



# HHS Public Access

Author manuscript

*Environ Sci Technol.* Author manuscript; available in PMC 2024 February 07.

Published in final edited form as:

*Environ Sci Technol.* 2023 February 07; 57(5): 2042–2053. doi:10.1021/acs.est.2c08121.

## Distinguishing Exposure to Secondhand and Thirdhand Tobacco Smoke among U.S. Children using Machine Learning: NHANES 2013-2016

Ashley L. Merianos<sup>\*a</sup>, E. Melinda Mahabee-Gittens<sup>b</sup>, Timothy M. Stone<sup>c</sup>, Roman A. Jandarov<sup>c</sup>, Lanqing Wang<sup>d</sup>, Deepak Bhandari<sup>d</sup>, Benjamin C. Blount<sup>d</sup>, Georg E. Matt<sup>e</sup>

<sup>a</sup>School of Human Services, University of Cincinnati, P.O. Box 210068, Cincinnati OH, 45221, USA

<sup>b</sup>Division of Emergency Medicine, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, 3333 Burnet Avenue, MLC 2008, Cincinnati, OH, 45229, USA

<sup>c</sup>Division of Biostatistics and Bioinformatics, Department of Environmental and Public Health Sciences, College of Medicine, University of Cincinnati, Kettering Lab Building, 160 Panzeca Way, Cincinnati, OH, 45267-0056, USA

<sup>d</sup>Tobacco and Volatiles Branch, Division of Laboratory Sciences, National Center for Environmental Health, Centers for Disease Control and Prevention, 4770 Buford Hwy NE, Atlanta, GA, 30341, USA

<sup>e</sup>Department of Psychology, College of Sciences, San Diego State University, 9245 Sky Park Court, Suite 225, San Diego, CA, 92123, USA

### Abstract

While the thirdhand smoke (THS) residue from tobacco smoke has been recognized as a distinct public health hazard, there are currently no gold standard biomarkers to differentiate THS from secondhand smoke (SHS) exposure. This study used machine learning algorithms to assess which combinations of biomarkers and reported tobacco smoke exposure measures best differentiate children into three groups: no/minimal tobacco smoke exposure (NEG); predominant THS exposure (TEG); and mixed SHS and THS exposure (MEG). Participants were 4,485 nonsmoking 3-17-year-olds from the National Health and Nutrition Examination Survey 2013-2016. We fitted and tested random forest models, and the majority (76%) of children were classified in NEG, 16% were classified in TEG, and 8% were classified in MEG. The final classification model based on reported exposure, biomarker, and biomarker ratio variables had a prediction accuracy of 95%. This final model had prediction accuracies of 100% for NEG, 88% for TEG, followed by 71% for MEG. The most important predictors were the reported number of household smokers, serum cotinine, serum hydroxycotinine, and urinary 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL). In the absence of validated biomarkers

\* ashley.merianos@uc.edu, Phone: 513-556-6753, Fax: 513-556-3898.

Author Contributions

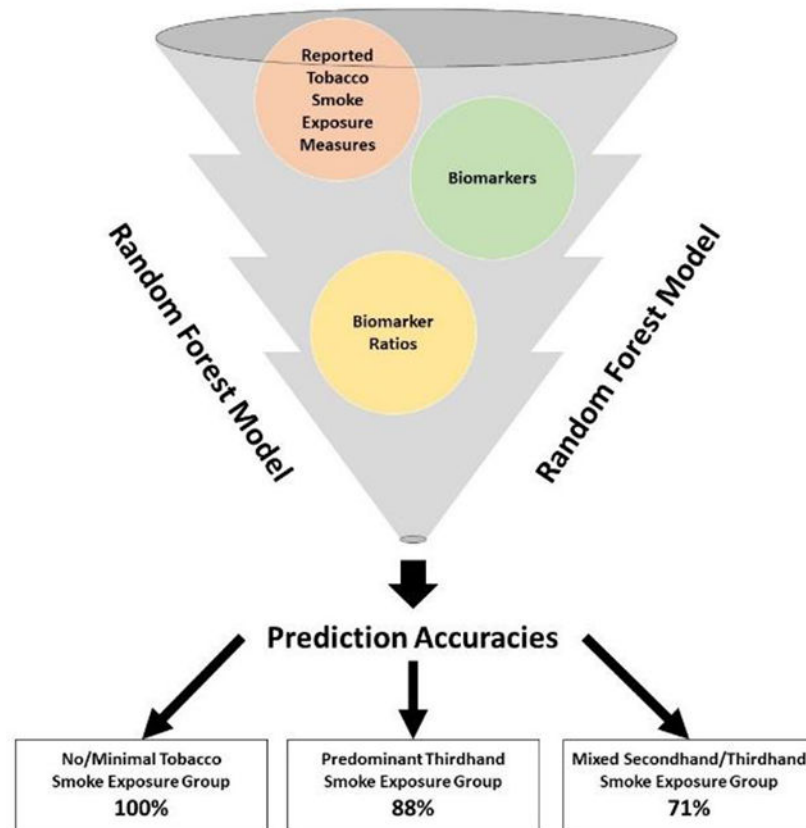
The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

The authors declare no competing financial interest.

specific to THS, comprehensive biomarker and questionnaire data for tobacco smoke exposure can distinguish children exposed to SHS and THS with high accuracy.

## Graphical Abstract

### Classifying Children into Tobacco Smoke Exposure Groups



## Keywords

Tobacco smoke exposure; thirdhand smoke; biomarkers; child; machine learning; National Health and Nutrition Examination Survey

## Introduction

Thirdhand smoke (THS) is a distinct public health hazard resulting from aged secondhand smoke (SHS) left behind from active smoking that adsorbs to surfaces, accumulates in dust, and can become embedded in furnishings.<sup>1</sup> Whereas SHS is inhaled from freshly emitted tobacco smoke, THS can be inhaled, ingested, and/or dermally absorbed from aged SHS for days to years after smoking ended.<sup>1</sup> THS remains in the environment and can react with ambient oxidants (e.g., nitrous acid, ozone) to create multiple harmful byproducts such as tobacco-specific nitrosamines (TSNAs) and ultrafine particles that are re-emitted and re-suspended in the air.<sup>2</sup> The THS pollutant mixture contains, among others, tobacco-specific

(e.g., nicotine, TSNA, nicotine) and tobacco non-specific pollutants such as polycyclic aromatic hydrocarbons (PAHs) and volatile organic compounds (VOCs) known to cause harm to humans.<sup>3-7</sup>

Nicotine intake from exposure to SHS and THS can be measured using several metabolites including cotinine, the widely used main proximate metabolite of nicotine, which is converted to other metabolites such as trans-3'-hydroxycotinine.<sup>8</sup> Compared to cotinine assessment only to measure exposure to nicotine, total nicotine equivalents (TNEs) is considered a gold standard biomarker for estimating daily nicotine intake since TNEs are less impacted by the substantial individual variability of nicotine metabolism pathways.<sup>9</sup> TNE is the sum of the free and glucuronide conjugated forms of nicotine and its metabolites including cotinine and trans-3'-hydroxycotinine.<sup>9</sup> Additionally, the measurement of nicotine intake via cotinine can underestimate exposure to other toxicants such as 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK),<sup>10</sup> which is the most potent carcinogenic TSNA that is rapidly metabolized to urinary 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL) with similar carcinogenicity.<sup>11</sup>

