Published in final edited form as:

Prev Med. 2020 November ; 140: 106218. doi:10.1016/j.ypmed.2020.106218.

Cluster Analysis of Urinary Tobacco Biomarkers among U.S. Adults: Population Assessment of Tobacco and Health (PATH) Biomarker Study (2013-2014)

Ban Majeed, PhD¹, Daniel Linder, PhD¹, Thomas Eissenberg, PhD², Yelena N. Tarasenko, DrPH³, Danielle Smith, MPH⁴, David L Ashley, PhD⁵

¹Department of Population Health Sciences, Medical College of Georgia, Augusta University, Augusta, GA

²Center for the Study of Tobacco Products, Department of Psychology, Virginia Commonwealth University, Richmond, VA

³Division of Biostatistics, Epidemiology and Environmental Health Sciences, Jiann-Ping Hsu College of Public Health, Georgia Southern University, Statesboro, GA

⁴Department of Health Behavior, Roswell Park Comprehensive Cancer Center, Buffalo, NY

⁵Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, GA

Abstract

Tobacco use delivers nicotine, tobacco-specific nitrosamines (TSNAs), volatile organic compounds (VOCs), and polycyclic aromatic hydrocarbons (PAHs), which are metabolized and excreted in urine offering useful biomarkers of exposure. Previous studies compared individual toxicants across tobacco users. Based on a group of biomarkers, cluster analysis was used to define tobacco toxicant exposure profiles. Clusters with distinct exposure profiles, were determined and described, based on levels of urinary biomarkers of exposure to nicotine, TSNAs, VOCs, and PAHs among a national sample of current, established, adult tobacco users, and examine the association of use behavior and cluster membership. The PATH Biomarker Wave 1 data were analyzed. Current established tobacco users with complete urinary biomarker data were included (N=6,724). User groups included cigarette smokers, users of electronic cigarette (ECIG), smokeless tobacco (SLT), and dual and poly tobacco users. Cluster analysis, pairwise comparisons, and multinomial logistic regression were conducted. Cigarette smokers were primarily in clusters with high biomarker concentrations across all groups, but actual concentrations were associated with smoking quantity. A cluster with high TSNAs but low levels of PAHs and VOCs was heavily populated by SLT users. Exclusive ECIG users, depending on use

Corresponding author: Ban Majeed, Assistant Professor, Division of Epidemiology, Department of Population Health Sciences, Medical College of Georgia, Augusta University. Physical Address: 1120 15th Street, AE-1036, Pavilion I, Augusta, GA 30912-4900. Phone: 706-721-0806. Fax: 706-721-6294. bmajeed@augusta.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

frequency, were predominantly in clusters with low biomarker concentrations, except for one cluster that had relatively high TSNAs. Clusters heavily populated by dual and poly tobacco users were the same as those heavily populated by cigarette smokers. Ten exposure profiles (clusters) were determined and linked to tobacco use behavior. Findings could inform future research and policy initiatives.

Keywords

cluster analysis; urinary biomarkers; exposure profile; smoking; smokeless tobacco; ECIG; dual; poly use

1. Introduction

Tobacco use, particularly cigarette smoking, is the leading cause of preventable death and disease in the United States (USDHHS 2014, Gallaway et al, 2018). The morbidity and mortality caused by cigarette smoking is attributed to long-term exposure to thousands of toxicants in tobacco smoke (USDHHS 2010). In addition to nicotine, smokers are exposed to carcinogenic tobacco-specific nitrosamines (TSNAs), volatile organic compounds (VOCs), and polycyclic aromatic hydrocarbons (PAHs). Following absorption and distribution, these toxicants are excreted in urine, unchanged or as metabolites, offering useful biomarkers of exposure (USDHHS 2010).

This study focuses on a set of tobacco-related exposure biomarkers because they distinguish between tobacco users from non-users; have a dose-response relationship with amount of product used; respond to changes in use behavior and product characteristics; and represent harmful and potentially harmful constituents (HPHC) (Chang et al, 2017). For example, all tobacco users are exposed to TSNAs, such as N'-nitrosonomotine (NNN), which causes oral cavity cancer, and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), which causes lung, pancreatic, oral, prostate, and cervical cancers (USDHHS 2010). Though not specific to tobacco because they could be environmental or occupational in origin, VOCs and PAHs classified as HPHC have been linked to both malignant and non-malignant lung diseases, such as COPD, emphysema, chronic bronchitis and asthma (USDHHS 2010).

Tobacco toxicant exposure varies depending on type and number of products used (Goniewicz et al, 2017, Nollen et al, 2017, Wang et al, 2019). For example, in some studies, exclusive use of electronic nicotine delivery systems (electronic cigarettes, ECIGs) was associated with lower levels of exposure to toxicants including TSNAs, VOCs, and PAHs, relative to exclusive combustible cigarette smoking (Shahab et al, 2017, Goniewicz et al, 2018). Similar findings were found over time when cigarette smokers completely switched to exclusive ECIG use (Goniewicz et al, 2017). Urinary TSNA biomarker concentrations were much lower among nicotine replacement therapy (NRT)-only and ECIG-only users compared to current cigarette smokers (Shahab et al, 2017). But, compared to cigarette smokers, users of smokeless tobacco products (SLT) were observed to have higher concentrations of urinary biomarkers of nicotine and TSNAs and lower concentrations of PAH and VOC biomarkers (Cheng et al, 2020). Additionally, those who either smoked cigarettes-only or cigars-only exhibited similar levels of tobacco toxicant exposure (Chang

et al, 2019). VOC biomarkers followed a similar pattern and were markedly lower in NRT-only and ECIG-only users than in cigarette smokers (Shahab et al, 2017). Furthermore, biomarkers of PAHs were detected in higher concentrations in users of combusted tobacco products than in users of non-combusted products (e.g. ECIG and/or SLT) (Wang et al, 2019). In sum, urinary biomarker concentrations differ substantially by product type and frequency of use.

Previous studies evaluating the impact of tobacco product use on biomarkers of exposure used conventional techniques (Nollen et al, 2017, Chang et al, 2019, Wang et al, 2019), such as pairwise t-tests and analysis of variance, to examine group differences in geometric means of individual tobacco biomarkers. These conventional techniques are useful in describing how a particular biomarker may vary across groups. Yet, results based on these techniques may not yield a comprehensive view of tobacco toxicant exposure. Thus, in the current study, we used cluster analysis, a technique that defines exposure profiles based on concentrations of biomarkers as a group or a mixture to provide a more data-driven depiction of tobacco toxicant exposure. Cluster analysis has not been applied to classify tobacco users according to biomarker exposure, though it has proven useful in other fields, led to new insights about disease pathophysiology, and provided new approach to group patients who shared similar clinical outcomes (van Bochove et al, 2012, Ahmad et al, 2016, Guo et al, 2017). Similarly, current study results, based on levels of well-established tobacco biomarkers, can inform future research and policy initiatives. In terms of research, using data-driven statistical methodology to cluster tobacco users presents a useful and efficient method to generate hypotheses about the relationship between cluster specific toxicant profile and product characteristics (e.g. flavor and nicotine content) and adverse health outcomes. Given the important role of tobacco biomarker research across the Food and Drug Administration (FDA) regulatory authorities (Chang et al, 2017), resultant observations could inform development and evaluation of product standards. For example, cluster specific toxicant profile and its association with product type, could serve as reference (baseline) information when assessing the impact of product changes (i.e. reduction or elimination of additives or constituents or other components) on toxicant exposure and in turn on health risks of the tobacco product.

