.7	81.0	62.4
Woken by SOB	6.50	0	3.75	12.0	47.6	41.2
Frequent coughing during the day or at night in the winter	39.6	0	76.7	49.3	95.2	90.6
Coughing on most days	6.31	0	8.59	6.72	36.5	30.6
Woken by coughing	16.8	0	28.9	17.6	79.4	51.8
Usually bringing up phlegm/mucous during the day or at night in the winter	28.1	0	51.7	36.4	77.8	71.8
Bringing up phlegm/mucous on most days	5.76	0	9.53	5.04	25.4	25.9
Asthma attack (any)	3.89	0	0	0.84	0	89.4
Number of asthma attacks						
1–5 attacks	2.86	0	0	0.56	0	65.9
More than 5 attacks	0.94	0	0	0	0	22.4
Taking medications for asthma (any)	6.40	0	0.16	9.24	27.0	92.9
Using fast-acting or rescue bronchodilators for asthma	5.47	0	0	5.88	27.0	85.9
Increasing usage of fast-acting or rescue bronchodilators over a period	2.76	0	0	0.28	3.17	62.4
Using inhaled steroids for asthma	3.79	0	0	3.64	7.94	69.4
Increasing usage of inhaled steroids over a period	1.72	0	0	0.56	0	38.8
Using oral steroids for asthma	2.32	0	0	0.56	3.17	50.6
Increasing usage of oral steroids over a period	1.58	0	0	0.56	1.59	34.1
Missing workdays due to asthma (any)	3.15	0.11	0.16	0.84	1.59	68.2
Number of missed workdays due to asthma						
Missing 1–5 workdays	1.77	0	0	0.84	1.59	37.6
Missing more than 5 workdays	1.23	0	0	0	0	29.4
Overnight hospitalization for asthma (any)	0.34	0	0	0	0	8.24
Urgent treatment for asthma (any)	2.27	0	0	0	1.59	52.9
Number of urgent treatments						
1–2 times urgent treatments	1.58	0	0	0	1.59	36.5
More than 3 times urgent treatments	0.69	0	0	0	0	16.5
<u>Composite Health Outcomes (not included in clustering)</u>						
Asthma score (mean \pm SD)	0.48 \pm 1.02	0	0.17 \pm 0.43	1.22 \pm 0.95	3.37 \pm 1.14	2.66 \pm 1.55
Asthma score = 1	13.6	0	12.0	51.3	3.17	17.6
Asthma score = 2	5.67	0	2.03	18.2	22.2	27.1
Asthma score = 3	3.10	0	0.16	7.84	30.2	17.6
Asthma score = 4	1.72	0	0	2.80	23.8	11.8
Asthma score = 5	1.43	0	0	0	20.6	18.8
Bronchial hyper-responsiveness-related symptoms	25.7	4.41	31.7	45.7	79.4	78.8
Allergic status	50.9	33.6	55.9	69.7	84.1	89.4
Physician-diagnosed asthma	11.0	3.28	4.06	18.8	31.7	95.3
Allergic status X physician-diagnosed asthma						
No allergy and no physician-diagnosed asthma	46.9	64.7	42.8	26.9	12.7	1.18
Allergy but no physician-diagnosed asthma	42.1	32.0	53.1	54.3	55.6	3.53
No allergy but physician-diagnosed asthma	2.22	1.69	1.25	3.36	3.17	9.41
Allergy and physician-diagnosed asthma	8.77	1.58	2.81	15.4	28.6	85.9
<u>Other Variables (not included in clustering)</u>						
Age in year (mean \pm SD)	48.6 \pm 11.4	48.9 \pm 11.6	47.8 \pm 11.6	48.2 \pm 11.1	49.4 \pm 9.27	51.6 \pm 9.12
Female	76.0	72.8	74.7	80.4	90.5	89.4
Smoking						
Former	11.1	9.49	10.8	13.4	19.0	15.3
Current	5.57	4.63	5.78	6.72	12.7	3.53

(continued on next page)

Table 1 (continued)

Variable	Overall (n = 2030)	HC-1: No symptoms (n = 885)	HC-2: Winter cough/ phlegm (n = 640)	HC-3: Mild asthma symptoms (n = 357)	HC-4: Undiagnosed/untreated asthma (n = 63)	HC-5: Asthma attacks/ exacerbations (n = 85)
Chronic obstructive pulmonary disease	1.33	0.79	0.47	1.12	6.35	10.6
Emphysema	0.64	0.23	0.31	0.84	4.76	3.53

HC, asthma health cluster; SOB, shortness of breath; SD, standard deviation.

2.2.2. Cleaning and disinfecting activity variables

Information on CDAs in the last 12 months was collected using six modules in the survey questionnaire: 1. Tasks and products used in sterilizing medical instruments; 2. Tasks and products used during cleaning fixed surfaces, equipment or instruments; 3. Combined task/product use while working in a medical or clinical laboratory; 4. Tasks and products used on patients such as chemicals, adhesives, antiseptics, alcohols or solvents; 5. Combined task/product use while administering aerosolized medications; and 6. Combined task/product use while working as a dental assistant; all modules included information on the frequency and duration of tasks and product use. These six modules had more than 250 variables, and the current study mainly used indicator variables (0 versus 1) and ordinal frequency scores (0: no use, 1/low: use 1–3 days per week, and 2/high: use 4–7 days per week) of tasks and product-use for statistical analyses. In addition, we grouped the raw indicator variables into 14 cleaning and disinfecting product application variables which were used in exposure clustering including: high-level disinfectants, alcohols, chlorine bleach, enzymes, formaldehyde, detergents, floor wax stripper, glass cleaners, phenolics, quaternary ammonium compounds (quats), skin-wipes used on patients, clinical or medical laboratory products, aerosolized medications, and dental products (Table 2). These 14 cleaning and disinfecting product application variables were combinations of tasks and product-use (e.g., variable chlorine bleach was defined as: use of chlorine bleach to clean and disinfect fixed-surface, equipment or instruments). This grouping approach enabled the assignment of total VOC (TVOC) exposures to participants who performed these 14 product application tasks based on our previous exposure study described below (Su et al., 2018).

2.2.3. Frequency-weighted total volatile organic compound (TVOC) exposures

In a previous study, we modeled the associations between TVOC exposures and the 14 CDAs using personal exposure measurements and observations of CDAs collected from 100 healthcare workers at four hospitals that were not part of the present study (Su et al., 2018). Two measures of the TVOC exposures were used as the outcome variable in the predictive models which included indicator variables (1/0) of the 14 CDAs as predictors: 1) the sum of 14 VOCs (ethanol, acetone, 2-propanol, methylene chloride, hexane, chloroform, benzene, methylmethacrylate, toluene, ethylbenzene, m,p-xylene, o-xylene, α -pinene, and d-limonene), and 2) a subset sum of 11 VOCs (which excluded the three most dominant VOCs: ethanol, acetone and 2-propanol). A task-exposure matrix (TEM) was created that included parameter estimates (β) for TVOC14 and TVOC11 for each of the 14 CDAs. The parameter estimates for the 14 CDAs were assigned to participants in the epidemiologic study if they reported performing a particular CDA in their questionnaire responses. Frequency-weighted TVOC14 and TVOC11 exposures were estimated for each participant using weekly frequency fractions reported in the questionnaire (0/7, 2/7, and 5.5/7 day/week), and the parameter estimates obtained from the predictive models. The

equation used to calculate the two TVOC exposures is shown below:

$$\text{TVOC14 or TVOC11 Exposure} = e^{(\text{Intercept} + (\beta_1 \times \text{PA}_1 \times F_1) + \dots + (\beta_n \times \text{PA}_n \times F_n))}$$

where $\beta_1 \dots \beta_n$ are parameter estimates (i.e., coefficients) for 14 CDAs obtained in the linear mixed effect models using log-transformed TVOC14 or TVOC11 measurements, $\text{PA}_1 \dots \text{PA}_n$ are 14 CDAs, and $F_1 \dots F_n$ are weekly frequency fractions for each product application.