Children are most vulnerable to tobacco smoke exposure (TSE), the cumulative SHS and THS exposure to tobacco, and its related effects due to size- and age-specific behaviors and activity patterns.<sup>2</sup> Relative to adults, children's biological and behavioral differences place them at increased risk of interacting with THS pollution in their environment (e.g., house dust on floors), such as their inhalation patterns (e.g., faster breathing rate relative to body weight), smaller body sizes, and pica and crawling behaviors.<sup>11-13</sup> Concerning prevalence of TSE, approximately 35% of U.S. children ages 3-17 years have detectable cotinine levels indicative of TSE, and rates are higher among children who are non-Hispanic Black and of lower socioeconomic status.<sup>14</sup> Consequently, when compared to adults, U.S. children exposed to tobacco smoke have nearly three-fold higher NNK exposure.<sup>15</sup> Additionally, NNAL positively correlates with cotinine among children, and thus, children with higher reported TSE have higher NNAL and cotinine concentrations.<sup>16</sup> Prior research also suggests a gradient relationship between biomarkers and home TSE with children who live with smokers who smoke indoors (home SHS+THS proxy) having the highest NNAL levels, followed by children living with smokers who do not smoke indoors (predominant home THS proxy), and then children living with nonsmokers (no home TSE proxy) having the lowest NNAL levels.<sup>17</sup>

Currently, there are no gold standard TSE biomarkers to differentiate exposure to THS from exposure to SHS. Methods to distinguish THS from SHS in nonsmokers include assessing self-report, THS-related main metabolites, and NNAL/cotinine ratios.<sup>1</sup> In addition to the common limitations surrounding self-reported measures such as recall and social desirability biases,<sup>18</sup> questionnaire items cannot be exclusively relied upon to assess THS exposure since parents who are nonsmokers and who try to protect their children from TSE may be unaware of THS pollutants that can remain in environments long after smoking has ceased.<sup>19</sup> Concerning potential THS-specific measures, the TSNA 4-(methylnitrosamino)-4-(3-pyridyl)butanal (NNA) is not typically found in freshly emitted smoke, but is produced when the combustion sources, including nicotine, react with nitrous oxide in the environment.<sup>1</sup> Therefore, NNA is difficult to measure in THS samples likely

due to this aldehyde's high chemical reactivity during combustion that makes it difficult for NNA to subsist.<sup>1</sup> While there are other TSE markers not specific to THS such as TNE, ratios may be of interest if the proportion of the two biomarkers differ between SHS and THS. For example, the NNAL/cotinine ratio may assist in differentiating between SHS and THS exposure due to: (1) NNK concentrations increasing or remaining stable over time due to reacting with ambient oxidants while nicotine concentrations decrease;<sup>20</sup> and (2) the varying elimination kinetics with NNAL having a greater half-life compared to cotinine.<sup>21</sup> Therefore, the NNAL/cotinine ratio is posited to be lower among children with SHS exposure, but higher among children with THS exposure due to the premise of greater dermal exposure, ingestion, and inhalation of THS as well as behaviors such as hand-to-mouth behavior and playing on the floor.<sup>1</sup> Similar to urinary NNAL only, children have higher NNAL/cotinine ratio levels compared to adults.<sup>22</sup> However, more research is needed to assess biomarkers in ratio form with TNE, which TNE2 (i.e., molar concentration cotinine + trans-3'-hydroxycotinine sum) has been strongly correlated with cotinine and trans-3'-hydroxycotinine among U.S. individuals.<sup>23</sup> Further, while PAH hydroxyfluorene metabolites<sup>24,25</sup> and VOC acrylonitrile metabolites<sup>26</sup> are highly selective to smoke exposure, more research is needed to assess the contributions that these PAHs and VOCs may have to SHS and THS in individual and ratio forms with TNE2. Thus, it is crucial to examine comprehensive, highly selective biomarkers of TSE to assess children's exposure to tobacco smoke and to potentially further differentiate children by TSE type (i.e., SHS or THS) to improve TSE group classification.

The study objective was to compare multiple tobacco-specific (e.g., TNE2, NNAL) and non-specific (e.g., PAHs, VOCs) biomarkers of exposure with reported TSE patterns to assess which biomarker profiles differentiate U.S. children ages 3-17 years into the following three groups: (1) no/minimal TSE group (NEG): lives with nonsmokers, no reported SHS exposure; (2) predominant THS exposure group (TEG): lives with tobacco smokers, no reported SHS exposure; and (3) mixed SHS and THS exposure group (MEG): lives with tobacco smokers, reported SHS exposure. Children in the NEG who are thought to be protected from TSE can be minimally or incidentally exposed to THS.<sup>19</sup> However, it is important to assess the known and systematic source of living with smokers to distinguish the TEG from the NEG, despite sharing the criteria of no reported SHS exposure. This rationale is due to children in the TEG having a chronic, increased magnitude of exposure to THS due to household smokers carrying THS pollutants into these children's homes compared to children in the NEG who do not live with smokers.<sup>19</sup> We hypothesized that comprehensive questionnaire and biomarker data would differentiate by TSE type and improve TSE classification, and that biomarkers in ratio form with TNE2 (e.g., NNAL/TNE2) would further differentiate by TSE type and improve TSE classification.

## Materials and Methods

### Participants and Procedures

The present study relies on a secondary analysis of National Health and Nutrition Examination Survey (NHANES) data (2013-2014 and 2015-2016) that are representative of the total U.S. pediatric population living in the 50 U.S. States and District of Columbia.

NHANES survey and analytic methods are available elsewhere.<sup>27–29</sup> In brief, participants were recruited for each cross-sectional cycle using a complex, multistage probability cluster design, with overall response rates of 71% in 2013–2014 and 61% in 2015–2016.<sup>29</sup> We limited our analysis to combined public-use data from these two consecutive survey cycles,<sup>30,31</sup> due to having the most recent and complete laboratory data available at time of analysis (e.g., PAHs and VOCs were not yet publicly available for the 2017–2018 cycle). The two NHANES cycles were approved by the National Center for Health Statistics ethics review board. Additionally, the University of Cincinnati's institutional review board approved this study (#2020-0350) with an exempt determination due to the use of secondary, publicly available data with no participant consent required.

We delimited our analysis to 4,485 nonsmoking children who were ages 3–17 years and had serum cotinine results available for analysis. Child age was limited to 3–17 years for whom NHANES collected household interview and medical examination data including biological samples to analyze the biomarkers of interest (e.g., serum cotinine). NHANES did not measure the biomarkers included in this study on participants <3 years old, and therefore, the younger group was excluded from analyses ( $n=2,758$ ). Children missing data on household smokers and home TSE were also excluded prior to analyses ( $n=70$ ). Additionally, children 12 years old ( $n=3,038$ ) were asked if they smoked any tobacco product (i.e., cigarettes, electronic cigarettes [e-cigarettes], cigars, little cigars/cigarillos, hookah, waterpipes) in the past 5-days or if they smoked cigarettes in the past 30-days, and were excluded if they reported past 5-day tobacco use or past 30-day cigarette smoking ( $n=94$ ).