This study aims to determine and describe groups (clusters) with distinct exposure profiles based on levels of urinary biomarkers of exposure to nicotine, TSNAs, VOCs, and PAHs among a national sample of current, established, adult tobacco users. Furthermore, the association of product use behavior with cluster membership will be examined.

2. Methods

Study Sample

Data are from the Restricted-Use Files of the Population Assessment of Tobacco and Health (PATH) Biomarker (Core) Wave 1 (2013-2014), a nationally representative longitudinal cohort study launched to assess tobacco use and health outcomes among the non-institutionalized adult civilian population (Hyland et al, 2017). PATH uses a four-stage stratified sampling method to sample over 150,000 mailing addresses, yielding at baseline a cohort of 32,320 adult participants with a response rate of 74%. A probability subsample of

the baseline adult cohort (N=11,522) from the Wave 1 PATH Biomarker Core provided urine specimens that were analyzed for tobacco biomarkers. The rationale for using Wave 1 biomarker data, instead of a more recent (Wave 2 or 3) is that only Wave 1 biomarker study data were available for analysis at the time that this study was designed and implemented. This secondary data analysis study consisted of adults who had completed the adult interview study and reported current established tobacco use, had urine creatinine values within the (normal) reference range (10—370 mg/dL)(Aylward et al, 2014, Goniewicz et al, 2018), and had complete data on tobacco use and a set of urinary biomarkers of exposure to eight chemicals representative of major classes of harmful constituents in tobacco products: 1) TSNAs, namely, NNAL and N'-nitrosonomicotine (NNN); 2) PAHs, metabolites of naphthalene, 1-naphthol (1-NAP), and fluorene, 3-hydroxyfluorene (3-FLU); 3) VOCs, metabolites of acrolein, N-acetyl-S-(2-carboxyethyl)-L-cysteine (CEMA), acrylonitrile, N-acetyl-S-(2-cyanoethyl)-L-cysteine (CYMA), and 1,3-butadiene, N-acetyl-S-(4-hydroxy-2-buten-1-yl)-L-cysteine (MHB3); and 4) nicotine, as measured by total nicotine equivalents (TNE2), based on the molar sum of cotinine and trans-3'-hydroxycotinine. TNE2 was set to missing if either of these analytes was missing. After excluding adults with incomplete information on tobacco use, biomarkers of interest and those with out-of-range creatinine values, the resulting analytic sample size was 6,724 participants. To control for potential variations based on renal function, urinary biomarkers were normalized for creatinine by dividing the biomarker value by the urinary creatinine value. Details on all measured urinary biomarkers, biospecimen collection, and laboratory procedures can be found elsewhere (Goniewicz et al, 2018, USDHHS 2019). Augusta University IRB granted this project a non-human subject designation.

Measures

Tobacco User Type—Based on self-reported tobacco use behavior, study participants were categorized into these mutually exclusive groups: exclusive combustible cigarette smokers; exclusive ECIG users; exclusive SLT (e.g. loose snus, moist snuff, dip, spit, or chewing tobacco, snus, and dissolvable tobacco) users; dual (cigarette and ECIG) users; and poly tobacco users, users of any combination of two or more tobacco products except dual use of cigarettes and ECIGs. Few individuals were exclusive users of traditional cigars, cigarillos, filtered cigars, pipe tobacco, and hookah, so these users were combined into other combustible tobacco (OCT).

Current established cigarette smoker was defined as having ever smoked a cigarette, smoked more than 100 cigarettes in lifetime, and currently smokes every day or some days. Current established users of ECIGs and SLT were defined as individuals who reported ever, fairly regular, and current (every day or some days) use of the particular tobacco product.

Cigarette smoking quantity (SQ) was determined for exclusive cigarette smokers, dual, and poly users who also smoked cigarettes. SQ was computed by multiplying 30 by the number of cigarettes smoked per day (CPD) for daily smokers or multiplying the number of days smoked in past 30 days by CPD for non-daily smokers. Means for number of days used ECIGs or SLT in past 30 days were estimated using equivalent variables from PATH.

Recency of tobacco use was categorized into use ‘in past hour’, ‘sometime today’, and ‘yesterday or earlier’.

Demographic Characteristics—Self-reported demographic characteristics included: sex; age (18-24, 25-34, 35-54, and 55 and above); race/ethnicity (non-Hispanic White, non-Hispanic Black, Other, non-Hispanic (including Asian and multiracial), and Hispanic; education; and region.

Statistical Analysis

Cluster analysis was performed using the R package **mclust** (Scrucca et al, 2016). This clustering approach, based on the finite mixture of multivariate Gaussian formulation (Fraley et al, 2002), offers several advantages to competing methodologies that analyze individual biomarkers. Namely, it allows for flexible within cluster covariance structures and is likelihood based, thus, the classical fit indices may be used to select optimal models. The Bayesian information criterion (BIC) to select the final model (optimal number of clusters) was used, following stepwise process: (i) fit a single cluster model under all 14 possible variance structures allowed by the model and chose the best model from these 14 models using BIC, (ii) increase the number of clusters by 1 (i.e., a two-cluster model) and fit all 14 possible variance structures, choosing the best model from these 14 via BIC. We stopped this process of adding one new cluster to the model when there was no further improvement in BIC.

Descriptive and regression analyses were performed using SVY procedures in Stata 15.1 (Stata Corp LLC, College Station, Texas). Geometric means (and 95% Confidence Intervals, CI) of creatinine-adjusted biomarker concentrations per cluster were computed estimated. We conducted multinomial logistic regression analysis, where cluster was the dependent variable and tobacco user type was the independent variable, controlling for demographic characteristics. Predicted probabilities of tobacco user types per cluster were estimated using the Stata command **margins**. For continuous variables other than biomarker concentrations, we present means, 95% CI, and pairwise comparisons (comparing means across clusters). The significance level was set at 0.05 for all analyses. We employed the appropriate urinary biomarker weighting variables to adjust for the PATH Urinary Biomarker study design and non-response and used balanced repeated replication (BRR) with Fay’s adjustment set to 0.3 for variance estimation (Hyland et al, 2017, USDHHS 2018).

To describe toxicant exposure profile, we ranked biomarker means among clusters, using a scale from 1-10; 1: highest value and 10: lowest value among all clusters. This ranking approach provided the basis for cluster labeling and description.