2.2.4. Hierarchical clustering

Hierarchical clustering with Ward's minimum variance method (Ward and Joe, 1963) was applied to systematically group participants with similar patterns of the 27 respiratory health variables and frequency scores of 14 product applications (i.e., input variables), separately. Input variables were not normalized or scaled because all product application variables were on the same scale and had similar ranges, while the asthma health variable were all indicator variables with values of 0/1, with the exception of three indicator variables with values 0, 1, or 2. The number of clusters were determined using scree plots and dendrograms. For resulting asthma health clusters (HCs), the proportion of workers with physician-diagnosed asthma, positive allergic status, BHR-related symptoms, COPD, emphysema, their demographics and smoking status and average asthma scores were calculated. For exposure clusters (ECs), distributions of frequency-weighted TVOC exposures were also calculated.

Word clouds were used to visualize the complex information of resulting clusters. For participants in each cluster, a text file was created using the positive responses to the input variables. For example, if Participant 1 in HC-1 had a positive response to Question 1 “wheezing or whistling”, then the data cell of Question 1 (column) and Participant 1 (row) would be filled as “wheezing or whistling” (i.e., positive responses would be replaced by keywords from their corresponding questions) in the text file of HC-1. A list of keywords and their frequencies for each cluster was generated, and then a word cloud was created based on the list. The size of words in the cloud depended on their frequencies (i.e., larger words indicating higher frequencies). We labelled the word clouds using the most frequently highlighted keywords to represent the underlying characteristic of the clusters (HCs), which include complex combination/pattern of variables. In addition to 14 input variables, the frequencies of job and facility were also included in the text files of ECs, and presented in word clouds.

2.2.5. Multinomial logistic regression

The associations of HCs with exposure variables including individual product applications, and ECs (one variable with multiple categories) were separately evaluated using multinomial logistic regression that models nominal dependent variables (Engel, 1988). All models were adjusted for age, gender, education, smoking status and allergic status, and incorporated sampling weights to adjust for potential selection and non-participation bias (Caridi et al., 2019). The goodness-of-fit of the model was evaluated using max-rescaled R^2

Table 2

Fractions (shown in %) of product applications, estimated VOC exposures and occupational characteristics by resulting exposure clusters.

Variable	Group	Overall (n = 2030)	EC-1: No products (n = 630)	EC-2: House-keeping/ Chlorine (n = 382)	EC-3: Patient care (n = 500)	EC-4: General cleaning/ laboratory (n = 368)	EC-5: Disinfection products (n = 150)
Product Applications (included in clustering)							
High-level disinfectants	Low: 1–3 days/week	3.15	0	1.83	4.60	3.80	13.3
	High: 4–7 days/week	4.38	0	0.26	0.60	0.54	55.3
Alcohols	Low: 1–3 days/week	15.1	0	11.0	28.2	23.9	23.3
	High: 4–7 days/week	26.2	0	20.9	22.4	67.9	60.0
Chlorine bleach	Low: 1–3 days/week	15.5	3.17	19.4	22.2	20.7	22.7
	High: 4–7 days/week	22.2	0	59.9	4.40	36.4	44.0
Enzymes	Low: 1–3 days/week	3.89	0	4.45	2.00	1.09	32.0
	High: 4–7 days/week	5.22	0	8.38	2.00	0.82	40.7
Formaldehyde	Low: 1–3 days/week	1.03	0	0.52	0.20	0.82	10.0
	High: 4–7 days/week	0.79	0	0	0	0.82	8.67
Detergents	Low: 1–3 days/week	4.33	0	12.6	2.80	2.99	10.0
	High: 4–7 days/week	7.64	0	33.2	0.40	0.54	16.0
Floor wax stripper	Low: 1–3 days/week	4.68	0	19.6	0.60	1.36	8.00
	High: 4–7 days/week	0.94	0	3.40	0	0	4.00
Glass cleaners	Low: 1–3 days/week	6.40	0	26.7	0.60	4.35	6.00
	High: 4–7 days/week	7.29	0	35.1	0	1.09	6.67
Phenolics	Low: 1–3 days/week	2.12	0	3.93	2.00	1.63	8.00
	High: 4–7 days/week	2.76	0	7.33	1.40	1.36	10.7
Quaternary ammonium compounds	Low: 1–3 days/week	0.49	0.32	1.05	0.20	0.54	0.67
	High: 4–7 days/week	1.58	0	5.50	0.80	0.82	2.67
Skin-wipes used on patients	Low: 1–3 days/week	13.5	0	4.19	41.6	5.98	18.7
	High: 4–7 days/week	14.4	0	1.83	41.0	12.8	22.0
Clinical or medical laboratory products	Low: 1–3 days/week	0.94	0	0.79	0	3.26	2.67
	High: 4–7 days/week	1.77	0	0.52	0	8.97	0.67
Aerosolized medications	Low: 1–3 days/week	9.75	0	1.83	31.2	5.43	10.0
	High: 4–7 days/week	8.28	0	1.57	25.6	7.07	5.33
Dental products	Low: 1–3 days/week	0.84	0	0.52	0.40	1.90	4.00
	High: 4–7 days/week	0.69	0	0.26	0	2.17	3.33
Estimated Exposures^a and Occupational Characteristics (not included in clustering)							
Total 14 VOCs (score/1000)	Mean	3.00	1.23	1.98	3.12	3.76	10.3
	Standard deviation	6.54	0.01	1.18	2.09	3.38	21.6
Total 11 VOCs (score)	95th percentile	6.65	1.22	4.55	8.39	6.89	55.3
	Mean	29.2	9.07	9.51	13.5	21.5	235
Job	Standard deviation	213	0.27	4.25	6.14	102	740
	95th percentile	26.6	9.12	17.1	23.4	21.1	2145
Job	Central supply worker	2.02	0.95	1.31	0	0.82	18.0
	Dental assistant	1.58	0.63	0.79	0	3.80	7.33
	Environmental service worker, or housekeeper	18.4	13.8	65.7	0.20	5.43	10.0
	Lab technician, technologist, or assistant	8.18	5.40	8.12	1.20	23.9	4.67
	Licensed practical or licensed vocational nurse	14.6	4.76	3.66	36.2	13.9	14.0
	Nursing assistant	34.6	66.5	13.4	14.4	35.6	19.3
	Operating room technician	3.10	3.33	1.31	2.80	0.82	13.3
	Registered nurse	13.8	3.97	4.97	35.0	12.8	9.33
	Respiratory therapist or respiratory technician	3.69	0.63	0.79	10.2	2.99	4.00
	Facility						
Facility	Hospital	50.6	31.3	64.4	53.0	57.3	72.7
	Nursing home	38.7	59.7	28.3	32.2	31.0	17.3
	Hospital with nursing home	6.70	6.67	3.14	9.40	6.52	7.33
	Other	3.94	2.22	4.19	5.40	5.16	2.67

EC, exposure cluster; VOC, volatile organic compounds.

