

Clustering analysis of volatile organic compound biomarkers with tobacco exposure and the association with cardiovascular health outcomes in an observation study cohort

Juan Zhao¹, Haoyun Hong¹, Joseph Zhai¹, Remy Poudel¹, Sanjay Srivastava², Andrew C. Stokes³, Pawel K. Lorkiewicz², Tian Jiang¹, Rose Marie Robertson¹, Aruni Bhatnagar², Jennifer L. Hall¹, Naomi M. Hamburg⁴, Rachel J. Keith²

ABSTRACT

INTRODUCTION Volatile organic compounds (VOCs) are toxic compounds found in tobacco smoke. Despite research on cigarette generated single VOCs, scant evidence exists on the mixtures of VOCs associated with different tobacco products. We aimed to explore whether distinct VOC exposure profiles exist among users of combustible cigarettes, e-cigarettes, and non-users, and to assess their associations with cardiovascular (CV) health markers.

METHODS Participants who self-reported use of e-cigarettes, cigarettes, or no tobacco (n=348; mean age 26 ± 7 years) enrolled in The Cardiovascular Injury due to Tobacco Use (CITU) 2.0 study from July 2018 to July 2023 at two US sites (Boston, MA, and Louisville, KY). VOC metabolites were analyzed in urine one-hour post-use of a tobacco product via ultraperformance liquid chromatography. We applied unsupervised K-Means clustering on the creatinine-adjusted VOC metabolite data and explored the association between each cluster and blood pressure, adjusting for age, sex, and race.

RESULTS The clustering analysis identified two distinct clusters. Cluster 1 (302 individuals, 86.8%) was characterized by low VOC metabolite levels with individuals predominantly e-cigarette users (59.3%), non-users (29.1%), and a smaller proportion of cigarette smokers (11.6%). Cluster 2 (46 individuals, 13.2%) had higher levels of VOC metabolites including CYMA, HPMMA, MHBMA3, and 3HPMA, and included most of the individuals who used cigarettes (91.3%). After adjustment for age, sex, and race, Cluster 2 was associated with a higher heart rate ($\beta=3.29$; 95% CI: -0.26–6.84; $p<0.05$) compared to Cluster 1. No significant differences were observed for systolic ($\beta=-0.66$; 95% CI: -4.60–3.28) or diastolic blood pressure ($\beta=0.34$; 95% CI: -2.51–3.2) between clusters.

CONCLUSIONS These findings suggest that cigarette-induced VOC exposure may not impact cardiovascular function after acute exposure. Additionally, VOC exposure profiles vary across tobacco product types, suggesting that regulatory assessments of tobacco products could consider exposure patterns rather than product types. Clustering analyses may offer a powerful tool to assess the safety and risks of new and emerging tobacco products based on real-world exposure patterns.

AFFILIATION

¹ American Heart Association, Dallas, United States

² American Heart Association Tobacco Regulation and Addiction Center, University of Louisville, Louisville, United States

³ Department of Global Health, Boston University School of Public Health, Boston, United States

⁴ Vascular Biology Section, Whitaker Cardiovascular Institute, Boston University School of Medicine, Boston, United States

CORRESPONDENCE TO

Rachel J. Keith. American Heart Association Tobacco Regulation and Addiction Center, University of Louisville, 302 E Muhammad Ali Blvd, Suite 200, Louisville KY 40202, United States
E-mail: Rachel.Keith@louisville.edu

KEYWORDS

volatile organic compounds, e-cigarettes, combustible cigarettes, cardiovascular health, hazardous and potentially harmful chemicals

Received: 6 November 2024

Revised: 29 January 2025

Accepted: 3 February 2025

INTRODUCTION

Volatile organic compounds (VOCs) are ubiquitous hazardous and potentially

hazardous chemicals that are emitted from various sources including industrial emissions and household products¹, and are also found in high levels in tobacco products². Tobacco-induced VOC exposure can be inhaled directly from active use or indirectly from secondhand exposure to tobacco products such as combusted tobacco³ and electronic cigarettes (e-cigarettes)^{3,4}. These hazardous compounds have been associated with several health outcomes, including respiratory disease⁵, cancer⁶, and heart disease⁷⁻⁹. Understanding VOC exposure is critical to determining the health effects of tobacco products and potentially offering the FDA an innovative strategy to assess the risks of new tobacco products.

While existing research has largely focused on individual VOCs^{6,10,11}, real-world exposure involves mixtures of multiple compounds, which may have a different effect as a mixture, for example synergistic or antagonistic effects on health. Analyzing VOCs individually overlooks these complex interactions, limiting our understanding of their cumulative impact. Identifying the exposure patterns from different tobacco products is crucial for policymakers to assess health risks and ensure appropriate product standards¹². However, traditional regulatory approaches such as substantial equivalence focus on broad tobacco product classes with similar design features rather than the specific exposure patterns and risks posed by individual product use^{13,14}.

With the growing understanding of the importance of mixture models there remains a gap in literature of mixture evaluation of the health effects caused by hazardous and potentially harmful compounds associated with tobacco products, such as VOCs¹⁵⁻¹⁷. Given the potential for additive health effects within mixtures, health assessments derived from single-source analyses may not capture real world relevance. Thus, it is important for policymakers to understand how different tobacco products and use patterns are associated with specific patterns of co-exposure.