3. Results

Participants

Participant characteristics are presented in Table 1. Of the 6,724 tobacco users, 57.4% were males, most aged 35-54 years (37.1%, 95% CI 34.9%–39.3%), and 68.7% (66.2%–71.1%) were non-Hispanic White. Overall, 63.4% were exclusive cigarette smokers, of whom 82.6% (80.1%–84.8%) smoked daily; 3.0% were exclusive ECIG users, of whom 44.2% (39.9%–

48.6%) used ECIGs daily; 5.4% were dual users and the majority of them 78.5% (73.3%–83.0%) smoked cigarettes daily. Lastly, poly users constituted 15.4%, of whom 84.0% (80.9%–86.7%) smoked cigarettes.

Cluster description

The aforementioned stepwise process resulted in 11 clusters, since no improvement in the BIC was achieved once the 12 cluster models were considered (Supplemental Table 1). Because of the small size of one of the clusters ($n=33$), tobacco users in this cluster were assigned to the cluster with the nearest ‘centroid,’ the most representative point within the cluster. Hence, the reported results are from ten clusters. Geometric means of the creatinine-normalized biomarkers per cluster are shown in Table 2. Ranking of biomarker means among clusters, indicated using a scale from 1-10; 1: highest value and 10: lowest value among all clusters, provided the basis for labeling the clusters. Table 3 shows cluster number (1-10), label, size based on percentage of study participants within each cluster, description, and ranking of biomarker means. The largest—with 29.4% of tobacco users—cluster 7 was named ‘very high’ because it contained biomarker means that ranked as the second or third highest among all clusters. The second largest, cluster 3 (~20% of tobacco users) contained biomarkers toward the middle for all concentrations with rankings ranging from 5 to 6, hence the name ‘middle.’ Cluster 1, third largest (12.3%) was called ‘extremely high’ because it included the highest or second highest means of biomarker concentrations among all clusters. In Table 3, the clusters are listed in order of cluster size, i.e. the percentage of study participants.

Cluster membership

Controlling for demographic characteristics, cluster membership varied significantly by tobacco user type. All comparisons are made in reference to cluster 7. Females were significantly less likely to be in clusters 3 and 5 and more likely to be in clusters 1 and 9 than males. Relative to those aged 18-24, adults aged 35-54 and 55+ years were significantly more likely to be in clusters 1 and 9, and significantly less likely to be in clusters 2, 3, 4, 6, and 8. Relative to non-Hispanic Whites, Non-Hispanic Black tobacco users were significantly more likely to be in clusters 2, 3, 4, and 6 (Supplemental Table 2).

Table 4 depicts the predicted probabilities of user type by cluster based on the adjusted multinomial logistic regression model. Of exclusive cigarette smokers, 33.2% were in cluster 7, 24.2% were in cluster 3, 12.8% were in cluster 1, and 11.3% were in cluster 2. Of exclusive ECIG users, 30.6% were in cluster 5, 15.3% in cluster 6, 12.3% in cluster 3, and 11.6% in cluster 8. Of exclusive SLT users, 44.5% were in cluster 5, and 17.4% were in the cluster 10. Of dual users, 33.9% were in cluster 7, 24.7% in cluster 3, and 13.1% in cluster 1. Lastly, of poly users, 32.1% were in cluster 7, 16.4% in cluster 1, 15.8% in cluster 3, and 11.1% in cluster 2.

Use frequency and quantity

Among exclusive cigarette smokers, mean SQ varied significantly by cluster (Table 5). Mean SQ in cluster 1 was significantly higher than that in clusters 2, 3, and 7. Similar differences were observed in means of CPD by cluster (Supplemental Table 3). As shown in Table 6,

mean number of days of ECIG use in past 30 days was significantly lower in cluster 8 than in clusters 3, 5, and 6. Pertaining to SLT use, there was no significant difference between clusters in number of days used SLT in past 30 days.

Time of last cigarette smoked before sample collection varied significantly by cluster (Supplemental Table 4). Of cigarette smokers in cluster 1, 87.8% smoked ‘in the past hour’ (vs 69.7% in cluster 2, and 64.8% in cluster 3). Use of SLT ‘in the past hour’ was reported by 68.5% (62.0%–74.3%) of SLT users in cluster 5, and 76% (65.0%–85.1%) of users in cluster 10.

4. Discussion

Based on concentrations of a set of urinary tobacco biomarkers, current established tobacco users were grouped in ten clusters, with a distinct exposure profile. Cluster membership was associated with type and frequency of tobacco product used, indicating differential toxicant exposure profile as a function of use behavior. Exclusive cigarette smokers were primarily found in four clusters: *extremely high*, *very high*, *PAH&VOC high*, and *middle*. This variation was mainly driven by differences in smoking quantity. The observation that SQ means were higher in the first two clusters than in the latter ones explains the difference in biomarker concentrations in these clusters. Means of the nicotine biomarker (TNE2) in *PAH&VOC high* and *middle* clusters were similar to those previously reported in cigarette smokers (Goniewicz et al, 2018). In contrast, the observed levels of TNE2 in *extremely high* and *very high* clusters were two- or three-fold higher than those reported in prior research (Goniewicz et al, 2018). Given that cotinine concentration correlates with CPD (Chang et al, 2017), the lower CPD in *PAH&VOC high* and *middle* clusters compared to *extremely high* and *very high* clusters helps explain the variation in TNE2 in these clusters. In this study, 24% of exclusive cigarette smokers were in *middle* cluster, which was marked by concentrations consistent with the biomonitoring levels in smokers from US National Health and Nutrition Examination Survey (CDC 2019), suggesting that on average, biomarker concentrations in smokers have similar pattern to that in *middle* cluster. Other factors, than quantity of use, may have also contributed to variation in exposure profiles, such as cigarette design (filter length and ventilation), tobacco type (burley, bright, or oriental), smoking topography, and metabolism (Ashley et al, 2010, USDHHS 2010, Rostron et al, 2019). Taken together, these findings indicate that cigarette smokers are a heterogeneous group in terms of levels of exposure to tobacco constituents, depending on intensity of use.

Exclusive ECIG users, were found predominantly in four clusters: *TSNA&TNE2 high/PAH&VOC low*, *very low*, *middle*, and *extremely low* clusters. The finding that ECIG users in the *extremely low* cluster used ECIGs on fewer days than those in other clusters provides evidence that exposure profile in the former, is associated with sporadic use. The presence of ECIG users in *TSNA&TNE2 high/PAH&VOC low* cluster was surprising, because ECIG use has been associated with minimal exposure to TSNA (Konstantinou et al, 2018). However, this high level of TSNA could have resulted from impurities of ECIG or liquid used. Furthermore, given that amount of nicotine delivered is a function of ECIG design (Voos et al, 2019), the observed high level of TNE2 in *TSNA&TNE2 high/PAH&VOC low* cluster maybe attributed to high nicotine content and more advanced/efficient design. When

interpreting these results, one should keep in mind the most commonly used brands at the time of data collection (e.g. blu cigs, NJOY, logic, and eSmoke) and that these results may not represent most modern types of ECIGs, such as JUUL. In sum, results indicate that ECIG users differ in terms of toxicant exposure and considerations should be given to product characteristics and use behavior when assessing exposure.