We compared children ages 3–17 years who were included ( $N=4,485$ ) versus excluded due to missing serum cotinine results ( $n=1,411$ ), and found sociodemographic differences based on child age, child race/ethnicity, caregiver age, caregiver education level, household income to federal poverty level ratio (FPL), number of household members, and number of child household members. Specifically, children who were excluded from this study based on not having serum cotinine data available for analysis were more likely to be non-Hispanic White (57% versus 51%,  $p<0.001$ ), have caregivers with an education college graduate (37% versus 27%,  $p<0.001$ ), in the highest FPL category 350% (33% versus 27%,  $p<0.001$ ), and have two children in the household (41% versus 35%,  $p<0.001$ ). Additionally, children who were excluded from this study were more likely to be younger ( $M=8.1$  versus  $M=10.5$ ,  $p<0.001$ ), have caregivers who were younger ( $M=39.5$  versus  $M=41.3$ ,  $p<0.001$ ), and live with a lower number of household members ( $M=4.5$  versus  $M=4.7$ ,  $p<0.001$ ). No differences were found based on child sex ( $p=0.524$ ), caregiver sex ( $p=0.977$ ), family home ownership status ( $p=0.199$ ), and number of household rooms ( $p=0.754$ ) when comparing 3–17-year-olds excluded versus included in this study based on serum cotinine data available for analysis.

## Measures

NHANES 2013–2016 data used in this study were derived from questionnaires and laboratory analysis as described below.

## NHANES questionnaire data

**Household smokers and home TSE**—As a first step, we evaluated TSE in the household via three items from the questionnaire’s “Smoking – Household Smokers” section to distinguish children into the three TSE groups. This section was administered to adult proxy respondents for 0-11-year-olds while 12-17-year-olds self-reported the number of residents who lived in the household and smoked tobacco products. Responses options were “no one in household is a smoker,” “one household member is a smoker,” “two household members are smokers,” and “three or more household members are smokers.” Children who did not live with a tobacco smoker were preliminarily classified into the NEG prior to SHS exposure assessment in other locations (e.g., cars), as described below.

If one or more household residents smoked tobacco, then two follow-up questions were asked about the number of: (1) residents who lived in the household who smoked tobacco products inside the home (0, 1, and 2 smokers); and (2) days the resident(s) smoked inside the home in the past 7-days (range 0-7). To distinguish children into the TEG versus the MEG, children who lived with tobacco smokers who did not smoke inside the home (i.e., 0 smokers smoked tobacco inside the home and smoked inside the home on 0 days in the past 7-days) were preliminarily classified into the TEG prior to SHS exposure assessment in other locations. Children who lived with tobacco smokers who smoked inside the home (i.e., 1 smoker smoked tobacco inside the home and smoked tobacco inside the home on 1 day in the past 7-days) were classified into the MEG.

**SHS exposure in the past 7-days**—To further distinguish children into the NEG, TEG, and MEG based on exposure in locations other than the home, we evaluated SHS exposure in various environments in the past 7-days via four items from the questionnaire’s “Smoking – SHS Exposure” section. For each SHS exposure item, a preliminary filter yes/no question was asked to determine whether the child spent time in the particular environment during the past 7-days (i.e., restaurant, car, another home, and any other indoor area). If the child was in an environment (e.g., rode in a car), then a follow-up yes/no question was asked about whether someone else smoked tobacco products while the participant was in that respective environment in the past 7-days. Participants with a “yes” response were considered exposed to SHS in the environment.

**Sociodemographics**—We assessed sociodemographics from the questionnaire’s “Demographics” and “Housing Characteristics” sections. Demographic-related items were administered to adult proxy respondents for 0-15-year-olds while 16-17-year-olds self-reported the following information: child age, sex, and race/ethnicity (i.e., non-Hispanic White, non-Hispanic Black, non-Hispanic Other/Multiracial, Hispanic); caregiver age, sex, and education level (i.e., high school graduate/equivalent, some college, college graduate); FPL (<185%, 185-349%, 350%, unspecified including don’t know/refused), and total number of members (1 - 7) and children (1, 2, 3) in the household. The household reference person, who was most frequently a parent and henceforth referred to as caregiver, was the adult household member who owned or rented the home. FPL is the ratio of family income to annual poverty thresholds of U.S. Department of Health and Human Services that

accounts for family size and geographic location to determine financial eligibility of federal programming (e.g., Head Start).<sup>29</sup>

Housing characteristic-related items were administered to the adult proxy respondents and responses were applied for all other household members. Housing characteristics included the home ownership status (owned/being bought, rented/other arrangement) and total number of rooms in the household (1 - 13).

### NHANES laboratory data

Details on NHANES 2013-2016 specimen collection/processing and laboratory methods used on 3-year-olds for tobacco-specific and non-specific biomarkers are available elsewhere.<sup>32,33</sup>

**Serum cotinine and hydroxycotinine**—We assessed cotinine and trans-3'-hydroxycotinine (i.e., hydroxycotinine) measured in serum, the widely used primary nicotine metabolites with excretion half-lives of about 15-20 hours.<sup>34</sup> NHANES collected serum cotinine and hydroxycotinine among participants ages 3 years, and analyzed the serum for nicotine metabolites using an isotope-dilution high-performance liquid chromatography/atmospheric pressure chemical ionization tandem mass spectrometric (HPLC-APCI MS/MS) method. The lower limit of detections (LLODs) for serum cotinine and serum hydroxycotinine were 0.015 ng/mL. To further distinguish children into the TSE groups after using questionnaire data, we assessed serum cotinine levels among the NEG to biochemically confirm this group had no/minimal cotinine levels.

**Urinary TNE2**—We assessed urinary TNE2, which refers to the molar sum of major nicotine metabolites, and has shown to strongly correlate with nicotine dose.<sup>35</sup> In NHANES, TNE2 is calculated as (total cotinine/176.2151) + (total hydroxycotinine/192.2145) nmol/mL.<sup>36,37</sup> In order to calculate TNE2, total cotinine and hydroxycotinine were measured in urine. Urinary cotinine is about 5-10 times more concentrated than that of serum cotinine, and urinary hydroxycotinine is 2-4 times more concentrated than urinary cotinine.<sup>38</sup> NHANES collected urine to assess cotinine and hydroxycotinine among participants ages 6 years in the 2013-2014 cycle and ages 3 years in the 2015-2016 cycle, and analyzed urine for nicotine metabolites using an isotope-dilution high performance liquid chromatography/electrospray ionization tandem mass spectrometric (HPLC-ESI-MS/MS) method.<sup>39,40</sup> The LLODs for urinary cotinine and urinary hydroxycotinine were 0.03 ng/mL.

**Urinary 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL)**—We assessed total NNAL measured in urine, the main metabolite of the TSNA carcinogen NNK that itself has similar carcinogenicity and a long half-life of about 40-45 days.<sup>41,42</sup> NHANES collected urine to assess NNAL among participants ages 6 years in the 2013-2014 cycle, and analyzed urine for NNAL using an isotope-dilution HPLC-ESI MS/MS.<sup>43,44</sup> The LLOD for urinary NNAL was 0.60 pg/mL. Urinary TSNA data were limited to the 2013-2014 cycle because the 2015-2016 cycle was unavailable at the time of analysis.