^a Estimated exposures to total VOCs were obtained using parameters for product applications from predictive models developed in our previous study, and weighted by frequencies of product applications in this study; total 14 VOCs included ethanol, acetone, 2-propanol, methylene chloride, hexane, chloroform, benzene, methyl-methacrylate, toluene, ethylbenzene, m,p-xylene, o-xylene, α -pinene, and d-limonene, and total 11 VOCs excluded the three most dominant compounds: ethanol, acetone and 2-propanol.

(Nagelkerke, 1991; Shtatland et al., 2002). The HC and EC with the least positive responses to symptoms and CDAs, respectively, were assigned as reference groups. The individual product applications were indicated variables (yes versus no).

Hierarchical clustering was conducted using JMP version 12

software (SAS Institute Inc., Cary, NC). Multinomial logistic regression and descriptive analyses were conducted using SAS version 9.4 software (SAS Institute Inc., Cary, NC). The word clouds were generated in R 3.3.1 using the “ggplot2”, “tm” and “wordcloud” packages (Feinerer and Hornik, 2017; Fellows, 2014; R Core Team, 2016; Wickham, 2009).

3. Results

3.1. Asthma health clusters (HCs)

The hierarchical clustering identified five HCs for the 12-month asthma symptoms/care. The distributions of asthma symptoms/care variables, composite health outcomes, and other variables among five identified HCs are shown in Table 1. The clusters were labelled based on their prevailing symptom(s). All participants in HC-1 ($n = 885$) had no asthma symptoms in the last 12 months (except one participant who had missed workdays due to asthma; this participant had physician-diagnosed asthma but no asthma attacks in the last 12 months; to avoid confusion, this observation was not shown in the word cloud). HC-1 also had a 0 for average asthma scores and across almost all variables used in clustering, and the smallest fractions of participants with BHR-related symptoms, allergies, and emphysema among all HCs. This cluster was labelled “no symptoms”. In HC-2 ($n = 640$), participants had a lower average asthma score than the overall average of 0.48, and also has the lowest fraction of participants with the classic asthma symptoms such as wheeze, SOB and chest tightness excluding HC-1. About one-third and more than half of participants had BHR-related symptoms and allergies, respectively, and a larger fraction of this cluster reported cough and phlegm symptoms compared to HC-3. Additionally, HC-2 had the smallest fraction of participants who had COPD as well as the lowest fraction of physician diagnosed asthma excluding HC-1. This cluster was labelled “winter cough/phlegm”. In HC-3 ($n = 357$), more than half of participants had average asthma scores ≥ 1 . About 45% and 70% of participants had BHR-related symptoms and allergies, respectively. HC-3 also had a higher fraction of participants with the classic asthma symptoms and physician diagnosed asthma compared to HC-2, but a smaller fraction for medication or urgent care use. This cluster was labelled “mild asthma symptoms”. HC-4 ($n = 63$) had the highest average asthma score, and the largest fractions for BHR-related symptoms, classic asthma symptoms of wheeze, SOB and chest tightness, and emphysema among all HCs. HC-4 also had the highest percentages of former/current smokers ($> 30\%$). With only 31% physician-diagnosed asthma, and smaller fraction of participants seeking medical attention compared to HC-5, this cluster likely represents “undiagnosed or untreated asthma” and was labelled as such. HC-5 ($n = 85$) had the second highest average asthma score, but the largest fractions for allergies, physician-diagnosed asthma, COPD, and exacerbations as indicated by the need for a variety of asthma treatments. This cluster was labelled “asthma attacks/exacerbations”. Note, unlike conventional approaches which may use a specific set of variables to define severity, in this data-driven approach, symptoms may be present in all clusters but in different proportions and in different combinations with other symptoms. Thus HCs are labelled (given meaning) based on prevailing symptoms, and cannot be defined by any set of specific variables as in the conventional approach.

3.2. Exposure clusters (ECs)

The distributions of product application variables, frequency-weighted TVOC exposures, job, and facility among five identified ECs are shown in Table 2. In EC-1 ($n = 630$), 66.5% of the participants were nursing assistants, and 59.7% of the participants worked in nursing homes. The fraction of participants in EC-1 reporting positive responses to 14 product application variables was 0 for all except 3.17% for low chlorine bleach and 0.32% for low quats. This cluster was labelled “no products”. Similar to HC results, while there were many overlapped product applications among ECs, EC-2 to EC-5 represented different levels and types of exposures and occupations. We named EC-2 to EC-5 based on the most common product applications in each EC as follows:

“housekeeping/chlorine” in EC-2, “patient care” in EC-3, “general cleaning/laboratory” in EC-4, and “disinfection products” in EC-5.

In EC-2 ($n = 382$), the most common product applications were high chlorine bleach (59.9%), high glass cleaners (35.1%), and high detergents (33.2%). More than 65% of the participants were environmental service workers or housekeepers. While EC-2 had the lowest means and medians of frequency-weighted scores for both TVOV14 and TVOC11 among all ECs (except EC-1), it had the highest fraction of participants using chlorine bleach; TVOCs do not reflect chlorine bleach exposures. In EC-3 ($n = 500$), both high and low skin-wipes (more than 40%) and aerosolized medications (more than 25%), and low alcohols (28.2%) were the most common product applications. Licensed practical nurses or licensed vocational nurses, and registered nurses accounted for more than 70% of the total EC-3 participants. Additionally, 51 out of 75 respiratory therapists or respiratory technicians were included in EC-3. In EC-4 ($n = 368$), high alcohols (67.9%) and chlorine bleach (36.4%) were the most common product applications. In fact, other product applications were reported in small number of the participants in EC-4. Nursing assistants and lab technicians accounted for about 60% of the participants. The last EC, EC-5 ($n = 150$), was the most “exposed” group. More than half of the participants in EC-5 used high alcohols and high-level disinfectants, and more than 40% used high chlorine bleach and enzymes. About 50% of the participants in EC-5 were nursing assistants, licensed practical nurses or licensed vocational nurses, and central supply workers. Not surprisingly, EC-5 had the highest means and medians of frequency-weighted scores for TVOC14 and TVOC11 among all ECs. Extremely high values of TVOC exposures were observed in upper percentiles (e.g., the 95th percentile). EC-5 also had the highest percentage of the participants working in hospitals (72.7%).