TSNA&TNE2 high/PAH& VOC low cluster depicts exposure profile typical of that associated with SLT use: high TSNA and nicotine, and low combustion-byproducts (Cheng et al, 2020). SLT users were also found in *TSNA&TNE2 extremely high* cluster, where levels of biomarkers of TSNA and nicotine were the highest among all clusters. Such elevated levels may have resulted from heavy use (Etemadi et al, 2019). However, in this study, there was no significant difference between clusters in number of days used SLT in past 30 days. Other reasons may explain these differences, include flavor (Kaur et al, 2018), brand and tobacco type (Ashley et al, 2010), and use topography (Muhammad-Kah et al, 2011).

Similar to exclusive cigarette smokers, dual and poly tobacco users were mainly observed in *extremely high*, *PAH& VOC high*, *very high*, and *middle* clusters. The SQ in dual and poly users significantly varied by cluster indicating that differences in SQ was associated with different exposure profiles. Furthermore, the same clusters which contained the largest percentages of dual and poly users also contained the largest percentages of exclusive cigarette smokers. Table 5 also reveals that, in these clusters, dual and poly users smoked a similar number of cigarettes as was found with exclusive cigarette smokers. This suggests that the exposure profiles in these clusters were mainly driven by cigarette smoking, confirming the notion that dual and poly use may not always reflect lower toxicant exposure or reduced frequency of use of the individual product (Choi et al, 2017).

Two clusters merit further examining: *High 1-NAP* cluster, as reflected by its label, this cluster had the highest concentration of 1-NAP, but not of 3-FLU, pointing to a non-tobacco source of exposure, e.g. carbaryl pesticide (Maroni et al, 2000, Li et al, 2010). *Extremely low* cluster had similar concentrations to those reported previously in non-tobacco users (Goniewicz et al, 2018, CDC 2019). These low levels may be a function of prolonged lag time between exposure and sampling, low number of cigarettes smoked, or product characteristic.

This study has limitations. First, because biomarker concentrations are most highly associated with recent use, they may not accurately reflect the self-reported tobacco user type. Yet, it is likely that the recent product used was the same as reported in users of one type of product. Second, the implemented clustering algorithm yielded the described clusters based on tobacco users with complete data only, thus, the results may not generalize to other tobacco users. Third, the findings among ECIG users are based on device types and characteristics prevalent during 2013-2014 and do not necessarily reflect the rapidly evolving market of ECIG that is currently dominated by a USB flash drive-like ECIGs, JUUL (Hsu et al, 2018, King et al, 2018).

These limitations notwithstanding, this study, employing a data-driven segmentation method, cluster analysis, presents exposure profiles of different subgroups of tobacco users. These descriptive profiles, incorporating observations of previous studies (Choi et al, 2017, Wang et al, 2019, Cheng et al, 2020) provide novel views and approximate comprehensive representation of the complex realities of tobacco toxicant exposure. Exclusive cigarette smokers were primarily found in clusters with high biomarker concentrations across all four groups, but actual concentrations were associated with smoking quantity. A cluster that contained high TSNAs but low levels of PAHs and VOCs was heavily populated by SLT users. Exclusive ECIG users, depending on use frequency, were predominantly in clusters with low biomarker concentrations, except for one cluster that had relatively high TSNAs. The clusters that were heavily populated by dual and poly tobacco users were the same as those heavily populated by exclusive cigarette smokers.

These observations, based on levels of well-established tobacco biomarkers, can inform future research and policy initiatives. The findings highlight the heterogeneity that exists within tobacco use, including those who share the same product type, raising the question as to whether risk associated with a tobacco product can be appropriately assessed without taking into consideration the user and product characteristics. The results also raise the possibility that the data driven approach of using cluster analysis to classify tobacco users in terms of their toxicant exposure, and in turn risk, might generate hypotheses worth testing regarding impact of product characteristics on toxicant exposure among users of the same product (ECIG, or cigarettes) and the same intensity of use.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: Dr. Eissenberg's effort is supported by the National Institute on Drug Abuse of the National Institutes of Health under Award Number U54DA036105 and the Center for Tobacco Products of the U.S. Food and Drug Administration. The content is solely the responsibility of the authors and does not necessarily represent the views of the NIH or the FDA.

Conflict of interest statement:

Dr. Eissenberg is a paid consultant in litigation against the tobacco industry and also the electronic cigarette industry and is named on a patent for a device that measures the puffing behavior of electronic cigarette users. David L. Ashley has received funds for work done for the World Health Organization Tobacco Free Initiative, as a Special Government Employee of the U.S. Food and Drug Administration, as a consultant for Pfizer, and as an independent contractor for McKing Consulting.

References

- Ahmad T, Desai N, Wilson F, Schulte P, Dunning A, Jacoby D, Allen L, Fiuzat M, Rogers J, Felker GM, O'Connor C and Patel CB (2016). "Clinical Implications of Cluster Analysis-Based Classification of Acute Decompensated Heart Failure and Correlation with Bedside Hemodynamic Profiles." *PLoS One* 11(2): e0145881 10.1371/journal.pone.0145881 [PubMed: 26840410]
- Ashley DL, O'Connor RJ, Bemert JT, Watson CH, Polzin GM, Jain RB, Hammond D, Hatsukami DK, Giovino GA, Cummings KM, McNeill A, Shahab L, King B, Fong GT, Zhang L, Xia Y, Yan X and McCraw JM (2010). "Effect of differing levels of tobacco-specific nitrosamines in cigarette smoke

on the levels of biomarkers in smokers.” *Cancer Epidemiol Biomarkers Prev* 19(6): 1389–1398. DOI:10.1158/1055-9965.EPI-10-0084 [PubMed: 20501750]

Aylward LL, Hays SM, Smolders R, Koch HM, Cocker J, Jones K, Warren N, Levy L and Bevan R (2014). “Sources of variability in biomarker concentrations.” *J Toxicol Environ Health B Crit Rev* 17(1): 45–61. doi: 10.1080/10937404.2013.864250 [PubMed: 24597909]

Centers for Disease Control and Prevention (CDC). 2019 Fourth National Report on Human Exposure to Environmental Chemicals Updated Tables. Available: <https://www.cdc.gov/exposurereport/index.html> (accessed 3.30.20).

Chang CM, Rostron BL, Chang JT, Corey CG, Kimmel HL, Sosnoff CS, Goniewicz ML, Edwards KC, Hatsukami DK, Wang Y, Del Valle-Pinero AY, Yang M, Travers MJ, Amstein S, Taylor K, Conway K, Ambrose BK, Borek N, Hyland A, Wang L, Blount BC and van Bommel DM (2019). “Biomarkers of Exposure among U.S. Adult Cigar Smokers: Population Assessment of Tobacco and Health (PATH) Study Wave 1 (2013-2014).” *Cancer Epidemiol Biomarkers Prev* 28(5): 943–953. doi: 10.1158/1055-9965.EPI-18-0539 [PubMed: 30733305]

Chang CM, Edwards SH, Arab A, Del Valle-Pinero AY, Yang L and Hatsukami DK (2017). “Biomarkers of Tobacco Exposure: Summary of an FDA-Sponsored Public Workshop.” *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 26(3): 291–302.