We assessed urinary NNAL separately and as the NNAL/TNE2 ratio to test our hypothesis that ratios may help to distinguish THS from SHS in children.<sup>22</sup> This is posited due to the

increase in urinary NNAL while nicotine levels rapidly decrease over time.<sup>20</sup> Thus, as SHS ages, NNAL levels potentially increase due to nicotine reacting with nitrous acid resulting in increased NNAL/nicotine ratios found in real-world data.<sup>1</sup>

**Urinary 2-hydroxyfluorene and 3-hydroxyfluorene**—We assessed 2-hydroxyfluorene and 3-hydroxyfluorene measured in urine that are PAH metabolites of fluorene; hydroxyfluorenes are the PAH metabolites that most effectively differentiate smokers from nonsmokers.<sup>24</sup> NHANES collected urine to assess PAH metabolites among participants ages 6 years in the 2013-2014 cycle and ages 3 years in the 2015-2016 cycle, and analyzed urine using isotope dilution high performance liquid chromatography-tandem mass spectrometry (SPE-HPLC-MS/MS).<sup>45</sup> The LLODs for urinary 2-hydroxyfluorene and 3-hydroxyfluorene were 8.00 ng/L. Since fluorene can have other possible sources, we also assessed both PAHs normalized to TNE2.

**Urinary N-acetyl-S-(2-cyanoethyl)-L-cysteine (2CyEMA)**—We assessed 2CyEMA measured in urine that is a VOC metabolite of acrylonitrile, which is found in tobacco smoke and other sources including acrylic, plastics, and resins,<sup>46</sup> and are selective to tobacco smoke.<sup>47</sup> NHANES collected urine to assess VOC metabolites among participants ages 6 years in the 2013-2014 cycle and ages 3 years in the 2015-2016 cycle, and analyzed urine using ultra-performance liquid chromatography coupled with electrospray tandem mass spectrometry (UPLC-ESI/MSMS).<sup>48</sup> The LLOD for urinary 2CyEMA was 0.50 µg/L. Since VOCs can also have other sources, we assessed 2CyEMA normalized to TNE2.

### Statistical analysis

We used the survey package in R Statistical Software (version 4.0.5.)<sup>49</sup> and adhered to NHANES analytic guidelines<sup>29</sup> for all analyses. We used examination sample weights to adjust for survey non-response and selection probability to yield estimates that reflect the U.S. 3-17-year-old population. We also used NHANES-provided primary sampling units and strata variables to account for the clustered design for variance estimation. Before any statistical modeling, we assessed the distributional properties of all variables including the biomarkers, which underwent logarithmic transformations to control for non-normality and heterogeneity of variances.

As a preliminary step, we classified children into the three TSE groups using questionnaire data on household smokers and home TSE (Figure 1). As a second step, we re-classified children into TSE groups based on their reported SHS exposure patterns in other locations (e.g., car). The associations between tobacco-specific biomarker levels and preliminary TSE groups were then investigated as a validation step using univariate and bivariate analyses to ensure the accuracy of the group labels using questionnaire data. After validating group membership, we assessed child sociodemographic characteristics overall and based on TSE groups. We report unweighted counts and weighted percentages and 95% confidence intervals (CIs) for categorical variables (e.g., child sex) and weighted means and standard errors (SEs) for continuous variables (e.g., child age). We also computed descriptive statistics including frequencies and weighted percentages and 95% CIs for the reported TSE



variables within each TSE group. We report geometric means (GeoMs) and 95% CIs for the log-transformed biomarker variables overall, and within each TSE group.

As a second step, we used the random forests machine learning method, which ensembles multiple decision trees to make predictions and build a “forest”.<sup>50</sup> Random forest models are an effective ensemble tool that incorporate randomness via employing bagging or bootstrap aggregation to generate each tree. Random forests randomly sample predictors by creating bootstrap replicates called bagged predictors, which this subset is then tested to create splits in each tree. This method reduces correlation between the decision trees and enhances the accuracy and robustness of the ensemble when random features are applied.<sup>50</sup> To account for missing values, we used random forest imputation with the model parameters set to run 500 trees to maximize the training set. A key advantage of random forest imputation is the model’s inability to extrapolate or impute values beyond the observed effect. This is useful, especially for continuous biomarker data that are lower bound by a LLOD. The imputation step was performed using the missRanger package in R while accounting for the sample design and applying NHANES-provided weighting, and the resulting dataset was used for all subsequent downstream analysis.

We fitted and tested random forest models using a cross-validation technique of an 80% training split to develop the model and 20% testing split to assess each model’s performance. The NHANES-provided weights were no longer applied for this step due to using the random sampling method described above and using these data for classification purposes. Instead, for the random forest models, each sample was inversely weighted according to the frequency of their respective TSE group to avoid potential classification problems resulting in imbalances from the initial varying TSE group sample sizes. For example, the smaller MEG was weighted more heavily than the larger NEG. We entered selected variables a priori and entered each predictor independently to produce variable importance scores using the Gini impurity index that reflect the overall influence, or importance, of each predictor variable in the model.

Based on the main random forest model results that included 16 variables, and is described in detail below, we conducted sensitivity analyses and fitted six additional random forests with varying predictors selected a priori to assess prediction accuracy for the models including the following predictors: (1) all five reported TSE predictors; (2) four reported TSE predictors excluding the predictor of number of household smokers; (3) all 11 biomarker and biomarker ratio predictors; (4) all seven biomarker predictors (excluding biomarker ratios); (5) all four biomarker ratio predictors; and (6) seven biomarker and biomarker ratio predictors excluding serum cotinine, serum hydroxycotinine, urinary TNE2, and urinary NNAL. The second sensitivity model excluded the predictor of the number of household smokers since this was the most important predictor in the main random forest model. The sixth sensitivity model excluded the predictors of serum cotinine, serum hydroxycotinine, urinary TNE2, and urinary NNAL, which were also of importance in the main random forest model.

As a third and final step, we fitted a logistic regression model using the four reported TSE and biomarker predictors that were identified as important to predict variation in

MEG versus TEG membership (reference category). The logistic regression model was also evaluated using an 80% training and 20% testing split.

## RESULTS

The mean (SE) age of the 4,485 nonsmoking children was 10.5 (0.01) years, and 49% were female (Table 1).

The majority of children were non-Hispanic White (51%) followed by Hispanic (25%), non-Hispanic Black (14%), and non-Hispanic Other race/Multiracial (10%). Concerning caregiver characteristics, their mean (SE) age was 41.3 (0.03) years, 50% were female, and 40% completed an education of high school graduate/equivalent. Concerning household characteristics, FPL varied with 43% in the lowest <185% level and 27% in the highest 350% level. Thirty-nine percent of children lived in rented homes or had other arrangements, and the mean (SE) number of household rooms and members were 6.7 (0.01) and 4.8 (0.01), respectively. The majority of households had two child residents (35%) and 3 child residents (48%).

### Sociodemographic Characteristics based on Child TSE Group Membership

Child race/ethnicity differed based on TSE group membership (see Table 1). Concerning caregiver characteristics, caregiver age, sex, and education level differed based on TSE group membership. Concerning household characteristics, FPL, family home ownership status, and number of household members, child household members, and household rooms differed based on TSE group membership.

### TSE Patterns Overall and based on Child TSE Group Membership

A total of 76% ( $n=3,369$ ) of children were classified in the NEG, 16% ( $n=701$ ) were classified in the TEG, and 8% ( $n=415$ ) were classified in the MEG. Irrespective of child TSE group membership, 76% lived with no household smokers, and 14% and 10% lived with one and 2 smokers, respectively (Table 2). Concerning locations of TSE, 12% had car TSE, 6% had other home TSE, 1% had restaurant TSE, and 3% had other indoor area TSE.

### Biomarker Levels Overall and based on Child TSE Group Membership

Overall biomarker and biomarker ratio levels are presented in Table 3. The highest biomarker levels were observed in the MEG followed by the TEG and then the NEG with the exception of 2- and 3-hydroxyfluorene, which were higher in the MEG followed by the NEG and then the TEG. For biomarker ratio levels with TNE2 as the denominator, ratio levels were highest in the NEG followed by the TEG and then the MEG.