3.3. Associations between HCs and exposures (individual product applications and ECs)

Using “no symptoms” cluster (HC-1) as the reference group, the effects of individual product applications (i.e., one product application in one model) on HCs are shown in Fig. 1 and Table A1. Six out of 14 product applications had statistically significant positive associations with at least one HC. The use of alcohols to sterilize medical instruments or clean and disinfect fixed-surface equipment or instruments was associated with “mild asthma symptoms”, and “asthma attacks/exacerbations” clusters. The use of chlorine bleach to sterilize medical instruments or clean and disinfect fixed-surface equipment or instruments was associated with “mild asthma symptoms”, “undiagnosed/untreated asthma”, and “asthma attacks/exacerbations” clusters. The use of enzymes to clean medical instruments or clean and disinfect fixed-surface equipment or instruments was associated with “undiagnosed/untreated asthma”, and “asthma attacks/exacerbations” clusters. The use of detergents to clean and disinfect fixed-surface equipment or instruments was associated with “undiagnosed/untreated asthma” cluster. Skin-wipes for use on patients was associated with “mild asthma symptoms”, and “asthma attacks/exacerbations” clusters. Administering aerosolized medications was associated with “asthma attacks/exacerbations” cluster.

Using “no products” cluster (EC-1) as the reference group, the effects of ECs for product applications on HCs are shown in Fig. 2 and Table A2. Except for the “winter cough/phlegm” cluster (HC-2), the other HCs showed significant associations with different ECs. The “housekeeping/chlorine” cluster was associated with “undiagnosed/untreated asthma” cluster (OR = 3.11, 95% CI = 1.46–6.63, p -value = 0.003) and “asthma attacks/exacerbations” cluster (OR = 2.71, 95% CI = 1.25–5.86, p -value = 0.011). The “patient care” cluster was associated with “asthma attacks/exacerbations” cluster

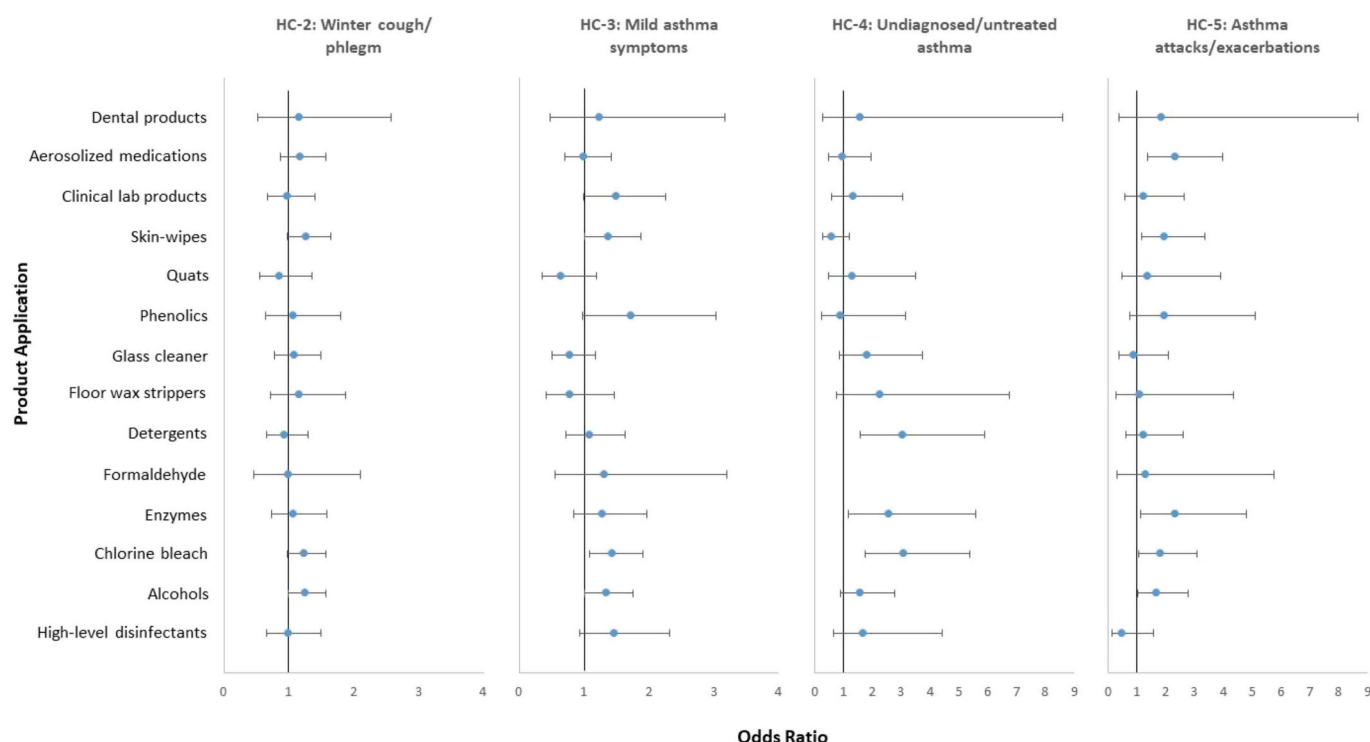


Fig. 1. Associations* between resulting asthma health clusters and individual product applications.

HC, asthma health cluster; quats, quaternary ammonium compounds.

*Each model included one main risk factor (i.e., one product application in one model) and adjustments for age, sex, education, smoking and allergic status. The figure shows the results for 14 individual multinomial logistic regression models.

(OR = 2.34, 95% CI = 1.16–4.72, p-value = 0.018). The “general cleaning/laboratory” cluster was associated with “asthma attacks/exacerbations” cluster (OR = 2.35, 95% CI = 1.07–5.13, p-value = 0.033). The “disinfection products” cluster was associated with “mild asthma symptoms” cluster (OR = 1.81, 95% CI = 1.09–2.99, p-value = 0.021) and “undiagnosed/untreated asthma” cluster (OR = 3.42, 95% CI = 1.24–9.39, p-value = 0.017).

For covariates (Table A2), allergic status had significant associations with all HCs, especially the “undiagnosed/untreated asthma” cluster (OR = 9.40, 95% CI = 4.54–19.4, p-value < 0.001) and “asthma attacks/exacerbations” cluster (OR = 12.7, 95% CI = 6.05–26.7, p-value < 0.001). Comparing with non-smokers, former and current smokers were significantly associated with the “mild asthma symptoms” cluster (OR = 1.66, 95% CI = 1.08–2.57, p-value = 0.021 for former smokers) and “undiagnosed/untreated asthma” cluster (OR = 3.05, 95% CI = 1.37–6.76, p-value = 0.006 for former smokers; OR = 3.58, 95% CI = 1.36–9.37, p-value = 0.010 for current smokers).