This study aims to explore whether distinct VOC patterns are associated with the use of e-cigarettes and combustible cigarettes. Using data from the Cardiovascular Injury due to Tobacco Use 2.0 (CITU 2.0) study, a wide-ranging demographic cohort inclusive of individuals who currently use cigarettes, e-cigarettes, or have never used tobacco, we employed

an unsupervised clustering method to identify VOC exposure profiles and assessed their association with cardiovascular disease (CVD) health markers.

METHODS

Study design

Participants were enrolled in the CITU 2.0 study from July 2018 to July 2023 at two US sites (Boston, MA, and Louisville KY). The participants were briefly self-reported healthy adults who currently used e-cigarettes (use for past 3 months and at least 3 times per week), cigarettes (>100 cigarettes and use for past 3 months and at least 3 times per week), cigarillos (use for past 3 months and at least 3 times per week) or non-users who had never used any tobacco products (<100 lifetime uses of any tobacco). Each institutional review board approved CITU 2.0 and all participants provided written consent (BU #H-32613 and UL #13.0590).

Inclusion and exclusion criteria

At the time of conducting this analysis, 365 participants enrolled in the CITU 2.0 cohort. Exclusion criteria included missing demographic data needed for adjustment and use of cigarillos on the day of their visit or use of a product that may have contained cannabis instead of a nicotine vape liquid (Supplementary file Figure S1). We also excluded individuals without urinary measures of VOC measures at 1 hour post-exposure or without urinary creatinine. Participants who were part of the cigarillo groups were excluded because of the small sample size (n=7).

Study protocol

Study visits were scheduled after an 8-h food fast and a 6-h tobacco fast. All study visits occurred before 11 a.m. to limit effects due to circadian changes. Each visit included a structured interview on demographics, socioeconomic, lifestyle, health, family history of heart disease, allergies, and tobacco use. Detailed self-reported tobacco use history was collected using a modified version of the National Health Interview survey on tobacco use¹⁸ and surveys harmonized with the PhenX toolkit to include detailed information on ENDS and non-traditional tobacco products. Fasting blood and urine samples were collected, as

well as vascular health indices. All vascular function studies were completed after 10 min of supine positioning. A centralized laboratory at the University of Louisville processed and performed urinary and blood measurements. All surveys were collected and kept in Research Electronic Data Capture (REDCap), a secure web application for building and managing online surveys and databases^{19,20}.

Acute tobacco exposure session

After completion of the baseline measures, participants were asked to complete a structured use protocol for smoking, vaping, or sham (those who do not use tobacco products) within our specially designed exposure rooms. Participants were asked to bring in their typical product for this session, with dual users bringing an e-cigarette. Those who would smoke were asked to use a completely combustible cigarette for ≤ 10 min. Those who would vape were asked to bring a new vape or filled device of their most common current product, flavor, and nicotine strength. They were then asked to complete one 3-s puff every 30 s for 10 min. If participants reported they were unable to handle the nicotine content from 2 puffs/min, they could reduce their puffs to 1/min. Those who did not use any tobacco products were asked to inhale on a straw for three seconds every 30 s for 10 min. Urine and vascular measures were collected as described one and two hours after the exposure session²¹.

VOC measurements

Standard clean catch urine specimens were obtained from participants and stored at 4°C. The samples were transported for long-term storage and mass spectrometric analysis at the University of Louisville. A total of 23 urinary metabolites of tobacco-induced aldehydes and other VOCs were quantified with a modified version of the mass spectrometry method developed by the Centers for Disease Control and Prevention (CDC)²² and described in detail by Lorkiewicz et al.²³. The analysis was performed on an ACQUITY UPLC core system and a Quattro Premier XE triple quadrupole mass spectrometer coupled with an electrospray source (Waters, Inc, MA). Urine (25 μ L) was mixed with 15 mM ammonium acetate (975 μ L) containing a mixture of internal standards and filtered through a 0.2 mm polytetrafluoroethylene

membrane. Two microliters of the sample were applied on ACQUITY UPLC HSS T3 column (150 mm \times 2.1 mm, 1.8 μ m; Waters, Inc) maintained at 40°C and preequilibrated with ammonium acetate (15 mM, pH 6.8; solvent A) at a flow rate of 0.45 mL/min. The binary gradient started with 3% solvent A at 0 min and was linearly increased to 5% solvent B (acetonitrile) at 1.3 min, 10% B at 2.0 min, 30% B at 3.35 min, and 40% at 4.36 min. The gradient was then decreased to 15% B at 4.7 min, 10% B at 5.0 min, and 3% B at 5.36 min. The samples were analyzed, both in positive and negative ion modes. Three multiple reaction monitoring procedures were conducted for each analyte: one for quantification, one for confirmation, and one for stable isotope labeled analogous internal standard. At least 12 data points across the peaks were used for the quantitation of peak area. Analytes in urine samples were quantified using peak area ratio (analyte to internal standard) based on 10-point-standard curves that were run before and after the urine samples. TargetLynx quantification application manager software (Waters, Inc.) was used for peak integration, calibration, and quantification. The concentration ranges determined for this method are comparable to those reported in the CDC method for the VOC metabolites^{22,23}. Similarly, the reproducibility of the method was satisfactory, with relative SDs below 8% for VOC metabolites and alkaloids (5.5% for cotinine). Additional validation shows comparability in terms of sensitivity, accuracy, and precision compared to the CDC.

The concentration values of analytes obtained from ultra-performance liquid chromatography-tandem mass spectrometer were normalized to the creatinine level, which was measured on a COBAS MIRA-plus analyzer (Roche, NJ) with Infinity Creatinine Reagent (Thermo Fisher Scientific, MA)²³.