### Main Random Forest Model Prediction Accuracy

The final random forest model selected for classifying children into the three TSE groups used all 16 reported TSE, biomarker, and biomarker ratio predictor variables and had a prediction accuracy of 95% (Table 4). Compared to the sensitivity model results presented below, this model had the highest NEG (100%) and MEG (71%) prediction accuracies, and the second highest TEG (88%) prediction accuracy. Concerning variable importance, the top

four important predictors in the final model were: number of household smokers (1,027), serum cotinine (333), serum hydroxycotinine (242), and urinary NNAL (93) (Figure 2).

### **Sensitivity Random Forest Models Prediction Accuracy**

As a sensitivity analysis and based on the main prediction model's results, we fitted six additional models with different variable selections (see Table 4). The two prediction models including reported TSE data only, with and without the number of household smokers variable, had the next highest overall model prediction accuracies of 93% and 76%, respectively. The model with all reported TSE data had the highest TEG prediction accuracy of 91%, but the reported TSE data model without the number of household smokers predictor variable had the lowest TEG prediction accuracy of 1%.

Next, biomarkers only (excluding ratios) had the next highest overall model prediction accuracy of TSE group membership (75%), and the models with biomarkers and biomarker ratios with (71%) and without (72%) serum cotinine, serum hydroxycotinine, urinary TNE2, and urinary NNAL metabolites had similar overall prediction accuracies. However, the model with biomarkers only had the second lowest TEG prediction accuracy of 9%. The model with the four biomarker ratios only had the lowest overall model (64%), NEG (78%), and MEG (17%) prediction accuracies (see Table 4).

### **Logistic Regression Model to Distinguish MEG versus TEG Membership**

Figure 3 presents a forest plot of logistic regression model results using the top four variables of importance in the main random forest model to predict variation in TSE group membership and distinguish between MEG versus TEG membership. The overall model prediction accuracy was 79% for these two exposure groups, with specific prediction accuracies of MEG and TEG membership of 86% and 74%, respectively. Children who lived with 2 smokers were more likely to be in the MEG (OR=1.8, 95% CI=1.4-2.2,  $p<0.001$ ) compared to children who lived with only one smoker. With every one log-unit increase of serum cotinine (OR=1.2, 95% CI=1.1-1.4,  $p<0.001$ ), serum hydroxycotinine (OR=1.3, 95% CI=1.2-1.5,  $p<0.001$ ), and urinary NNAL (OR=1.1, 95% CI=1.0-1.2,  $p<0.001$ ), children were at increased odds to be in the MEG versus the TEG.

## **DISCUSSION**

We applied machine learning methodology to compare multiple tobacco-specific and non-specific biomarkers with reported TSE patterns collected during NHANES 2013-2016 to assess which exposure profiles best differentiate U.S. children ages 3-17 years into TSE groups. As hypothesized, our main random forest model that was developed and validated using the comprehensive list of questionnaire item and biomarker predictors demonstrated good performance with high accuracy of about 95%. Specifically, the top four variables of importance included the reported measure of number of household smokers, the two nicotine metabolites of serum cotinine and hydroxycotinine, and the NNK metabolite of urinary NNAL. While biomarker ratios included in the main random forest model assisted in differentiating by TSE group and improved classification, the urinary NNAL/TNE ratio ranked ninth for importance. This particular ratio was ranked lower than the PAH and VOC

metabolite ratios normalized to TNE2. This finding was unanticipated since the NNAL/cotinine ratio, or in this case the NNAL/TNE2 ratio, is expected to be higher among those primarily exposed to THS and lower among those primarily exposed to SHS.<sup>1</sup> This is due to the premise that dust is a major source of THS exposure, and nicotine found in dust decreases more rapidly compared to NNK over time.<sup>51</sup> Additionally, nicotine and its metabolites are metabolized more rapidly with shorter half-lives compared to the NNK metabolite of NNAL.<sup>21</sup> Therefore, one possible explanation for this finding was that we only had NNAL data available from the NHANES 2013-2014 cycle. It is also important to note that while the sensitivity random forest model using the four biomarker ratio predictors only had an overall model prediction accuracy of 64%, this model had the lowest prediction accuracies for the NEG (78%) and the MEG (17%) compared to the main random forest and other sensitivity models.

Our main random forest model showed the highest prediction accuracy for the NEG and MEG, but not the TEG, indicating that key predictors are different among the MEG and TEG. Specifically, the sensitivity model with reported variables had higher prediction accuracy for the TEG, but the second lowest for the MEG. Conversely, the three sensitivity models including all biomarker predictors and all biomarker and biomarker ratio predictors with and without serum cotinine, serum hydroxycotinine, urinary TNE2, and urinary NNAL had higher model prediction accuracy for the MEG than the TEG. This shows that a combination of questionnaire and biomarker variables are key predictors to consider for TSE group membership, especially to distinguish the TEG from the MEG. It is important to note that while we used all tobacco-specific and non-specific biomarkers provided by NHANES that were available for analysis, there are environmental THS pollution markers that may further distinguish TEG from MEG membership. Concerning potential environmental markers that may further classify TSE groups and can be measured in indoor environments, the first studies to measure children's exposure to THS in indoor environments using biomarkers (e.g., urine cotinine) and environmental markers (e.g., house dust) found varying THS pollution levels among children of smokers and nonsmokers, and that THS can last six months after cessation.<sup>51-53</sup> Research indicates that the prevalence of THS is high in smokers' children who had high levels of THS pollutants on their hands, even when they lived in homes with no smoking allowed indoors.<sup>54-56</sup> A more recent study used a combined assessment of urinary cotinine as the internal biomarker and hand wipe nicotine as the external marker of contact with nicotine pollution in children's microenvironments, and results showed that these distinctive exposure profiles were differentially associated with child health.<sup>57</sup> Specifically, there were associations found between child hand nicotine and illnesses while controlling for cotinine, age, and race/ethnicity, but no association was found while assessing urinary cotinine as the independent variable of interest and while controlling for hand nicotine, age, and race/ethnicity. Thus, future research should consider expanding on this work by adding environmental markers of TSE, such as dust and surface nicotine, nicotelline, and TSNAs as well as hand wipe nicotine levels, to increase the prediction accuracies of the NEG versus TEG versus MEG memberships.