Cheng YC, Reyes-Guzman CM, Christensen CH, Rostron BL, Edwards KC, Wang L, Feng J, Jarrett JM, Ward CD, Xia B, Kimmel HL, Conway K, Leggett C, Taylor K, Lawrence C, Niaura R, Travers MJ, Hyland A, Hecht SS, Hatsukami DK, Goniewicz ML, Borek N, Blount BC and van Bommel DM (2020). “Biomarkers of Exposure among Adult Smokeless Tobacco Users in the Population Assessment of Tobacco and Health Study (Wave 1, 2013-14).” *Cancer Epidemiol Biomarkers Prev*. DOI: 10.1158/1055-9965.EPI-19-0766

Choi K, Sabado M, El-Toukhy S, Vogtmann E, Freedman ND and Hatsukami D (2017). “Tobacco Product Use Patterns, and Nicotine and Tobacco-Specific Nitrosamine Exposure: NHANES 1999-2012.” *Cancer Epidemiol Biomarkers Prev* 26(10): 1525–1530. doi: 10.1158/1055-9965.EPI-17-0338. [PubMed: 28710077]

Etemadi A, Poustchi H, Chang CM, Blount BC, Calafat AM, Wang L, De Jesus VR, Pourshams A, Shakeri R, Shiels MS, Inoue-Choi M, Ambrose BK, Christensen CH, Wang B, Murphy G, Ye X, Bhandari D, Feng J, Xia B, Sosnoff CS, Kamangar F, Brennan P, Boffetta P, Dawsey SM, Abnet CC, Malekzadeh R and Freedman ND (2019). “Urinary Biomarkers of Carcinogenic Exposure among Cigarette, Waterpipe, and Smokeless Tobacco Users and Never Users of Tobacco in the Golestan Cohort Study.” *Cancer Epidemiol Biomarkers Prev* 28(2): 337–347. doi: 10.1158/1055-9965.EPI-18-0743 [PubMed: 30622099]

Fraley C and Raftery AE (2002). “Model-Based Clustering, Discriminant Analysis, and Density Estimation.” *Journal of the American Statistical Association* 97(458): 611–631. 10.1198/016214502760047131

Gallaway MS, Henley SJ, Steele CB, Momin B, Thomas CC, Jamal A, Trivers KF, Singh SD and Stewart SL (2018). “Surveillance for Cancers Associated with Tobacco Use -United States, 2010-2014.” *MMWR Surveill Summ* 67(12): 1–42. doi: 10.15585/mmwr.ss6712a1.

Goniewicz ML, Smith DM, Edwards KC, Blount BC, Caldwell KL, Feng J, Wang L, Christensen C, Ambrose B, Borek N, van Bommel D, Konkel K, Erives G, Stanton CA, Lambert E, Kimmel HL, Hatsukami D, Hecht SS, Niaura RS, Travers M, Lawrence C and Hyland AJ (2018). “Comparison of Nicotine and Toxicant Exposure in Users of Electronic Cigarettes and Combustible Cigarettes.” *JAMA Network Open* 1(8): e185937–e185937. doi:10.1001/jamanetworkopen.2018.5937 [PubMed: 30646298]

Goniewicz ML, Gawron M, Smith DM, Peng M, Jacob P 3rd and Benowitz NL (2017). “Exposure to Nicotine and Selected Toxicants in Cigarette Smokers Who Switched to Electronic Cigarettes: A Longitudinal Within-Subjects Observational Study.” *Nicotine Tob Res* 19(2): 160–167. doi: 10.1093/ntr/ntw160 [PubMed: 27613896]

- Guo Q, Lu X, Gao Y, Zhang J, Yan B, Su D, Song A, Zhao X and Wang G (2017). "Cluster analysis: a new approach for identification of underlying risk factors for coronary artery disease in essential hypertensive patients." *Sci Rep* 7: 43965. doi: 10.1038/srep43965 [PubMed: 28266630]
- Hsu G, Sun JY and Zhu S-H (2018). "Evolution of Electronic Cigarette Brands From 2013-2014 to 2016-2017: Analysis of Brand Websites." *Journal of medical Internet research* 20(3): e80–e80. [PubMed: 29530840]
- Hyland A, Ambrose BK, Conway KP, Borek N, Lambert E, Carusi C, Taylor K, Crosse S, Fong GT, Cummings KM, Abrams D, Pierce JP, Sargent J, Messer K, Bansal-Travers M, Niaura R, Vallone D, Hammond D, Hilmi N, Kwan J, Piesse A, Kalton G, Lohr S, Pharris-Ciurej N, Castleman V, Green VR, Tessman G, Kaufman A, Lawrence C, van Bommel DM, Kimmel HL, Blount B, Yang L, O'Brien B, Tworek C, Alberding D, Hull LC, Cheng YC, Maklan D, Backinger CL and Compton WM (2017). "Design and methods of the Population Assessment of Tobacco and Health (PATH) Study." *Tob Control* 26(4): 371–378. doi: 10.1136/tobaccocontrol-2016-052934 [PubMed: 27507901]
- Kaur G, Muthumalage T and Rahman I (2018). "Mechanisms of toxicity and biomarkers of flavoring and flavor enhancing chemicals in emerging tobacco and non-tobacco products." *Toxicol Lett* 288: 143–155. doi: 10.1016/j.toxlet.2018.02.025 [PubMed: 29481849]
- King BA, Gammon DG, Marynak KL and Rogers T (2018). "Electronic Cigarette Sales in the United States, 2013-2017." *JAMA* 320(13): 1379–1380. [PubMed: 30285167]
- Konstantinou E, Fotopoulou F, Drosos A, Dimakopoulou N, Zagoriti Z, Niarchos A, Makrynioti D, Kouretas D, Farsalinos K, Lagoumintzis G and Poulas K (2018). "Tobacco-specific nitrosamines: A literature review." *Food Chem Toxicol* 118: 198–203. doi: 10.1016/j.fct.2018.05.008 [PubMed: 29751076]
- Li Z, Romanoff LC, Lewin MD, Porter EN, Trinidad DA, Needham LL, Patterson DG Jr. and Sjodin A (2010). "Variability of urinary concentrations of polycyclic aromatic hydrocarbon metabolite in general population and comparison of spot, first-morning, and 24-h void sampling." *Journal of exposure science & environmental epidemiology* 20(6): 526–535. doi: 10.1038/jes.2009.41 [PubMed: 19707251]
- Maroni M, Colosio C, Ferioli A and Fait A (2000). "Biological Monitoring of Pesticide Exposure: a review. Introduction." *Toxicology* 143(1): 1–118. DOI: 10.1016/s0300-483x(99)00152-3 [PubMed: 10675783]
- Muhammad-Kah R, Liang Q, Frost-Pineda K, Mendes PE, Roethig HJ and Sarkar M (2011). "Factors affecting exposure to nicotine and carbon monoxide in adult cigarette smokers." *Regul Toxicol Pharmacol* 61(1): 129–136. doi: 10.1016/j.yrtph.2011.07.003 [PubMed: 21798300]
- Nollen NL, Mayo MS, Clark L, Cox LS, Khariwala SS, Pulvers K, Benowitz NL and Ahluwalia JS (2017). "Tobacco toxicant exposure in cigarette smokers who use or do not use other tobacco products." *Drug Alcohol Depend* 179: 330–336. doi: 10.1016/j.drugalcdep.2017.07.021 [PubMed: 28843083]
- Rostron BL, Corey CG, Chang JT, van Bommel DM, Miller ME and Chang CM (2019). "Associations of Cigarettes Smoked Per Day with Biomarkers of Exposure Among U.S. Adult Cigarette Smokers in the Population Assessment of Tobacco and Health (PATH) Study Wave 1 (2013-2014)." *Cancer Epidemiol Biomarkers Prev* 28(9): 1443–1453. doi: 10.1158/1055-9965.EPI-19-0013 [PubMed: 31239264]
- Scrucca L, Fop M, Murphy TB and Raftery AE (2016). "mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models." *R j* 8(1): 289–317. [PubMed: 27818791]
- Shahab L, Goniewicz ML, Blount BC, Brown J, McNeill A, Alwis KU, Feng J, Wang L and West R (2017). "Nicotine, Carcinogen, and Toxin Exposure in Long-Term E Cigarette and Nicotine Replacement Therapy Users: A Cross-sectional Study." *Ann Intern Med* 166(6): 390–400. doi: 10.7326/M16-1107. [PubMed: 28166548]
- USDHHS (2010). *How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General*. Atlanta (GA), Centers for Disease Control and Prevention (US) <https://www.ncbi.nlm.nih.gov/books/NBK53017/>
- USDHHS (2014). *The Health Consequences of Smoking: 50 Years of Progress. A Report of the Surgeon General*. Atlanta, GA, U.S. Department of Health and Human Services, Centers for

Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health <https://www.ncbi.nlm.nih.gov/pubmed/24455788> [dataset]

- USDHHS (2019). Population Assessment of Tobacco and Health (PATH) Study [United States] Restricted-Use Files, Inter-university Consortium for Political and Social Research [distributor] (ICPSR 36231) <https://www.icpsr.umich.edu/icpsrweb/NAHDAP/studies/36231>
- van Bochove K, van Schalkwijk DB, Parnell LD, Lai CQ, Ordovas JM, de Graaf AA, van Ommen B and Arnett DK (2012). “Clustering by plasma lipoprotein profile reveals two distinct subgroups with positive lipid response to fenofibrate therapy.” PLoS One 7(6): e38072. [PubMed: 22719863]
- Voos N, Goniewicz ML and Eissenberg T (2019). “What is the nicotine delivery profile of electronic cigarettes?” Expert Opin Drug Deliv 16(11): 1193–1203. doi: 10.1080/17425247.2019.1665647 [PubMed: 31495244]
- Wang Y, Wong LY, Meng L, Pittman EN, Trinidad DA, Hubbard KL, Etheredge A, Del Valle-Pinero AY, Zamoiski R, van Bommel DM, Borek N, Patel V, Kimmel HL, Conway KP, Lawrence C, Edwards KC, Hyland A, Goniewicz ML, Hatsukami D, Hecht SS and Calafat AM (2019). “Urinary concentrations of monohydroxylated polycyclic aromatic hydrocarbons in adults from the U.S. Population Assessment of Tobacco and Health (PATH) Study Wave 1 (2013-2014).” Environ Int 123: 201–208. doi: 10.1016/j.envint.2018.11.068. [PubMed: 30530162]

Highlights

- Cluster analysis is powerful to uncover diverse toxicant exposure profiles
- Heterogeneity in tobacco toxicant exposure exists, even in same product users
- Tobacco risk assessment should consider both user and product characteristics
- Clusters marked by high biomarkers were dominated by cigarette smokers
- SLT users were in clusters with high TSNA and nicotine and low PAH and VOC

Table 1.

Characteristics of current established tobacco users, Population Assessment of Tobacco and Health Wave 1 (2013-2014) (N=6,724)

Characteristic	Weighted % (95% CI)
Sex	
Male	57.4 (55.3, 59.6)
Female	42.6 (40.4, 44.7)
Age (years)	
18-24	17.5 (16.1, 19.0)
25-34	25.0 (23.1, 26.9)
35-54	37.1 (34.9, 39.3)
55 +	20.4 (18.6, 22.4)
Race/ethnicity	
White, NH	68.7 (66.2, 71.1)
Black, NH	13.8 (12.0, 15.7)
Other, NH	5.8 (5.0, 6.9)
Hispanic	11.7 (10.5, 13.1)
Education	
Less than high school	25.1 (23.5, 26.6)
High school graduate	27.5 (25.7, 29.5)
Some college/associate's degree	35.4 (33.6, 37.2)
Bachelor's/ advanced degree	12.0 (10.8, 13.4)
U.S. Census region	
Northeast	17.0 (15.4, 18.8)
Midwest	23.8 (21.2, 26.6)
South	40.4 (37.9, 43.1)
West	18.8 (17.0, 20.7)
Tobacco user type	
Exclusive combustible cigarette smokers	63.4 (62.2, 64.6)
Exclusive ECIG users	3.0 (2.6, 3.5)
Exclusive SLT users	6.4 (5.6, 7.2)
Dual (cigarette and ECIG) users	5.4 (4.8, 6.2)
Poly tobacco users	15.4 (14.4, 16.5)
Exclusive OCT smokers	6.3 (5.5, 7.2)

Notes. NH: Non-Hispanic. SLT: loose snus, moist snuff, dip, spit, or chewing tobacco, snus, and dissolvable tobacco. Poly: any two or more products except those included in dual users. OCT: Other combustible tobacco products include traditional cigars, cigarillos, filtered cigars, pipe tobacco, and hookah. Because of small sample size, data pertaining to OCT were suppressed.

Table 2.