Cluster analysis

We conducted a cluster analysis on VOC biomarkers from the CITU 2.0 data set, utilizing the KMeans algorithm to identify distinct biomarker profiles. VOC measurements that were obtained one hour after smoking were fed into the clustering method. To determine the optimal number of clusters that were derived, we used the Elbow method, and used principal component analysis (PCA) facilitated

visualization of cluster differentiation. To ensure data quality, VOCs with incomplete records or invalid entries – such as 12DCVMA, TCVMA, AMCC, and CEMA – were excluded. Additionally, any VOCs with >85% of values below detection limits were removed from analysis. After this preprocessing, a total of 12 metabolites (i.e. AAMA, 3HPMA, 2HPMA, MA, DHBMA, MHBMA3, PGA, HPMMA, 2MHA, 34MHA, BMA, and CYMA) were finally included in the analysis (Supplementary file Table S1). For values below detection limits, imputation was performed using values close to zero. To adjust urine dilution effects, all VOC values were normalized by dividing them by the creatine levels measured at the corresponding urinary time point.

Statistical analysis

Descriptive statistics were computed for each cluster. Univariate tests were performed to compare the difference between clusters with two-sample t-tests for continuous variables, and the Z-proportion test for categorical variables. To examine the association between the clusters and surrogate cardiovascular (CV) health outcomes (e.g. systolic and diastolic blood pressure, heart rate), we used generalized linear regression by adjusting age, gender, and race/ethnicity. We report the β co-efficient and 95% confidence interval (CI). The significance level was defined as a $p < 0.05$.

RESULTS

The study included 348 unique participants after applying the inclusion and exclusion criteria (Supplementary file Figure S1). The average age was 26.1 ± 7.1 years. Of the 348 participants, 77 were cigarette smokers, 183 were e-cigarette users, and 88 were non-users. Among the 348 participants, 54 were dual users, 53.2% were males, and 62.1% were non-Hispanic White.

Demographic and product use characteristics of clusters

The clustering process resulted in 2 distinct clusters based on VOC biomarker profiles (Supplementary file Figure S2). Table 1 and Supplementary file Figure S3 present the demographic and product use characteristics within each cluster. Most individuals

in Cluster 1 ($n=302$) were e-cigarette users (59.3%), with a smaller number of cigarette smokers (11.6%) and a few non-users (29.1%). In contrast, Cluster 2, with 46 individuals, was composed mostly of cigarette users (91.3%), with few using e-cigarettes (8.7%), and none reported as non-users or dual users.

Cluster 1 and Cluster 2 also demonstrate notable differences in demographic characteristics. Participants in Cluster 1 are likely to be younger (median age 22 years) and comprise 48.2% biological females (Table 1). The majority self-reported their race as White (60.3%), followed by Asian (26.8%), and a small percentage as Black (5.6%). Most individuals in this cluster self-report as non-Hispanic.

Table 1. Demographic characteristics for two clusters

Characteristics	Cluster 1 n (%)	Cluster 2 n (%)
Sample size, n	302	46
Age (years), median (IQR)	22 (7.0)	34 (9.75)*
Sex, Female	146 (48.2)	17 (37.0)
Race		
White	182 (60.3)	34 (73.9)
Black	17 (5.6)	8 (17.4)*
Asian	81 (26.8)	1 (2.2)*
Other	22 (7.3)	3 (6.5)
Ethnicity		
Hispanic, Latino or Spanish	38 (12.6)	0 (0.0)*
Education level		
Lower than high school	10 (3.4)	5 (11.1)*
High school	35 (11.9)	13 (28.9)*
Some college	113 (38.4)	11 (24.4)
College and higher	136 (46.3)	16 (35.6)
Product type use		
Cigarette	35 (11.6)	42 (91.3)*
E-cigarette	179 (59.3)	4 (8.7)*
Non-user	88 (29.1)	0 (0.0)*
Dual use	54 (17.9)	0 (0.0)*
Outcomes		
Systolic blood pressure (mmHg), median (IQR)	115.0 (17.0)	119.0 (15.5)*
Diastolic blood pressure (mmHg), median (IQR)	70.0 (12.0)	74.5 (13.8)*
Heart rate (bpm), median (IQR)	63.0 (13.0)	67 (13.0)*

Statistical test: Kruskal-Wallis for continuous variables; Z-proportion test for categorical variables. Significance level: $p < 0.05$. *Significant. IQR: interquartile range.

Education completed is varied, with 38.4% having some college education or higher (46.3%). In contrast, Cluster 2 tends to be older (median age 34 years) and consists of a smaller proportion of females (37.0%). The cluster has a larger percentage of individuals self-reported as White (73.9%) and Black (17.4%), but fewer as Asian (2.2%). No one from this cluster identifies as Hispanic, Latino, or Spanish. Cluster 2 also shows a lower proportion with some college education (24.4%) or higher degrees (35.6%).

In terms of health outcomes, Cluster 2 has higher median systolic (119.0 vs 115.0 mmHg), diastolic blood pressures (74.5 vs 70.0 mmHg), and heart rate (67 vs 63 bpm).