This study also contributes to the gap in the evidence base on the utility of applying machine learning methods to pediatric tobacco control efforts. A recent child TSE study used random forest modeling, but assessed the ability of questionnaires to predict the variation of nicotine

metabolites among Canadian infants.<sup>58</sup> This prior study demonstrated that parent-reported question models predicted 31% and 41% of the respective variation in urinary cotinine and hydroxycotinine;<sup>58</sup> thus, self-reported questionnaire data are useful for situations when serum cotinine and hydroxycotinine biomarkers are not measured. Our study expands on this prior work by assessing nicotine and other tobacco smoke-derived constituents that may be important to distinguish THS from SHS using comprehensive biomarker profiles, such as other widespread sources of TSE pollutants in children that have great implications to their health including NNK,<sup>59,60</sup> PAHs,<sup>61</sup> and VOCs.<sup>62</sup>

This is the first study to differentiate children into the NEG versus the TEG versus the MEG based on comprehensive biomarker profiles and questionnaire measures of TSE from the NHANES 2013-2016, with results generalizable to the U.S. 3-17-year-old population. Several limitations of this secondary analysis should be noted. NHANES is a cross-sectional survey, and therefore longitudinal associations could not be explored. NHANES 2013-2016 questionnaire items did not assess exposure to specific tobacco product types (e.g., cigarettes, cigars, e-cigarettes) or cannabis, and did not assess self-reported tobacco product use among children <12 years old. While we excluded children 12 years old if they had used any tobacco product in the past 5-days and/or if they smoked cigarettes in the past 30-days, social desirability bias may have occurred and children could have not accurately reported their primary tobacco use. We assessed serum cotinine levels to biochemically confirm NEG membership, but children in the TEG or MEG may not have accurately reported their primary tobacco use. Additionally, we found sociodemographic differences between 3-17-year-olds who were included and excluded in this study due to the availability of serum cotinine results for our analysis, which may have presented selection bias. For example, caregivers may have been less likely to agree to serum collection for younger children compared to older children, explaining why included children had a higher mean age of about 10 years old versus eight years old for excluded children. Thus, the original NHANES serum collection may have biased the results of this secondary analysis. Further, while e-cigarette use prevalence among U.S. adults is very low (4%) relative to combustible tobacco product use prevalence (15%),<sup>63</sup> we were unable to assess exposure to e-cigarette aerosol only. Prospective longitudinal studies are encouraged to assess the prevalence, contributions, and health risks of exposure to SHS and THS among children exposed to e-cigarette aerosol only.

In conclusion, the current study highlights that child TSE is complex and can have many routes, toxins, and hazards to children. Findings indicate that random forest modeling using comprehensive biomarker and questionnaire item predictors is a promising tool for predicting TSE group membership. The current study's prediction models with good performance identified significant features of SHS and THS exposure that should be considered by policymakers and researchers to develop practice-based remediation strategies and regulatory actions to reduce hazardous tobacco-related toxicant exposures in children.

## Funding Sources

This work was supported by the National Institute of Environmental Health Sciences (NIH Grant Numbers R21ES032161, R01ES027815, and R01ES030743), National Institute on Drug Abuse (NIH Grant Number

K01DA044313), and the California Tobacco-Related Disease Research Program (TRDRP Grant Number 28PT-0078).

**Disclaimer:**

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Additionally, the findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention/the Agency for Toxic Substances and Disease Registry.

**ABBREVIATIONS**

<b>TSE</b>	tobacco smoke exposure
<b>SHS</b>	secondhand smoke
<b>THS</b>	thirdhand smoke
<b>NEG</b>	no/minimal tobacco smoke exposure group
<b>TEG</b>	predominant thirdhand smoke exposure group
<b>MEG</b>	mixed secondhand and thirdhand smoke exposure group
<b>NNAL</b>	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol
<b>NHANES</b>	National Health and Nutrition Examination Survey
<b>PAHs</b>	polycyclic aromatic hydrocarbons
<b>VOCs</b>	volatile organic compounds
<b>NNK</b>	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone
<b>TNEs</b>	total nicotine equivalents
<b>TSNAs</b>	tobacco-specific nitrosamines
<b>2CyEMA</b>	N-acetyl-S-(2-cyanoethyl)-L-cysteine

**REFERENCES**

- Jacob P 3rd, Benowitz NL, Destailats H, Gundel L, Hang B, Martins-Green M, Matt GE, Quintana PJE, Samet JM, Schick SF, Talbot P, Aquilina NJ, Hovell MF, Mao JH, Whitehead TP. Thirdhand smoke: new evidence, challenges, and future directions. *Chem Res Toxicol.* 2017;30(1):270–294. doi:10.1021/acs.chemrestox.6b00343 [PubMed: 28001376]
- Matt GE, Quintana PJE, Destailats H, Gundel LA, Sleiman M, Singer BC, Jacob P 3rd, Benowitz N, Winickoff JP, Rehan V, Talbot P, Schick S, Samet J, Wang Y, Hang B, Martins-Green M, Pankow JF, Hovell MF. Thirdhand tobacco smoke: emerging evidence and arguments for a multidisciplinary research agenda. *Environ Health Perspect.* 2011;119(9):1218–1226. doi: 10.1289/ehp.1103500 [PubMed: 21628107]
- Sleiman M, Logue JM, Luo W, Pankow JF, Gundel LA, Destailats H. Inhalable constituents of thirdhand tobacco smoke: chemical characterization and health impact considerations. *Environ Sci Technol.* 2014;48(22):13093–13101. doi:10.1021/es5036333 [PubMed: 25317906]
- Schick SF, Farraro KF, Perrino C, Sleiman M, van de Vossenberg G, Trinh MP, Hammond SK, Jenkins BM, Balmes J. Thirdhand cigarette smoke in an experimental chamber: evidence of surface deposition of nicotine, nitrosamines and polycyclic aromatic hydrocarbons and de novo formation