Geometric means and 95% Confidence Interval of tobacco biomarkers per cluster, Population Assessment of Tobacco and Health Wave 1 (2013–2014) (N=6,724)

Cluster	TSNA (pg/mg creatinine)		PAH (ng/mg creatinine)		voc (ng/mg creatinine)			Nicotine (nmol/mg creatinine)	
	NNAL	NNN	1-NAP	3-FLU	CEMA	CYMA	MHB3	TNE2	
Cluster1	694.6 (656.1, 735.3)	35.8 (33.4, 38.4)	26.6 (25.7, 27.4)	1.4 (1.4, 1.5)	494.7 (471.3, 519.2)	393.5 (368.8, 419.9)	57.5 (368.8, 419.9)	82.5 (78.4, 86.8)	
Cluster2	138.5 (126.4, 151.9)	5.9 (5.5, 6.2)	11.3 (10.9, 11.9)	0.6 (0.6, 0.7)	267.9 (254.3, 282.2)	134.1 (122.5, 146.8)	21.7 (20.6, 22.8)	20.1 (17.6, 22.8)	
Cluster3	131.9 (125.6, 138.4)	6.5 (6.3, 6.8)	5.4 (5.2, 5.6)	0.33 (0.32, 0.34)	163.8 (159.3, 168.4)	68.5 (65.5, 71.7)	16.1 (15.6, 16.7)	22.7 (21.5, 24.0)	
Cluster4	23.0 (20.6, 25.6)	1.7 (1.6, 1.7)	3.1 (3.0, 3.3)	0.2 (0.18, 0.20)	111.4 (106.2, 116.8)	30.8 (27.6, 34.3)	7.3 (7.0, 7.7)	2.4 (2.0, 2.9)	
Cluster5	203.7 (161.9, 256.3)	17.4 (15.6, 19.3)	0.9 (0.8, 1.0)	0.1 (0.09, 0.11)	86.0 (81.0, 91.3)	1.5 (1.4, 1.6)	3.4 (3.2, 3.6)	47.5 (42.0, 53.8)	
Cluster6	12.6 (11.1, 14.2)	1.6 (1.5, 1.7)	0.9 (0.85, 0.95)	0.074 (0.70, 0.077)	77.5 (74.5, 80.8)	5.7 (5.2, 6.3)	3.5 (3.4, 3.7)	1.3 (1.1, 1.5)	
Cluster7	364.3 (351.6, 377.4)	18.8 (18.1, 19.4)	13.0 (12.7, 13.4)	0.8 (0.7, 0.8)	308.2 (300.1, 316.5)	189.1 (182.6, 195.8)	36.8 (35.8, 37.9)	57.0 (55.7, 58.5)	
Cluster8	1.59 (1.45, 1.75)	1.65 (1.55, 1.75)	0.73 (0.68, 0.78)	0.057 (0.056, 0.060)	82.04 (77.6, 86.6)	1.25 (1.18, 1.33)	3.31 (3.14, 3.48)	0.013 (0.012, 0.016)	
Cluster9	146.67 (105.5, 203.9)	8.18 (6.78, 9.87)	465.45 (328.79, 658.91)	0.40 (0.33, 0.52)	297.55 (253.3, 349.5)	81.78 (58.4, 114.49)	24.07 (20.08, 28.86)	19.3 (12.3, 30.4)	
Cluster10	1027.9 (835, 1265.3)	82.5 (70.9, 96.1)	3.4 (2.88, 4.03)	0.32 (0.28, 0.37)	142.65 (129.5, 157.1)	13.2 (9.7, 18.12)	10.04 (9.04, 12.2)	88.05 (77.88, 99.54)	

Notes. NNAL: 4-(methylnitrosamino)-1-(3-pyridyl)-1-butananol; NNN: N'-nitrosomonicotine; 1-NAP: 1-Naphthol; 3-FLU: 3-Hydroxyfluorene; CEMA: N-acetyl-S-(2-carboxyethyl)-L-cysteine; CYMA: N-acetyl-S-(2-cyanoethyl)-L-cysteine; MHB3: N-acetyl-S-(4-hydroxy-2-buten-1-yl)-L-cysteine; TNE2: Total Nicotine Equivalents 2.

Cluster label, size, and descriptions among current established tobacco users, Population Assessment of Tobacco and Health Wave 1 (2013–2014) (N=6,724)

Table 3.

Cluster number	Label	Size: Weighted % (95% CI)	Description	Biomarker Ranking
Cluster 7	Very high	29.4% (27.6, 31.2)	Largest cluster in size. On average, users in this cluster has the second or third highest mean concentrations of all biomarkers among all clusters	NNAL: 3, NNN: 3, 1-NAP: 3, 3-FLU: 2, CEMA: 2, CYMA: 2, MHB3: 2, TNE2: 3
Cluster 3	Middle	19.9% (18.5, 21.4)	Second largest cluster in size. Among all clusters, this cluster contains biomarkers toward the middle for all mean concentrations with rankings ranging from 5 to 6	NNAL: 6, NNN: 6, 1-NAP: 5, 3-FLU: 5, CEMA: 5, CYMA: 5, MHB3: 5, TNE2: 5
Cluster 1	Extremely high	12.3% (10.9, 13.9)	Third largest cluster in size. This cluster includes the highest or second highest means of biomarker concentrations among all clusters	NNAL: 2, NNN: 2, 1-NAP: 2, 3-FLU: 1, CEMA: 1, CYMA: 1, MHB3: 1, TNE2: 2
Cluster 2	PAH&VOC high	10.3% (9.2, 11.5)	Fourth largest cluster in size. Compared to other clusters, this cluster has the sixth and seventh highest mean concentrations of biomarkers of TSNA and TNE2 but the third and fourth highest mean concentrations of biomarkers of PAHs and VOCs	NNAL: 7, NNN: 7, 1-NAP: 4, 3-FLU: 3, CEMA: 4, CYMA: 3, MHB3: 4, TNE2: 6
Cluster 4	Low	6.8% (5.8, 8.0)	This cluster contains tobacco users with low mean concentrations of all biomarkers among all clusters; rankings ranged from 6 to 8	NNAL: 8, NNN: 8, 1-NAP: 7, 3-FLU: 7, CEMA: 7, CYMA: 6, MHB3: 7, TNE2: 8
Cluster 8	Extremely low	5.9% (4.8, 7.1)	This cluster has the lowest or second lowest biomarker concentrations compared to the rest of the clusters	NNAL: 10, NNN: 9, 1-NAP: 10, 3-FLU: 10, CEMA: 9, CYMA: 9, 3-FLU: 10, CEMA: 9, CYMA: 9, MHB3: 10, TNE2: 10
Cluster 6	Very low	5.4% (4.9, 6.1)	This cluster includes tobacco users with very low mean concentrations of all biomarkers among all clusters with rankings ranging from 8 to 10	NNAL: 9, NNN: 10, 1-NAP: 8, 3-FLU: 9, CEMA: 10, CYMA: 8, MHB3: 8, TNE2: 9
Cluster 5	TSNA&TNE2 high/PAH&VOC low	5.2% (4.7, 5.8)	This cluster has high mean biomarker concentrations of TSNA and TNE2 (rankings of 4) but low means of biomarker concentrations of PAHs and VOCs (rankings ranging from 8 to 10)	NNAL: 4, NNN: 4, 1-NAP: 9, 3-FLU: 8, CEMA: 8, CYMA: 10, MHB3: 9, TNE2: 4
Cluster 10	TSNA&TNE2 extremely high	2.4% (2.1, 2.9)	This cluster has the highest mean concentrations of biomarkers of TSNA and TNE2 among all the clusters but other biomarker concentrations are toward the middle or low end of the rankings ranging from 6 to 7	NNAL: 1, NNN: 1, 1-NAP: 6, 3-FLU: 6, CEMA: 6, CYMA: 7, MHB3: 6, TNE2: 1
Cluster 9	High 1-NAP	2.3% (1.8, 2.9)	Smallest cluster in size. It has the highest concentrations of 1-NAP compared to other clusters but all other biomarker concentrations were toward the middle of the rankings ranging from 3 to 7	NNAL: 5, NNN: 5, 1-NAP: 1, 3-FLU: 4, CEMA: 3, CYMA: 4, MHB3: 3, TNE2: 7