VOCs profiles by cluster

Figure 1 shows the standardized VOC values among each cluster. Generally, Cluster 2 shows higher median values across most VOCs, particularly prominent in compounds such as DHBMA, MHB3A, and BMA. This indicates a higher concentration of VOCs presenting in Cluster 2, which could correlate with a higher level of tobacco exposure. When comparing VOC distributions to the usage patterns, the high representation of cigarette smokers in Cluster 2 aligns with the higher VOC levels in this cluster. Outliers

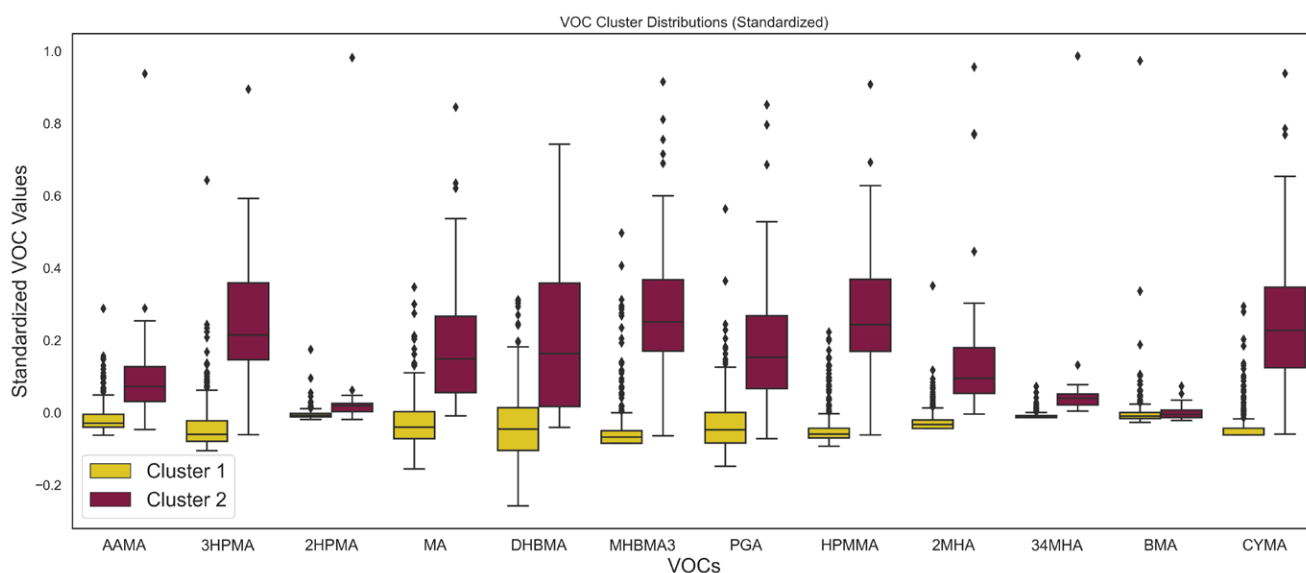
are present in both clusters but are more prevalent in Cluster 2, indicating individual variations in VOC metabolite excretion.

We employed PCA along with a scatter plot to visualize the distinction between the two clusters within the first two principal component (PC) dimensions, as shown in Figure 2. The first two PCs explained 57.17% variance. Cluster 1 appears predominantly concentrated, suggesting that VOCs characteristic of this cluster is more homogeneous which differentiates them from Cluster 2. The PCA loading plot (Supplementary file Figure S4) reveals the VOCs contribute most to the differentiation. Specifically, compounds such as CYMA, HPMMA, MHBMA3, and 3HPMA exhibit high positive loadings on the first principal component (PC1), underscoring their role in differentiating Cluster 1. Conversely, BMA shows a high positive loading on PC2, suggesting its association with Cluster 1. This bimodal distribution underscores distinct biomarker profiles, potentially reflecting different tobacco exposure statuses within the population.

Association between cluster and health outcomes

Table 2 contains the results of regression analysis that

Figure 1. Box plots of standardized volatile organic compound (VOC) values across two identified clusters. The distribution of each VOC is summarized through its median, quartiles, and outliers. The plots highlight the variance within and between the clusters for each VOC, indicating the different exposure or metabolic profiles that characterize each cluster



examined the association between health outcomes and two clusters, with Cluster 1 serving as the reference group. The unadjusted models indicated that Cluster 2 had significantly higher systolic blood pressure, diastolic blood pressure, and heart rate, with β coefficients of 4.21 (95% CI: 0.34–8.09), 3.70

(95% CI: 1.06–6.35) and 4.58 (95% CI: 1.36–7.79), respectively. After adjusting for age, gender, and race/ethnicity, the difference in heart rate was no longer statistically significant with a β of 3.29 (95% CI: -0.29–6.84).

DISCUSSION

Our findings underscore the significance of considering VOC mixtures in regulatory assessments. This approach yields a mixture-based understanding of hazardous and potentially harmful compounds potential health risks and has substantial implications for public health policies aimed at tobacco regulation. Using clustering analysis of the mixtures of VOCs, we identified two distinct clusters based on statistical determination of the optimal number of groups. The two clusters revealed a demographic divergence in tobacco product usage, inherent VOC exposure, and health risk factors. Specifically, Cluster 1, was defined by lower median levels of VOCs and included younger individuals who either have never used tobacco or have used e-cigarettes. Cluster 2 was defined by higher median levels of VOCs and included older individuals who use traditional cigarettes. We further examined the association between clusters and CVD risk factors. We found that participants in Cluster 2 were likely to have a higher heart rate, even after adjusting for demographic variables.