4. Discussion

4.1. Asthma symptom/care grouping

In occupational epidemiologic studies, asthma is often defined as a dichotomous outcome; however, it likely occurs along a continuous spectrum with varying degree of severity, phenotypes and subtypes (Deliu et al., 2016; Pekkanen et al., 2005). Using hierarchical clustering to group participants with similar profiles of asthma-related health, we identified five clusters that reflected unique symptom patterns and different predominant characteristics. HC-1 included participants with no symptoms. HC-2 had low prevalence of lower respiratory symptoms,

but had high prevalence of cough and phlegm production in the winter, and the highest proportion of those with no asthma and no allergies (except for HC-1). HC-3 had moderate prevalence of lower respiratory symptoms including cough and phlegm production, and likely represents mild asthma symptoms. HC-4 had the highest proportions of participants with respiratory symptoms, but moderate use of asthma medications and physician diagnosed asthma, and low proportion of urgent asthma treatment; this cluster likely represents undiagnosed/untreated asthma. Participants in HC-5 experienced more symptoms related to asthma attacks and exacerbation and use of urgent treatments and likely represents workers with asthma exacerbations. HC-4 and HC-5 represent workers at a more advanced stage of asthma than workers in HC-2 and HC3.

BHR-related symptoms and positive allergic status showed increasing trends in prevalence across HC-2 to HC-5; HC-2 to HC-4 were dominated by participants with allergies but no physician-diagnosed asthma, while HC-5 was predominantly participants with allergies and physician-diagnosed asthma. BHR-related symptoms and positive allergic status was common among cleaners with respiratory symptoms. Moreover, a recent review has reported co-morbidity of asthma and COPD especially after prolonged exposure or chronic asthma, which was observed in HC-4 and HC-5, clusters with the highest severity of asthma (Gibson and McDonald, 2015). Le Moual et al. (2005) defined asthma severity using a 7-grade (0–7) clinical score based on questions on frequency of asthma attacks, asthma symptoms and hospitalization and categorized the score to reflect mild or severe asthma. They observed significant associations between several classes of asthmagens including cleaning agents and severe asthma but not mild asthma among subjects with adult onset asthma. Severe asthma may be associated with different mechanism and phenotypes than mild asthma

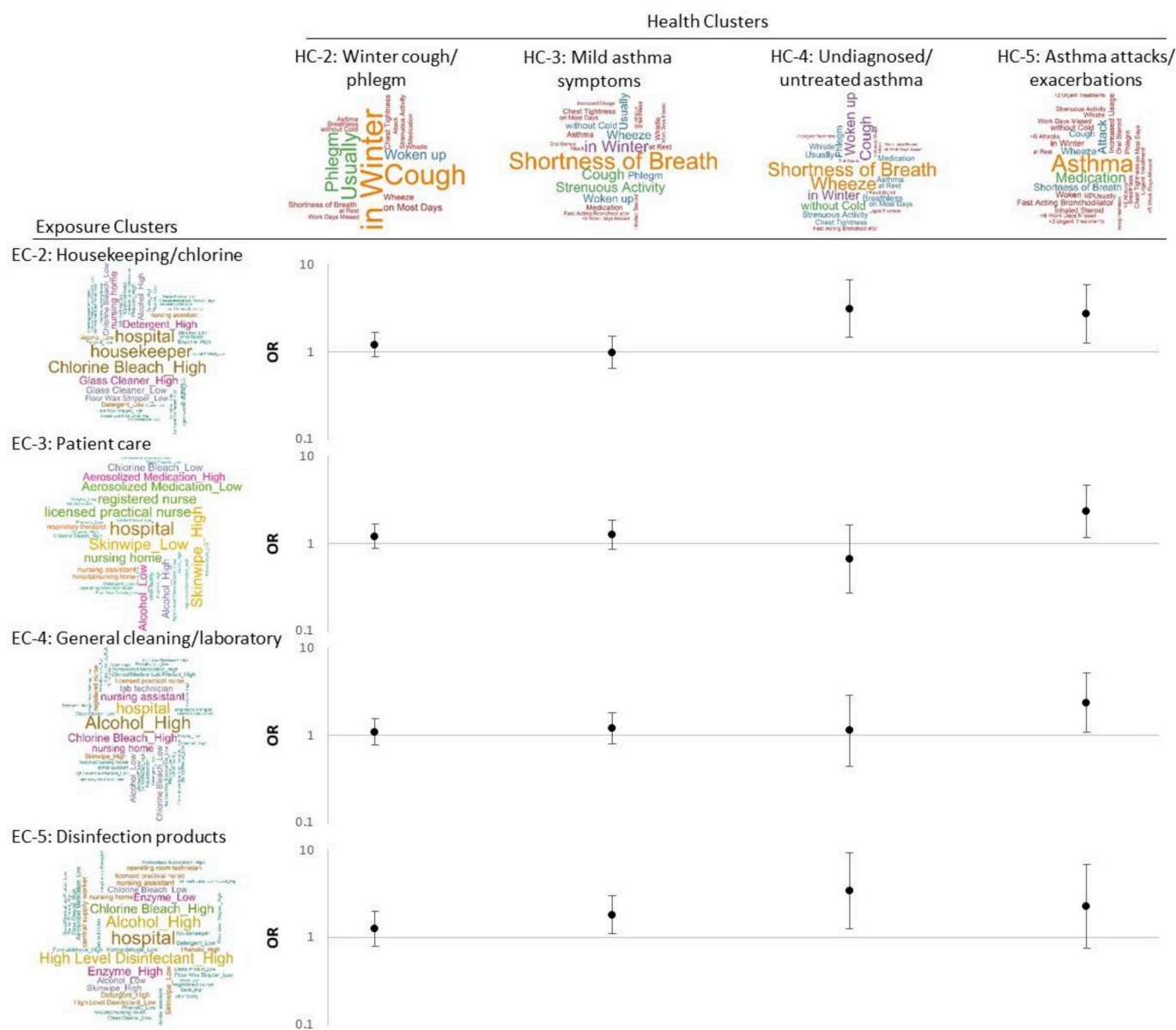


Fig. 2. Associations* between resulting clusters of asthma health outcomes and clusters of product applications.

HC, asthma health cluster; EC, exposure cluster, CI, confidence interval.

*The main risk factor was exposure clusters (one variable with 5 categories) included in one model with the adjustments for age, sex, education, smoking and allergic status. The figure shows the results for one multinomial logistic regression model.

(Wenzel et al., 2017). Clustering may help in focusing prevention efforts among those classified with having severe asthma to avoid disability in the future.

Several approaches have been used to explore patterns of asthma symptoms. A recent review summarized various grouping approaches utilized to characterize asthma subtypes (Deliu et al., 2016). The authors reviewed 41 asthma studies (23 studies recruited adult populations), and classified the grouping methods into two categories: PCA/factor analysis and mode-free approaches (e.g., hierarchical clustering with Ward's method). Additionally, Deliu et al. (2016) discussed a third approach to grouping model-based approaches (i.e., machine learning approaches) that were reviewed in another paper (Howard et al., 2015). While the identified asthma subtypes were based on data-driven approaches and needed to be interpreted cautiously (especially for clinical use), the asthma subtyping is still beneficial for hypothesis-generating research.