- of NNK. *Tob Control*. 2014;23(2):152–159. doi:10.1136/tobaccocontrol-2012-050915 [PubMed: 23716171]
5. Ramírez N, Vallecillos L, Lewis AC, Borrull F, Marcé RM, Hamilton JF. Comparative study of comprehensive gas chromatography-nitrogen chemiluminescence detection and gas chromatography-ion trap-tandem mass spectrometry for determining nicotine and carcinogen organic nitrogen compounds in thirdhand tobacco smoke. *J Chromatogr A*. 2015;1426:191–200. doi:10.1016/j.chroma.2015.11.035 [PubMed: 26684592]
  6. Hoh E, Hunt RN, Quintana PJE, Zakarian JM, Chatfield DA, Wittry BC, Rodriguez E, Matt GE. Environmental tobacco smoke as a source of polycyclic aromatic hydrocarbons in settled household dust. *Environ Sci Technol*. 2012;46(7):4174–4183. doi:10.1021/es300267g [PubMed: 22397504]
  7. Sleiman M, Gundel LA, Pankow JF, Jacob P 3rd, Singer BC, Destailats H. Formation of carcinogens indoors by surface-mediated reactions of nicotine with nitrous acid, leading to potential thirdhand smoke hazards. *Proc Natl Acad Sci U S A*. 2010;107(15):6576–6581. doi: 10.1073/pnas.0912820107. [PubMed: 20142504]
  8. Benowitz NL, Hukkanen J, Jacob III P. Nicotine chemistry, metabolism, kinetics and biomarkers. *Handb Exp Pharmacol*. 2009;192:29–60. doi: 10.1007/978-3-540-69248-5\_2.
  9. Benowitz NL, Bernert JT, Foulds J, Hecht SS, Jacob P 3rd, Jarvis MJ, Joseph A, Oncken C, Piper ME. Biochemical verification of tobacco use and abstinence: 2019 update. *Nicotine Tob Res*. 2020;22(7):1086–1097. doi:10.1093/ntr/ntz132 [PubMed: 31570931]
  10. Benowitz N, Goniewicz ML, Eisner MD, Lazcano-Ponce E, Zielinska-Danch W, Koszowski B, Sobczak A, Havel C, Jacob P 3rd. Urine cotinine underestimates exposure to the tobacco-derived lung carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone in passive compared with active smokers. *Cancer Epidemiol Biomarkers Prev*. 2010;19(11):2795–2800. doi:10.1158/1055-9965.EPI-10-0497 [PubMed: 20805316]
  11. Ferguson A, Penney R, Solo-Gabriele H. A review of the field on children’s exposure to environmental contaminants: a risk assessment approach. *Int J Environ Res Public Health*. 2017;14(3):265. doi:10.3390/ijerph14030265 [PubMed: 28273865]
  12. World Health Organization. Summary of principles for evaluating health risks in children associated with exposure to chemicals. Geneva, Switzerland: World Health Organization; 2011.
  13. Moya J, Bearer CF, Etzel RA. Children’s behavior and physiology and how it affects exposure to environmental contaminants. *Pediatrics*. 2004;113(4):996–1006. [PubMed: 15060192]
  14. Brody DJ, Lu Z, Tsai J. Secondhand smoke exposure among nonsmoking youth: United States, 2013-2016. *NCHS Data Brief*. 2019;348:1–8.
  15. Wei B, Blount BC, Xia B, Wang L. Assessing exposure to tobacco-specific carcinogen NNK using its urinary metabolite NNAL measured in US population: 2011-2012. *J Expo Sci Environ Epidemiol*. 2016;26(3):249–256. doi:10.1038/jes.2014.88 [PubMed: 25564369]
  16. Jeong SH, Jang BN, Kang SH, Joo JH, Park E. Association between parents’ smoking status and tobacco exposure in school-age children: assessment using major urine biomarkers. *Sci Rep*. 2021;11(1):4536. doi:10.1038/s41598-021-84017-y [PubMed: 33633242]
  17. Park MB. Living with parents who smoke predicts levels of toxicant exposure in children. *Sci Rep*. 2020;10(1):11173. doi:10.1038/s41598-020-66920-y [PubMed: 32636401]
  18. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. *J Multidiscip Healthc*. 2016;9(1):211–217. doi:10.2147/JMDH.S104807 [PubMed: 27217764]
  19. Matt GE, Merianos AL, Quintana PJE, Hoh E, Dodder NG, Mahabee-Gittens EM. Prevalence and income-related disparities in thirdhand smoke exposure to children. *JAMA Network Open*. 2022;5(2):e2147184. doi:10.1001/jamanetworkopen.2021.47184 [PubMed: 35129597]
  20. Schick SF, Glantz S. Concentrations of the carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone in sidestream cigarette smoke increase after release into indoor air: results from unpublished tobacco industry research. *Cancer Epidemiol Biomarkers Prev*. 2007;16(8):1547–1553. doi:10.1158/1055-9965.EPI-07-0210 [PubMed: 17684127]
  21. Benowitz NL, Nardone N, Jain S, Dempsey DA, Addo N, St. Helen G, Jacob P 3rd. Comparison of urine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol and cotinine for assessment of active and passive smoke exposure in urban adolescents. *Cancer Epidemiol Biomarkers Prev*. 2018;27(3):254–261. doi:10.1158/1055-9965.EPI-17-0671 [PubMed: 29475963]

22. Chao MR, Cooke MS, Kuo CY, Pan CH, Liu HH, Yang HJ, Chen SC, Chiang YC, Hu CW. Children are particularly vulnerable to environmental tobacco smoke exposure: evidence from biomarkers of tobacco-specific nitrosamines, and oxidative stress. *Environ Int.* 2018;120:238–245. doi:10.1016/j.envint.2018.08.006 [PubMed: 30103123]
23. Mazumder S, Shia W, Bendik PB, Achilihu H, Sosnoff CS, Alexander JR, Luo Z, Zhu W, Pine BN, Feng J, Blount BC, Wang L. Nicotine exposure in the U.S. population: total urinary nicotine biomarkers in NHANES 2015–2016. *Int J Environ Res Public Health.* 2022;19(6):3660. doi:10.3390/ijerph19063660 [PubMed: 35329347]
24. St. Helen G, Goniewicz ML, Dempsey D, Wilson M, Jacob P 3rd, Benowitz NL. Exposure and kinetics of polycyclic aromatic hydrocarbons (PAHs) in cigarette smokers. *Chem Res Toxicol.* 2012;25(4):952–964. doi:10.1021/tx300043k [PubMed: 22428611]
25. Suwan-ampai P, Navas-Acien A, Strickland PT, Agnew J. Involuntary tobacco smoke exposure and urinary levels of polycyclic aromatic hydrocarbons in the United States, 1999 to 2002. *Cancer Epidemiol Biomarkers Prev.* 2009;18(3):884–893. doi:10.1158/1055-9965.EPI-08-0939 [PubMed: 19258471]
26. St. Helen G, Jacob P 3rd, Peng M, Dempsey DA, Hammond SK, Benowitz NL. Intake of toxic and carcinogenic volatile organic compounds from secondhand smoke in motor vehicles. *Cancer Epidemiol Biomarkers Prev.* 2014;23(12):2774–2782. doi:10.1158/1055-9965.EPI-14-0548 [PubMed: 25398951]
27. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey methods and analytic guidelines. <https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx>
28. Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National Health and Nutrition Examination Survey: sample design, 2011–2014. *Vital Health Stat 2.* 2014;(162):1–33.
29. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey: Analytic Guidelines, 2011-2014 and 2015-2016. Atlanta, GA: Centers for Disease Control and Prevention; 2018.
30. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey 2013–2014 questionnaires, datasets, and related documentation. <https://wwwn.cdc.gov/nchs/nhanes/ContinuousNhanes/Default.aspx?BeginYear=2013>
31. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey 2015–2016 questionnaires, datasets, and related documentation. <https://wwwn.cdc.gov/nchs/nhanes/ContinuousNhanes/Default.aspx?BeginYear=2015>
32. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey (NHANES): MEC laboratory procedures manual. National Center for Health Statistics. January 2013. [https://www.cdc.gov/nchs/data/nhanes/nhanes\\_13\\_14/2013\\_MEC\\_Laboratory\\_Procedures\\_Manual.pdf](https://www.cdc.gov/nchs/data/nhanes/nhanes_13_14/2013_MEC_Laboratory_Procedures_Manual.pdf)
33. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey: MEC laboratory procedures manual. National Center for Health Statistics. January 2016. [https://wwwn.cdc.gov/nchs/data/nhanes/2015-2016/manuals/2016\\_mec\\_laboratory\\_procedures\\_manual.pdf](https://wwwn.cdc.gov/nchs/data/nhanes/2015-2016/manuals/2016_mec_laboratory_procedures_manual.pdf)
34. Tobacco and Volatiles Branch. Laboratory procedures manual: cotinine and hydroxycotinine serum and saliva. Centers for Disease Control and Prevention. <https://wwwn.cdc.gov/nchs/data/nhanes/2017-2018/labmethods/COT-J-MET-508.pdf>
35. Benowitz NL, Dains KM, Dempsey D, Yu L, Jacob P 3rd. Estimation of nicotine dose after low-level exposure using plasma and urine nicotine metabolites. *Cancer Epidemiol Biomarkers Prev.* 2010;19(5):1160–1166. doi:10.1158/1055-9965.EPI-09-1303 [PubMed: 20447913]
36. National Health and Nutrition Examination Survey 2013–2014 data documentation, codebook, and frequencies: cotinine, hydroxycotinine, & other nicotine metabolites and analogs - urine (UCOT\_H). May 2019. [https://wwwn.cdc.gov/Nchs/Nhanes/2013-2014/UCOT\\_H.htm](https://wwwn.cdc.gov/Nchs/Nhanes/2013-2014/UCOT_H.htm)
37. National Health and Nutrition Examination Survey 2015–2016 data documentation, codebook, and frequencies: cotinine, hydroxycotinine, & other nicotine metabolites and analogs - urine (UCOT\_I). September 2019. [https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/UCOT\\_I.htm](https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/UCOT_I.htm)
38. Tobacco and Volatiles Branch. Laboratory procedure manual: cotinine and hydroxycotinine (total). Centers for Disease Control and