Although the association of VOCs with traditional combustible tobacco is well-documented, the levels and combinations emitted by new products like e-cigarettes are less understood^{4,3}. The generation of VOCs is frequently associated with combustion, as seen with combustible cigarettes. However, these compounds are found in numerous products unrelated to combustion. E-electronic cigarettes were introduced to the market as a reduced-harm product, mainly due to the elimination of combustion and the suggested health benefits of reducing hazardous and potentially hazardous compounds associated with combustion and tar. E-cigarettes eliminate combustion by utilizing a heating coil to vaporize nicotine-containing liquid, and though certain devices result in lower levels of VOCs^{3,24}, others can produce VOC levels comparable to combustible cigarettes²⁵. Our work suggests that the majority of our e-cigarette users had lower median VOC levels and tended to

Figure 2. Scatter plot illustrating the distinction of two clusters in two dimensions of space through principal component analysis. The first principal component (PC1) captures the largest variance, the second principal component (PC2) accounts for the second-largest variance. Cluster 1 (yellow dots) is tightly grouped and its VOCs characteristics are more homogeneous than those of Cluster 2

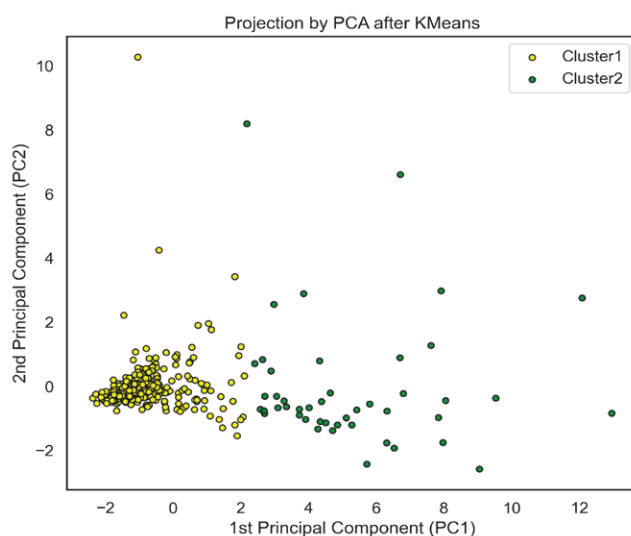


Table 2. Association analysis between two clusters and the health outcomes using a generalized linear model

Outcome	Cluster	Unadjusted β coefficients (95% CI)	Adjusted ^a β coefficients (95% CI)
Systolic blood pressure	Cluster 1 ®		
	Cluster 2	4.21 (0.34–8.09)*	-0.66 (-4.60–3.28)
Diastolic blood pressure	Cluster 1 ®		
	Cluster 2	3.70 (1.06–6.35)*	0.34 (-2.51–3.20)
Heart rate	Cluster 1 ®		
	Cluster 2	4.58 (1.36–7.79)*	3.29 (-0.26–6.84)

^a Adjusted for age, gender, and race/ethnicity. The significance level: $p < 0.05$.

*Significant. ® Reference category.

be closer aligned with a VOC profile of individuals with no tobacco exposure. For those e-cigarette users in Cluster 2, their use patterns and/or device may have exposed them to a greater level of VOCs thus grouping them into a cluster with a risk factor for CVD. Electronic cigarettes have changed rapidly since being introduced; thus, their exposure profile has changed over time and is likely to change in the future making these devices difficult to regulate as an overarching class. These findings suggest that individuals who use the same product class may have a different exposure profile and health risk, and thus the use of exposure patterns warrant consideration in future studies. From a regulatory standpoint, prioritizing the assessment of an individual's exposure to hazardous and potentially harmful chemicals, rather than grouping a highly diverse range of devices, could be a crucial regulatory strategy in navigating the dynamic landscape of this market as well as evaluating new and emerging tobacco products.

Moreover, substantial equivalence which is used to bring new tobacco products to market addresses tobacco product classes with similar design features but does not consider specific mixtures of exposure. Additionally, health effects comparisons for the substantial equivalence products are based on single chemical exposure, losing the complexity and variable nature of hazardous and potentially harmful chemicals that would vary between and within product classes. Consequently, when tobacco product groups alone are used to define groupings for health evaluations, the categorization does not consider the varied individual exposure within a product family like e-cigarettes. As demonstrated in this study, individuals who used the same type of product were separated into different clusters, which had different health outcomes, by their VOC exposure pattern. Given that many tobacco regulatory decisions are based on product classes, which results in mixed exposure to harmful constituents that differentially impact health, it is important to develop and explore more robust models of identifying health risks.

Mixed-exposure research can address the additive effects of chemicals, which could be synergistic or antagonistic, and is imperative to understand in health models. Users of tobacco and e-cigarettes are always exposed to a mixture of hazardous and potentially

harmful chemicals, but these types of exposures are exceedingly difficult to study from a statistical perspective²⁶.

Research suggests that when modeling mixtures of chemicals, additive and independent compound analyses are a popular methodology, but do not adequately reflect the models of exposure. Consequently, it cannot account for the antagonistic effects of certain chemicals, such as the antagonistic relationship between CO₂ and nicotine, seen with smoking. These models tend to underestimate the enhanced effects of mixtures on physiological responses²⁷. Our study used cluster analysis to explore mixtures of VOCs and health outcomes. Our cluster analysis generally separated the cohort based on several key factors including age and product of use. Other statistical approaches to modeling mixtures of exposure are available but have not been rigorously applied to tobacco science, leaving a gap in the literature for appropriate mixed-model methodology.