For the overall study population without clustering, positive allergic status was common (50.9%); participants with allergies had a higher risk of asthma (p -value < 0.0001; tested using chi-square test). While a literature review found equivocal evidence for the association of atopy with cleaning-related asthma (Folletti et al., 2017), one study found atopy to be common among those with “not so sudden onset” irritant asthma; they note that atopy status is an important risk factor for this type of asthma for pre-existing as well as new-onset asthma (Brooks et al., 1998). Asthma among cleaning occupations is thought to be predominantly irritant-induced (Vizcaya et al., 2011), but allergic asthma may also manifest among healthcare workers due to exposure to specific sensitizers such as quaternary ammonium compounds (Gonzalez et al., 2014), aldehydes, or products containing amines or fragrances (Zock et al., 2010; Labrecque, 2012), and indoor or environmental allergens (Simpson et al. (2001)), including molds and damp environments (Cox-Ganser et al., 2009; Kurth et al., 2017; Quirce

and Barranco, 2010). Thus both irritant and sensitizer asthma mechanisms may be important for asthma among healthcare workers. We also evaluated the health clusters in terms of the frequency of workers using products containing irritants, sensitizers, or both irritants and sensitizers, but did not observe significant differences among the clusters. In our analysis, the health clusters we observed are primarily characterized by the number and types of asthma symptoms, but we could not comment on asthma mechanisms or subtypes because our clustering variables were limited to symptoms variables only. Studies exploring asthma phenotypes and subtypes utilize different sources of information including questionnaires, detailed clinical observations, lung function testing and a variety of biomarkers of asthma mechanisms in cluster analysis (Deliu et al., 2016).

4.2. Cleaning and disinfecting product application grouping

While several individual product applications showed significant associations with HCs, grouping product applications was needed to avoid statistical issues and to represent mixed exposures (details described below). This study applied hierarchical clustering for variable/data reduction, and the resulting ECs showed moderate variability of product applications within occupations (especially for nursing occupations). This finding suggested that compared to job titles, product applications were more appropriate indicators of occupational exposures in the healthcare setting. Participants in the “no products” cluster (EC-1) had few positive responses to product application variables that resulted in low estimated TVOC exposures. Nurses were the most common occupations across all ECs. Among five ECs, participants in the “disinfecting” cluster (EC-5) had the highest concentrations for TVOC exposures. EC-5 had the higher proportions of high alcohol and chlorine bleach use, as well as high high-level disinfectant and enzyme use that were rarely used in other ECs. Compared with other ECs, EC-5 included more central supply workers. EC-2 had the highest proportion of use of chlorine bleach, glass cleaners, detergents and floor strippers, whose ingredients are not adequately captured by TVOC. Consequently, TVOC 14 and 11 were low for the EC-2 cluster that comprised mostly housekeepers. As shown in these clusters, product use is not independent, and multiple products are used simultaneously by workers making it difficult to evaluate their independent effects.

4.3. Associations between HCs and exposures

We found significant associations between five individual product applications and HC-5, the cluster with asthma exacerbations, including use of alcohol, chlorine releasing products, enzymatic cleaners, disinfectant wipes on patients and administration of aerosolized medications. HC-3 was associated with use of skin wipes, chlorine releasing products and alcohols, while HC-4 was associated with the use of detergents, enzymes and chlorine releasing products. For clustered exposures, HC-5 was associated with three ECs that reflected the highest exposures to alcohols, chlorine bleach, glass cleaners, detergents, use of skin wipes and aerosolized medications. HC-4 was significantly associated with EC-2 and EC-5 that comprised mixtures of housekeeping products and disinfecting products. HC-3 was associated with EC-5, reflecting a complex mixture of exposures that occur together and whose independent effects cannot be evaluated. These exposures have been previously reported as significant predictors of various asthma outcomes, e.g., use of bleach, enzymatic cleaner, cleaning medical instruments, general cleaning, administration of aerosolized medications, and using a variety of general purpose cleaning products (Arif et al., 2009; Delclos et al., 2007; Dumas et al., 2017; Vizcaya et al., 2011).

4.4. Strengths and limitations

This study collected comprehensive information on respiratory health and CDAs among a large population of healthcare workers, and estimated the effect of mixed risk factors on various asthma health components. The application of this data reduction approach avoided raising statistical issues, including multiple testing and multi-collinearity (due to the correlations among CDAs), and better represented the reality of health status (multiple symptoms occurring together) and exposure scenario (the use of mixed chemicals). Moreover, word clouds were creatively used to characterize participants in the resulting clusters, and to provide better visualization aids for the overlapped information among clusters, which enables more informed decisions on preventive actions. This study demonstrated that the combination of hierarchical clustering and word clouds was a useful approach to deal with a complex dataset that included a large number of potentially correlated variables.

As with any study utilizing self-reported exposures and health outcomes, there is a potential for information bias leading to differential or non-differential exposure misclassification. We developed the questionnaire based on standardized instruments, and asked questions mostly regarding routine activities (e.g., cleaning and disinfecting tasks and product-use) or recurring health events (e.g., usually coughed in the winter) in a reasonable recall period (i.e., last 12 months) to minimize such a bias (Althubaiti, 2016; Coughlin, 1990; Hassan, 2005). Participation weights were applied to adjust for potential bias due to the low response rate of the survey (Caridi et al., 2019); studies have shown that the participation rate is a weak indicator of the presence of bias (Galea and Tracy, 2007; Morton et al., 2012). Hierarchical clustering is a data-driven approach. The characterization of the corresponding variables (e.g., frequencies of allergies and diagnosed asthma) showed reasonable consistency with input variables by cluster. It suggested that hierarchical clustering grouped participants meaningfully.

5. Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health (NIOSH). Mention of any company or product does not constitute endorsement by NIOSH. In addition, citations to Web sites external to NIOSH do not constitute NIOSH endorsement of the sponsoring organizations or their programs or products. Furthermore, NIOSH is not responsible for the content of these Web sites. Its contents, including any opinions and/or conclusions are solely those of the authors.

6. Conclusions

This study examined the inherent patterns of asthma health and CDAs among healthcare workers in a complex survey dataset. We identified various combinations of cleaning and disinfecting product applications like using alcohols, bleach, high-level disinfectants, and enzymes to disinfect instruments and equipment as risk factors of mild (e.g., frequent coughing) to severe (e.g., asthma attacks) asthma health components. The unique pattern of symptoms that characterize the various degrees of asthma severity, and their association with exposure clusters can be used to better inform prevention strategies and treatment options to avoid disease progression to disability in the future. This study demonstrated that hierarchical clustering and word clouds were useful techniques for analyzing a complex dataset with a large number of potentially correlated variables to generate practical information that can be used to inform prevention.

Table A1

Results of the multinomial logistic regression models^a for asthma health clusters and individual product applications.