Notes. Clusters listed in order of their size, largest to the smallest; NNAL: 4-(mylthelnitrosamino)-1-(3-pyridyl)-1-butanol; NNN: N'-nitrosomnicotine; 1-NAP: 1-Naphthol; 3-FLU: 3-Hydroxyfluorene; CEMA: N-acetyl-S-(2-carboxyethyl)-L-cysteine; CYMA: N-acetyl-S-(4-hydroxy-2-buten-1-yl)-L-cysteine; TNE2: Total Nicotine Equivalents 2. PAH: polycyclic aromatic hydrocarbons; VOC: volatile organic compounds. Descriptive labels are for designed for descriptive purposes only, they are not meant to represent risk of adverse health outcomes.

Table 4.

Predicted probabilities of tobacco user type per cluster based on multinomial logistic regression model.

	Cigarette	ECIG	SLT	Dual	Poly
Cluster	Weighted % (95% CI)	Weighted % (95% CI)	Weighted % (95% CI)	Weighted % (95% CI)	Weighted % (95% CI)
Cluster 1	12.8 (11.0, 14.7)	<i>1.6 (0.1, 3.3)</i>	3.5 (1.2, 5.8)	13.1 (8.2, 18.1)	16.4 (13.0, 19.9)
Cluster 2	11.3 (9.9, 12.6)	6.1 (2.0, 10.2)	<i>2.0 (0.1, 4.2)</i>	10.8 (7.1, 14.6)	11.1 (8.6, 13.6)
Cluster 3	24.2 (22.0, 26.4)	12.3 (8.1, 16.6)	7.3 (3.8, 10.8)	24.7 (19.7, 29.6)	15.8 (13.6, 18.0)
Cluster 4	7.1 (5.4, 8.8)	7.6 (4.2, 10.9)	4.3 (0.8, 7.9)	6.0 (2.9, 9.1)	5.6 (4.2, 7.1)
Cluster 5	<i>0.1 (0.0, 0.2)</i>	30.6 (24.0, 37.2)	44.5 (38.5, 50.6)	<i>0.9 (0.1, 1.9)</i>	4.0 (2.7, 5.2)
Cluster 6	4.3 (3.5, 5.1)	15.3 (9.7, 20.9)	7.8 (4.3, 11.3)	4.0 (2.1, 5.8)	4.7 (3.7, 5.8)
Cluster 7	33.2 (30.7, 35.8)	8.8 (4.6, 13.0)	6.2 (3.7, 8.7)	33.9 (28.9, 39.0)	32.1 (26.5, 37.8)
Cluster 8	4.0 (2.8, 5.3)	11.6 (7.1, 16.2)	6.2 (2.9, 9.5)	1.3 (0.2, 2.4)	4.6 (3.2, 6.0)
Cluster 9	2.2 (1.5, 2.9)	3.0 (0.4, 5.6)	0.7 (0.2, 1.3)	3.6 (1.0, 6.3)	1.8 (0.5, 3.0)
Cluster10	0.7 (0.4, 1.1)	<i>3.1 (0.5, 6.7)</i>	17.4 (10.9, 23.8)	1.6 (0.4, 2.9)	3.9 (2.5, 5.2)

Notes. All but italicized cells are statistically significant. Because of limited table space, cluster numbers were used instead of cluster descriptive labels. Because of small sample size, data pertaining to OCT were suppressed.

Table 5.

Smoking quantity per cluster among exclusive smokers, dual and poly tobacco users.

	Exclusive cigarette smokers	Dual users	Poly users
	Weighted mean (95% CI)	Weighted mean (95% CI)	Weighted mean (95% CI)
Cluster 1	598.7 ^a (599.8, 637.6)	561.0 ^a (473.4, 648.5)	642.1 ^a (531.2, 753.1)
Cluster 2	372.8 ^b (340.7, 407.0)	337.4 ^{bc} (272.4, 402.4)	392.5 ^b (337.9, 447.1)
Cluster 3	320.0 ^c (294.2, 345.8)	280.9 ^{bc} (245.9, 314.7)	366.6 ^b (320.5, 412.8)
Cluster 4	132.4 (114.0, 150.7)	201.2 (174.7, 254.8)	175.3 (132.2, 218.4)
Cluster 5	16.2 (5.4, 27.0)	87.3 (0, 275.8)	56.5 (0, 116.1)
Cluster 6	96.7 (75.5, 118.0)	173.9 (109.8, 238.0)	121.5 (64.7, 178.2)
Cluster 7	509.6 ^d (488.8, 530.4)	453.4 ^d (390.4, 516.5)	517.3 ^{ad} (473.8, 560.7)
Cluster 8	66.6 (22.4, 110.9)	8.6 (4.5, 12.7)	83.9 (23.5, 144.4)
Cluster 9	574.1 (429.0, 719.2)	418.1 (342.9, 493.2)	267.5 (149.9, 385.1)
Cluster 10	372.2 (290.2, 454.1)	304.8 (160.2, 449.4)	289.0 (189.0, 389.1)

Notes, poly users who smoked cigarettes. Superscripts indicate significant differences in means across clusters (i.e., by column). If the values have common letter(s), there is no statistically significant difference between them; if the values do not have a common letter, there is a statistically significant difference between them.

Table 6.

Frequency of ECIG use per cluster among exclusive ECIG users.

Cluster	Weighted mean (95% CI)
Cluster 1	21.1 (11.4, 30.7)
Cluster 2	22.7 (13.9, 31.6)
Cluster 3	25.6 (22.1, 29.2) ^{ac}
Cluster 4	22.7 (17.3, 28.2)
Cluster 5	29.6 (28.9, 30.3) ^b
Cluster 6	22.7 (19.2, 26.2) ^{ac}
Cluster 7	24.2 (19.4, 28.9)
Cluster 8	10.4 (6.4, 14.5) ^d
Cluster 9	25.1 (16.4, 33.7)
Cluster 10	-

Notes, frequency of ECIG use was defined as number of days used ECIG in past 30 days. Superscripts indicate significant differences between the means per cluster. If the values have common letter(s), there is no statistically significant difference between them; if the values do not have a common letter, there is a statistically significant difference between them. Number of ECIG users in cluster 10 is too small to estimate weighted mean and 95% CI of frequency of use.