Our findings underscore the critical role of socioeconomic and demographic factors in shaping exposure to tobacco-related VOCs²⁸⁻³⁰. For example, younger populations, favoring e-cigarettes over traditional tobacco products, may show distinct VOC exposure profiles compared to older populations, as indicated by age-related usage patterns. Cluster 1 (e-cigarettes and non-users) also shows a more diverse racial composition, with the majority identifying as White, followed by a notable proportion of Asian individuals, and a smaller percentage as Black participants. Age and race are crucial considerations in tobacco regulatory science due to their demonstrated impact on nicotine metabolism³¹. Investigating demographic characteristics are also associated with alterations in metabolism of other harmful and potentially harmful constituents (HPHCs) such as VOCs is essential for understanding health effects in a diverse population. We also showed that Cluster 2 had a higher percentage of participants with high school or lower level of education. Lower education level is associated with reduced risk perceptions regarding tobacco use. This may result in reduced interactions with public health messaging and limited awareness of the increased exposure to hazardous and potentially harmful chemicals that may occur with different forms of tobacco products.

Strengths and limitations

This study possesses several strengths. The utilization of cluster analysis in this study enables the characterization of mixtures of toxic VOCs, rather than solely focusing on the products themselves. Our ability to group HPHCs into a mixture and assess the health effects is highly relevant to real-world exposures and could be an important tool to consider for future regulatory science. The clustering approach we used here offers a means to enhance regulatory knowledge regarding health effects based on mixed-exposure models. These models are likely to vary across different tobacco product classes and user profiles, as opposed to single compounds. Thus, this approach allows a more comprehensive evaluation of the concept of substantial equivalence for not only current products, but also future tobacco products trying to enter the market.

Our study design asked participants to use their preferred products, allowing evaluation of the most widely used products on the market. Study designs that provide a product are often limited because they are designed with products that are dated and may have lost popularity among users. This dynamic approach provided high-quality relevant data to the tobacco science field. Highly relevant exposures paired with an immediate assessment of CVD risk factors, such as heart rate and blood pressure, are highly generalizable. The health effects examined in our study are relevant to chronic disease and are reversible, allowing these biomarkers to reflect changes with exposure to new and emerging tobacco products, as well as transitions between them. These metrics provide a snapshot of potential long-term risks without the need for the lengthy follow-up required in prospective studies. Furthermore, given the potential for a prolonged latency period – often spanning ten years or more – between increased cardiovascular risk factors and an actual cardiac event, assessing these health endpoints enables the early detection of signs of cardiotoxicity, supporting timely decision-making regarding the health effects of new products.

While the cluster analysis offers an understanding of the potential risks for specific VOC mixtures, not just a product, there are limitations. The mixture of e-cigarette users and non-users into a single cluster could potentially mask the subtleties in health

changes between these two product use groups. The rapidly changing landscape of e-cigarette technology and formulations also suggests that these findings may be time-sensitive and subject to further evolution, though we did capture the exposure profiles of the most widely used products at the time of the study. Finally, self-reported data on tobacco product use could be subject to bias. While our study did not identify significant associations between VOC exposure clusters and cardiovascular risk factors, we saw trends that warrant further investigation. Longitudinal studies would be required to investigate the temporal relationships between VOC exposure and health outcomes, allowing for causal inference.

CONCLUSIONS

Our findings suggest that lower levels of mixtures of VOCs included both users of e-cigarettes and those who used no tobacco products. This indicates that e-cigarettes might be associated with a different VOC exposure profile compared to traditional cigarettes, but the inclusion of some e-cigarette users in Cluster 2 suggests a potential influence of individual use patterns on VOC levels, which needs further investigation. Future studies should consider longitudinal designs to track changes in VOC exposure over time, including more granular data collection methods to capture the frequency and intensity of tobacco product use. Such detailed analyses could provide clearer guidance for public health policy and regulatory measures aimed at reducing the harm from tobacco products.

REFERENCES

1. Li AJ, Pal VK, Kannan K. A review of environmental occurrence, toxicity, biotransformation and biomonitoring of volatile organic compounds. *Environmental Chemistry and Ecotoxicology*. 2021;3. doi:[10.1016/j.enceco.2021.01.001](https://doi.org/10.1016/j.enceco.2021.01.001)
2. Pazo DY, Moliere F, Sampson MM, et al. Mainstream smoke levels of volatile organic compounds in 50 U.S. domestic cigarette brands smoked with the ISO and Canadian intense protocols. *Nicotine Tob Res*. 2016;18(9):1886-1894. doi:[10.1093/ntr/ntw118](https://doi.org/10.1093/ntr/ntw118)
3. Keith RJ, Fetterman JL, Orimoloye OA, et al. Characterization of volatile organic compound metabolites in cigarette smokers, electronic nicotine device users, dual users, and nonusers of tobacco. *Nicotine Tob Res*. 2020;22(2):264-272. doi:[10.1093/ntr/ntz021](https://doi.org/10.1093/ntr/ntz021)
4. Lee MS, LeBouf RF, Son YS, Koutrakis P, Christiani DC.