Variable	Health Clusters; reference = HC-1: No symptoms (n = 885)												
	HC-2: Winter cough/phlegm (n = 640)				HC-3: Mild asthma symptoms (n = 357)			HC-4: Undiagnosed/untreated asthma (n = 63)			HC-5: Asthma attacks/exacerbations (n = 85)		
	n	OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI	
<u>Individual Product Applications</u>													
High-level disinfectants	156	0.99	0.66	1.50	1.47	0.94	2.31	1.69	0.65	4.42	0.48	0.14	1.57
Alcohols	916	1.25	1.00	1.57	1.34	1.01	1.76	1.56	0.88	2.77	1.69	1.03	2.77
Chlorine bleach	783	1.24	0.98	1.57	1.44	1.09	1.91	3.07	1.75	5.39	1.82	1.07	3.09
Enzymes	196	1.07	0.73	1.58	1.29	0.84	1.97	2.57	1.18	5.58	2.34	1.14	4.78
Formaldehyde	38	0.99	0.46	2.11	1.32	0.55	3.20	.	.	.	1.31	0.30	5.75
Detergents	253	0.93	0.67	1.30	1.08	0.72	1.63	3.04	1.56	5.90	1.24	0.60	2.59
Floor wax stripper	117	1.16	0.72	1.87	0.78	0.42	1.46	2.26	0.76	6.76	1.11	0.28	4.33
Glass cleaners	279	1.08	0.78	1.50	0.77	0.51	1.18	1.81	0.87	3.75	0.88	0.37	2.08
Phenolics	107	1.08	0.64	1.80	1.72	0.97	3.03	0.88	0.25	3.14	1.94	0.74	5.11
Quaternary ammonium compounds	125	0.86	0.55	1.37	0.64	0.35	1.19	1.29	0.47	3.49	1.37	0.48	3.90
Skin-wipes used on patients	571	1.27	0.98	1.65	1.37	1.00	1.87	0.57	0.27	1.19	1.96	1.15	3.36
Clinical or medical laboratory products	244	0.98	0.68	1.41	1.50	1.00	2.26	1.32	0.58	3.04	1.22	0.57	2.62
Aerosolized medications	395	1.17	0.88	1.57	0.99	0.70	1.42	0.97	0.48	1.94	2.34	1.38	3.98
Dental products	41	1.15	0.52	2.57	1.23	0.48	3.16	1.57	0.29	8.58	1.83	0.39	8.65

OR, odds ratio; CI, confidence interval; SOB, shortness of breath.

^a Each model included one main risk factor (i.e., one product application in one model) and adjustments for age, sex, education, smoking and allergic status. The table shows the results for 14 individual models.

Table A2

Results of the multinomial logistic regression model^a for asthma health and cleaning and disinfecting clusters.

Variable	Health Clusters; reference = HC-1: No symptoms (n = 885)												
	n	HC-2: Winter cough/phlegm (n = 640)			HC-3: Mild asthma symptoms (n = 357)			HC-4: Undiagnosed/untreated asthma (n = 63)			HC-5: Asthma attacks/exacerbations (n = 85)		
		OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI	
Age (year)													
40–49	485	0.85	0.62	1.16	0.89	0.60	1.30	1.85	0.76	4.52	1.40	0.59	3.31
50–59	697	0.84	0.62	1.13	1.03	0.72	1.48	1.52	0.62	3.71	2.78	1.29	5.97
60–83	360	0.75	0.53	1.05	0.71	0.45	1.10	0.81	0.28	2.33	1.57	0.63	3.90
18–39	488												
Sex													
Female	1542	1.07	0.82	1.40	1.43	1.01	2.01	4.97	2.05	12.1	2.81	1.28	6.17
Male	487												
Education													
< HS	97	1.62	0.91	2.86	1.43	0.71	2.87	0.93	0.23	3.81	2.64	0.90	7.76
HS or < College	1319	1.52	1.17	1.97	1.22	0.90	1.66	1.50	0.76	2.95	0.95	0.55	1.64
College +	594												
Smoking													
Former	226	1.16	0.80	1.69	1.66	1.08	2.57	3.05	1.37	6.76	2.00	0.97	4.14
Current	113	1.13	0.68	1.90	1.43	0.80	2.56	3.58	1.36	9.37	0.79	0.20	3.07
Never	1676												
Allergic Status													
Yes	1033	2.58	2.06	3.23	4.26	3.21	5.65	9.40	4.54	19.4	12.7	6.05	26.7
No	997												
EC-2:													
House-keeping/chlorine	382	1.20	0.87	1.66	0.98	0.64	1.51	3.11	1.46	6.63	2.71	1.25	5.86
EC-3:													
Patient care	500	1.21	0.88	1.65	1.26	0.86	1.85	0.66	0.27	1.63	2.34	1.16	4.72
EC-4:													
General cleaning/laboratory	368	1.09	0.78	1.53	1.20	0.80	1.81	1.13	0.44	2.87	2.35	1.07	5.13
EC-5:													
Disinfection products	150	1.26	0.79	2.01	1.81	1.09	2.99	3.42	1.24	9.39	2.27	0.76	6.84
EC-1:													
No products													

OR, odds ratio; CI, confidence interval; SOB, shortness of breath; HS, high school.

^a The main risk factor was exposure clusters (one variable with 5 categories) that included in one regression model with other covariates. The table shows the results for one model.

Acknowledgements

The authors would like to thank Laura M. Kurth and Girija Syamlal for their review of the manuscript. Funding for this project was provided by the National Institute for Occupational Safety and Health. Melissa C. Friesen was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, NCI (Z01 CP010122).