- Nicotine, aerosol particles, carbonyls and volatile organic compounds in tobacco- and menthol-flavored e-cigarettes. *Environ Health*. 2017;16(1):42. doi:[10.1186/s12940-017-0249-x](https://doi.org/10.1186/s12940-017-0249-x)
5. Alford KL, Kumar N. Pulmonary health effects of indoor volatile organic compounds-a meta-analysis. *Int J Environ Res Public Health*. 2021;18(4):1578. doi:[10.3390/ijerph18041578](https://doi.org/10.3390/ijerph18041578)
 6. Sarigiannis DA, Karakitsios SP, Gotti A, Liakos IL, Katsoyiannis A. Exposure to major volatile organic compounds and carbonyls in European indoor environments and associated health risk. *Environ Int*. 2011;37(4):743-765. doi:[10.1016/j.envint.2011.01.005](https://doi.org/10.1016/j.envint.2011.01.005)
 7. Riggs DW, Malovichko MV, Gao H, et al. Environmental exposure to volatile organic compounds is associated with endothelial injury. *Toxicol Appl Pharmacol*. 2022;437:115877. doi:[10.1016/j.taap.2022.115877](https://doi.org/10.1016/j.taap.2022.115877)
 8. McGraw KE, Konkle SL, Riggs DW, et al. Exposure to volatile organic compounds is associated with hypertension in black adults: The Jackson Heart Study. *Environ Res*. 2023;223:115384. doi:[10.1016/j.envres.2023.115384](https://doi.org/10.1016/j.envres.2023.115384)
 9. Haussmann HJ. Use of hazard indices for a theoretical evaluation of cigarette smoke composition. *Chem Res Toxicol*. 2012;25(4):794-810. doi:[10.1021/tx200536w](https://doi.org/10.1021/tx200536w)
 10. Xu X, Freeman NC, Dailey AB, Ilacqua VA, Kearney GD, Talbott EO. Association between exposure to alkylbenzenes and cardiovascular disease among National Health and Nutrition Examination Survey (NHANES) participants. *Int J Occup Environ Health*. 2009;15(4):385-391. doi:[10.1179/oeht.2009.15.4.385](https://doi.org/10.1179/oeht.2009.15.4.385)
 11. Ran J, Qiu H, Sun S, Yang A, Tian L. Are ambient volatile organic compounds environmental stressors for heart failure? *Environ Pollut*. 2018;242(Pt B):1810-1816. doi:[10.1016/j.envpol.2018.07.086](https://doi.org/10.1016/j.envpol.2018.07.086)
 12. Demonstrating the Substantial Equivalence of a New Tobacco Product: Responses to Frequently Asked Questions (Revised). Food and Drug Administration. Center for Tobacco Products. October 18, 2023. Accessed January 17, 2025. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/demonstrating-substantial-equivalence-new-tobacco-product-responses-frequently-asked-questions>
 13. Stratton K, Shetty P, Wallace R, Bondurant S, eds. *Clearing the Smoke: Assessing the Science Base for Tobacco Harm Reduction*. U.S. National Academies Press; 2001. doi:[10.17226/10029](https://doi.org/10.17226/10029)
 14. Cobb NK, Abrams DB. The FDA, e-cigarettes, and the demise of combusted tobacco. *N Engl J Med*. 2014;371(16):1469-1471. doi:[10.1056/NEJMp1408448](https://doi.org/10.1056/NEJMp1408448)
 15. Wu G, Gong S, He Y, Liu D. Smoking is associated with elevated blood level of volatile organic compounds: a population-based analysis of NHANES 2017-2018. *Arch Public Health*. 2023;81(1):55. doi:[10.1186/s13690-023-01070-x](https://doi.org/10.1186/s13690-023-01070-x)
 16. National Academies of Sciences, Engineering, and Medicine. Health and Medicine Division. Board on Population Health and Public Health Practice. *Public Health Consequences of E-Cigarettes*. In: Eaton DL, Kwan LY, Stratton K, eds. U.S. National Academies Press; 2018. doi:[10.17226/24952](https://doi.org/10.17226/24952)
 17. Dinakar C, O'Connor GT. The health effects of electronic cigarettes. *N Engl J Med*. 2016;375(14):1372-1381. doi:[10.1056/NEJMr1502466](https://doi.org/10.1056/NEJMr1502466)
 18. Parsons VL, Moriarity C, Jonas K, Moore TF, Davis KE, Tompkins L. Design and estimation for the national health interview survey, 2006-2015. *Vital Health Stat* 2. 2014;(165):1-53.
 19. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208. doi:[10.1016/j.jbi.2019.103208](https://doi.org/10.1016/j.jbi.2019.103208)
 20. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381. doi:[10.1016/j.jbi.2008.08.010](https://doi.org/10.1016/j.jbi.2008.08.010)
 21. Keith RJ, Fetterman JL, Riggs DW, et al. Protocol to assess the impact of tobacco-induced volatile organic compounds on cardiovascular risk in a cross-sectional cohort: Cardiovascular Injury due to Tobacco Use study. *BMJ Open*. 2018;8(3):e019850. doi:[10.1136/bmjopen-2017-019850](https://doi.org/10.1136/bmjopen-2017-019850)
 22. Alwis KU, Blount BC, Britt AS, Patel D, Ashley DL. Simultaneous analysis of 28 urinary VOC metabolites using ultra high performance liquid chromatography coupled with electrospray ionization tandem mass spectrometry (UPLC-ESI/MSMS). *Anal Chim Acta*. 2012;750:152-160. doi:[10.1016/j.aca.2012.04.009](https://doi.org/10.1016/j.aca.2012.04.009)
 23. Lorkiewicz P, Riggs DW, Keith RJ, et al. Comparison of urinary biomarkers of exposure in humans using electronic cigarettes, combustible cigarettes, and smokeless tobacco. *Nicotine Tob Res*. 2019;21(9):1228-1238. doi:[10.1093/ntr/nty089](https://doi.org/10.1093/ntr/nty089)
 24. St Helen G, Liakoni E, Nardone N, Addo N, Jacob P 3rd, Benowitz NL. Comparison of systemic exposure to toxic and/or carcinogenic volatile organic compounds (VOC) during vaping, smoking, and abstinence. *Cancer Prev Res (Phila)*. 2020;13(2):153-162. doi:[10.1158/1940-6207.CAPR-19-0356](https://doi.org/10.1158/1940-6207.CAPR-19-0356)
 25. Blair SL, Epstein SA, Nizkorodov SA, Staimer N. A real-time fast-flow tube study of VOC and particulate emissions from electronic, potentially reduced-harm, conventional, and reference cigarettes. *Aerosol Sci Technol*. 2015;49(9):816-827. doi:[10.1080/02786826.2015.1076156](https://doi.org/10.1080/02786826.2015.1076156)
 26. Bart S, Short S, Jager T, et al. How to analyse and account for interactions in mixture toxicity with toxicokinetic-toxicodynamic models. *Sci Total Environ*. 2022;843:157048. doi:[10.1016/j.scitotenv.2022.157048](https://doi.org/10.1016/j.scitotenv.2022.157048)
 27. Cedergreen N. Quantifying synergy: a systematic review of mixture toxicity studies within environmental toxicology. *PLoS One*. 2014;9(5):e96580. doi:[10.1371/journal.pone.0096580](https://doi.org/10.1371/journal.pone.0096580)