References

- Althubaiti, A., 2016. Information bias in health research: definition, pitfalls, and adjustment methods. *J. Multidiscip. Healthc.* 9, 211–217.
- Arif, A.A., Delclos, G.L., Serra, C., 2009. Occupational exposures and asthma among nursing professionals. *Occup. Environ. Med.* 66, 274–278.
- Bradding, P., Green, R.H., 2010. Subclinical phenotypes of asthma. *Curr. Opin. Allergy Clin. Immunol.* 10, 54–59.
- Burney, P., Luczynska, C., Chinn, S., Jarvis, D., 1994. The European community respiratory health survey. *Eur. Respir. J.* 7, 954–960.
- Caridi, M.N., Humann, M.J., Liang, X., Su, F.-C., Stefaniak, A.B., LeBouf, R.F., Stanton, M.L., Virji, M.A., Henneberger, P.K., 2019. Occupation and task as risk factors for asthma-related outcomes among healthcare workers in New York City. *Int. J. Hyg Environ. Health* 222 (2), 211–220.
- Coughlin, S.S., 1990. Recall bias in epidemiologic studies. *J. Clin. Epidemiol.* 43, 87–91.
- Cox-Ganser, J.M., Rao, C.Y., Park, J.-H., Schumpert, J.C., Kreiss, K., 2009. Asthma and respiratory symptoms in hospital workers related to dampness and biological contaminants. *Indoor Air* 19, 280–290.
- Delclos, G.L., Arif, A.A., Aday, L., Carson, A., Lai, D., Lusk, C., Stock, T., Symanski, E., Whitehead, L.W., Benavides, F.G., Antó, J.M., 2006. Validation of an asthma questionnaire for use in healthcare workers. *Occup. Environ. Med.* 63, 173–179.
- Delclos, G.L., Gimeno, D., Arif, A.A., Burau, K.D., Carson, A., Lusk, C., Stock, T., Symanski, E., Whitehead, L.W., Zock, J.-P., Benavides, F.G., Antó, J.M., 2007. Occupational risk factors and asthma among health care professionals. *Am. J. Respir. Crit. Care Med.* 175, 667–675.
- Deliu, M., Sperrin, M., Belgrave, D., Custovic, A., 2016. Identification of asthma subtypes using clustering methodologies. *Pulmonar. Therap.* 2, 19–41.
- Dodd, K., Mazurek, J., 2016. Asthma among employed adults, by industry and occupation - 21 States, 2013. *MMWR (Morb. Mortal. Wkly. Rep.)* 65, 1325–1331.
- Dumas, O., Wiley, A.S., Quinot, C., Varraso, R., Zock, J.-P., Henneberger, P.K., Speizer, F.E., Le Moual, N., Camargo Jr., C.A., 2017. Occupational exposure to disinfectants and asthma control in US nurses. *Eur. Respir. J.* 50, 1700237.
- Engel, J., 1988. Polytomous logistic regression. *Stat. Neerl.* 42, 233–252.
- Feinerer, I., Hornik, K., 2017. tm: text mining package. R package version 0.1-3. <https://CRAN.R-project.org/package=tm>.
- Fellows, I., 2014. Wordcloud: word clouds. R package version 2.5. <https://CRAN.R-project.org/package=wordcloud>.
- Ferris, B.G., 1978. Epidemiology standardization project (American thoracic society). *Am. Rev. Respir. Dis.* 118 1–120.
- Folletti, I., Siracusa, A., Paolucci, G., 2017. Update on asthma and cleaning agents. *Curr. Opin. Allergy Clin. Immunol.* 17, 90–95.
- Friedman-Jimenez, G., Harrison, D., Luo, H., 2015. Occupational asthma and work-exacerbated asthma. *Semin. Respir. Crit. Care Med.* 36, 388–407.
- Galea, S., Tracy, M., 2007. Participation rates in epidemiologic studies. *Ann. Epidemiol.* 17 (9), 643–653. <https://doi.org/10.1016/j.annepidem.2007.03.013>.
- Gibson, P.G., McDonald, V.M., 2015. Asthma–COPD Overlap 2015: Now We Are Six Thorax. Global Initiative for Asthma, 2018. Global Strategy for Asthma Management and Prevention. Available from: www.ginasthma.org.
- Gonzalez, M., Jégu, J., Kopferschmitt, M.C., Donnay, C., Hedelin, G., Matzinger, F., Velten, M., Guilloux, L., Cantineau, A., de Blay, F., 2014. Asthma among workers in healthcare settings: role of disinfection with quaternary ammonium compounds. *Clin. Exp. Allergy* 44 (3), 393–406.
- Hassan, E., 2005. Recall bias can be a threat to retrospective and prospective research designs. *Internet J. Epidemiol.* 3.
- Howard, R., Rattray, M., Prosperi, M., Custovic, A., 2015. Distinguishing asthma phenotypes using machine learning approaches. *Curr. Allergy Asthma Rep.* 15, 38.
- Jarvis, D., 2002. European community respiratory health survey II steering committee. The European Community Respirator. *Health Surv. II. Eur. Respirator. J.* 20, 1071–1079.
- Kurth, L., Virji, M.A., Storey, E., Framberg, S., Kallio, C., Fink, J., Laney, A.S., 2017. Current asthma and asthma-like symptoms among workers at a Veterans Administration Medical Center. *Int. J. Hyg Environ. Health* 220, 1325–1332.
- Labrecque, M., 2012. Irritant-induced asthma. *Curr. Opin. Allergy Clin. Immunol.* 12 (2), 140–144.
- Le Moual, N., Siroux, V., Pin, I., Kauffmann, F., Kennedy, S., 2005. Asthma severity and exposure to occupational asthrogens. *Am. J. Respir. Crit. Care Med.* 172, 440–445.
- Marbac, M., Sedki, M., Boutron-Ruault, M.-C., Dumas, O., 2018. Patterns of cleaning product exposures using a novel clustering approach for data with correlated variables. *Ann. Epidemiol.* 28, 563–569 e566.
- Mazurek, J.M., Weissman, D.N., 2016. Occupational respiratory allergic diseases in healthcare workers. *Curr. Allergy Asthma Rep.* 16, 77.
- Morton, S.M.B., Bandara, D., Robinson, E., Carr, P.E.A., 2012. In the 21st Century, what is an acceptable response rate? *Aust. N. Z. J. Public Health* 36 (2), 106–108.
- Nagelkerke, N.J.D., 1991. A note on a general definition of the coefficient of determination. *Biometrika* 78, 691–692.
- Pekkanen, J., Sunyer, J., Anto, J.M., Burney, P., 2005. Operational definitions of asthma in studies on its aetiology. *Eur. Respir. J.* 26, 28–35.
- Quirce, S., Barranco, P., 2010. Cleaning agents and asthma. *J. Investig. Allergol. Clin. Immunol.* 20, 542–550.
- R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Shtatland, E.S., Kleinman, K., Cain, E.M., 2002. One more time about R2 measures of fit in logistic regression. In: NESUG 15 Proceedings, pp. 222–226.
- Simpson, B.M., Custovic, A., Simpson, A., Hallam, C.L., Walsh, D., Marolia, H., Campbell, J., Woodcock, A., 2001. NAC Manchester Asthma and Allergy Study (NACMAAS): risk factors for asthma and allergic disorders in adults. *Clin. Exp. Allergy* 31, 391–399.
- Su, F.-C., Friesen, M.C., Stefaniak, A.B., Henneberger, P.K., LeBouf, R.F., Stanton, M.L., Liang, X., Humann, M., Virji, M.A., 2018. Exposures to volatile organic compounds among healthcare workers: modeling the effects of cleaning tasks and product-use. *Ann. Work Expos. Health* 62 (7), 852–870.
- Sunyer, J., Pekkanen, J., Garcia-Esteban, R., Svanes, C., Künzli, N., Janson, C., De Marco, R., Antó, J.M., Burney, P., 2007. Asthma score: predictive ability and risk factors. *Allergy* 62, 142–148.
- Tarlo, S.M., Balmes, J., Balkissoon, R., Beach, J., Beckett, W., Bernstein, D., Blanc, P.D., Brooks, S.M., Cowl, C.T., Daroowalla, F., Harber, P., Lemiere, C., Liss, G.M., Pacheco, K.A., Redlich, C.A., Rowe, B., Heitzer, J., 2008. Diagnosis and management of work-related asthma: American college of chest physicians consensus statement. *Chest* 134, 1S–41S.
- Vizcaya, D., Mirabelli, M.C., Antó, J.-M., Orriols, R., Burgos, F., Arjona, L., Zock, J.-P., 2011. A workforce-based study of occupational exposures and asthma symptoms in cleaning workers. *Occup. Environ. Med.* 68, 914–919.
- Ward, J., Joe, H., 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58, 236–244.
- Wenzel, S.E., Brillhart, S., Nowack, K., 2017. An invisible disease: severe asthma is more than just “bad asthma”. *Eur. Respir. J.* 50.
- Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.
- Wisniewska, M., Walusiak-Skorupa, J., 2014. Occupational allergy: respiratory hazards in healthcare workers. *Curr. Opin. Allergy Clin. Immunol.* 14, 113–118.
- Zock, J.-P., Vizcaya, D., Le Moual, N., 2010. Update on asthma and cleaners. *Curr. Opin. Allergy Clin. Immunol.* 10, 114–120.