28. Lim DH, Son YS, Kim YH, Kukkar D, Kim KH. Volatile organic compounds released in the mainstream smoke of flavor capsule cigarettes. *Environ Res.* 2022;209:112866. doi:[10.1016/j.envres.2022.112866](https://doi.org/10.1016/j.envres.2022.112866)
29. Smith DM, Shahab L, Blount BC, et al. Differences in exposure to nicotine, tobacco-specific nitrosamines, and volatile organic compounds among electronic cigarette users, tobacco smokers, and dual users from three countries. *Toxics.* 2020;8(4):88. doi:[10.3390/toxics8040088](https://doi.org/10.3390/toxics8040088)
30. Reilly SM, Goel R, Bitzer Z, et al. Effects of topography-related puff parameters on carbonyl delivery in mainstream cigarette smoke. *Chem Res Toxicol.* 2017;30(7):1463-1469. doi:[10.1021/acs.chemrestox.7b00104](https://doi.org/10.1021/acs.chemrestox.7b00104)
31. St Helen G, Benowitz NL, Ko J, et al. Differences in exposure to toxic and/or carcinogenic volatile organic compounds between Black and White cigarette smokers. *J Expo Sci Environ Epidemiol.* 2021;31(2):211-223. doi:[10.1038/s41370-019-0159-9](https://doi.org/10.1038/s41370-019-0159-9)

ACKNOWLEDGEMENTS

We thank all our participants of the CITU 2.0 study.

CONFLICTS OF INTEREST

The authors have each completed and submitted an ICMJE form for disclosure of potential conflicts of interest. The authors declare that they have no competing interests, financial or otherwise, related to the current work. J. Zhao and J.L. Hall report that since the initial planning of the work this study received funding U54 HL120163 and P50 HL120163. R.M. Robertson reports that since the initial planning of the work she received her salary from the American Heart Association (part-time employee) and that in the past 36 months there was a grant to her Institution from the NIH/CTP (U54 HL120163) American Heart Association Tobacco Center for Regulatory Science (A-TRAC 2.0) and that she received support for attending meetings and/or travel only from her employer and for work. Finally, she declares that in the past 36 months she was at the service for the RWJF Harold Amos Faculty Development Award as member of the National Advisory Committee. J.L. Hall reports that in the past 36 months she received grants P50 HL120163 and U54 HL120163. N. Hamburg and R.J. Keith report that since the initial planning of the work their Institutions received grants from the National Institutes of Health and the American Heart Association.

FUNDING

Research reported in this publication was supported in part by the National Heart, Lung, and Blood Institute of the National Institutes of Health and the Food and Drug Administration's Center for Tobacco Products (CTP) under award numbers P50 HL120163 and U54 HL120163. The content is solely the responsibility of the authors and does not necessarily represent official views of the National Institutes of Health, the Food and Drug Administration, or the American Heart Association.

ETHICAL APPROVAL AND INFORMED CONSENT

Ethical approval was obtained from the Institutional Review Boards of Boston University (Approval number: H-32613) and the University of Louisville (Approval number: 13.0590). Participants provided informed consent.

DATA AVAILABILITY

The data supporting this research are available from the authors on reasonable request.

AUTHORS' CONTRIBUTIONS

JZ and RJK: drafting of the manuscript. JZ, HH, JZhai, RP and JLH: statistical analysis. JH, HH, JZhai, SS, ACS, PKL, TJ, JLH, NMH, and RJK: reviewing of the manuscript. SS, RMR, AB, JLH, NMH and RJK: funding. SS, NMH, RJK and PKL: data acquisition. NMH and RJK: study design. All authors read and approved the final version of the manuscript.

PROVENANCE AND PEER REVIEW

Not commissioned; externally peer reviewed.