Prevention. April 10, 2018. [https://wwwn.cdc.gov/nchs/data/nhanes/2013-2014/labmethods/UCOT\\_H\\_UCOTS\\_H\\_Urinary\\_Cotinine\\_and\\_Hydroxycotinine\\_MET.pdf](https://wwwn.cdc.gov/nchs/data/nhanes/2013-2014/labmethods/UCOT_H_UCOTS_H_Urinary_Cotinine_and_Hydroxycotinine_MET.pdf)

39. Wei B, Feng J, Rehmani IJ, Miller S, McGuffey JE, Blount BC, Wang L. A high-throughput robotic sample preparation system and HPLC-MS/MS for measuring urinary anatabine, anabasine, nicotine and major nicotine metabolites. *Clinica Chim Acta*. 2014;436:290–297. doi:10.1016/j.cca.2014.06.012
40. Bernert JT, Harmon TL, Sosnoff CS, McGuffey JE. Use of cotinine immunoassay test strips for preclassifying urine samples from smokers and nonsmokers prior to analysis by LC-MS-MS. *J Anal Toxicol*. 2005;29(8):814–818. doi:10.1093/jat/29.8.814 [PubMed: 16374940]
41. Hecht SS, Stepanov I, Carmella SG. Exposure and metabolic activation biomarkers of carcinogenic tobacco-specific nitrosamines. *Acc Chem Res*. 2016;49(1):106–114. doi:10.1021/acs.accounts.5b00472 [PubMed: 26678241]
42. Goniewicz ML, Havel CM, Peng MW, Jacob P 3rd, Dempsey D, Yu L, Zielinska-Danch W, Koszowski B, Czogala J, Sobczak A, Benowitz NL. Elimination kinetics of the tobacco-specific biomarker and lung carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol. *Cancer Epidemiol Biomarkers Prev*. 2009;18(12):3421–3425. doi:10.1158/1055-9965.EPI-09-0874 [PubMed: 19959691]
43. Tobacco and Volatiles Branch. Tobacco-specific nitrosamines. Centers for Disease Control and Prevention. [https://wwwn.cdc.gov/nchs/data/nhanes/2013-2014/labmethods/TSNA\\_H\\_MET.pdf](https://wwwn.cdc.gov/nchs/data/nhanes/2013-2014/labmethods/TSNA_H_MET.pdf)
44. Xia B, Xia Y, Wong J, Nicodemus KJ, Xu M, Lee J, Guillot T, Li J. Quantitative analysis of five tobacco-specific N-nitrosamines in urine by liquid chromatography-atmospheric pressure ionization tandem mass spectrometry. *Biomed Chromatogr*. 2014;28(3):375–384. doi:10.1002/bmc.3031 [PubMed: 24127240]
45. Organic Analytical Toxicology Branch. Eight monohydroxy-polycyclic aromatic hydrocarbons: 1-hydroxynaphthalene, 2-hydroxynaphthalene, 2-hydroxyfluorene, 3-hydroxyfluorene, 1-hydroxyphenanthrene, 2- & 3-hydroxyphenanthrene, 1-hydroxypyrene. Centers for Disease Control and Prevention. [https://wwwn.cdc.gov/nchs/data/nhanes/2013-2014/labmethods/PAH\\_H\\_MET\\_Aromatic\\_Hydrocarbons.pdf](https://wwwn.cdc.gov/nchs/data/nhanes/2013-2014/labmethods/PAH_H_MET_Aromatic_Hydrocarbons.pdf)
46. Re-evaluation of some organic chemicals, hydrazine and hydrogen peroxide. Proceedings of the IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, France, 17–24 February 1998. *IARC Monogr Eval Carcinog Risks Hum*. 1999;71(Pt 1):1–315. [PubMed: 10507919]
47. Bhandari D, Zhang L, Zhu W, De Jesús VR, Blount BC. Optimal cutoff concentration of urinary cyanoethyl mercapturic acid for differentiating cigarette smokers from nonsmokers. *Nicotine Tob Res*. 2022;24(5):761–767. doi:10.1093/ntr/ntab224 [PubMed: 34718769]
48. Tobacco and Volatiles Branch. Volatile organic compounds (VOCs) metabolites. Centers for Disease Control and Prevention. [https://wwwn.cdc.gov/nchs/data/nhanes/2013-2014/labmethods/UVOC\\_H\\_MET.pdf](https://wwwn.cdc.gov/nchs/data/nhanes/2013-2014/labmethods/UVOC_H_MET.pdf)
49. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013. <http://www.R-project.org/>.
50. Breiman L Random forests. *Mach Learning*. 2001;45(1):5–32.
51. Matt GE, Quintana PJE, Zakarian JM, Hoh E, Hovell MF, Mahabee-Gittens EM, Watanabe K, Datuin K, Vue C, Chatfield DA. When smokers quit: exposure to nicotine and carcinogens persists from thirdhand smoke pollution. *Tob Control*. 2016;26(5):548–556. doi:10.1136/tobaccocontrol-2016-053119 [PubMed: 27655249]
52. Matt GE, Quintana PJE, Zakarian JM, Fortmann AL, Chatfield DA, Hoh E, Uribe AM, Hovell MF. When smokers move out and non-smokers move in: residential thirdhand smoke pollution and exposure. *Tob Control*. 2011;20(1):e1. doi:10.1136/tc.2010.037382
53. Matt GE, Quintana PJE, Hovell MF, Bernert JT, Song S, Novianti N, Juarez T, Floro J, Gehrman C, Garcia M, Larson S. Households contaminated by environmental tobacco smoke: sources of infant exposures. *Tob Control*. 2004;13(1):29–37. doi: 10.1136/tc.2003.003889 [PubMed: 14985592]
54. Mahabee-Gittens EM, Matt GE, Jandarov RJ, Merianos AL. Hand nicotine and cotinine in children exposed to cigars: a pilot study. *Tob Regul Sci*. 2021;7(3):170–176. doi:10.18001/trs.7.3.2 [PubMed: 34423080]

55. Mahabee-Gittens EM, Merianos AL, Hoh E, Quintana PJE, Matt GE. Nicotine on children's hands: limited protection of smoking bans and initial clinical findings. *Tob Use Insights*. 2019;12:1179173X18823493. doi:10.1177/1179173X18823493
56. Mahabee-Gittens EM, Merianos AL, Matt GE. Preliminary evidence that high levels of nicotine on children's hands may contribute to overall tobacco smoke exposure. *Tob Control*. 2018;27(2):217–219. doi:10.1136/tobaccocontrol-2016-053602 [PubMed: 28360145]
57. Mahabee-Gittens EM, Merianos AL, Jandarov RA, Quintana PJE, Hoh E, Matt GE. Differential associations of hand nicotine and urinary cotinine with children's exposure to tobacco smoke and clinical outcomes. *Environ Res*. 2021;202:111722. doi:10.1016/j.envres.2021.111722 [PubMed: 34297932]
58. Parks J, McLean KE, McCandless L, de Souza RJ, Brook JR, Scott J, Turvey SE, Mandhane PJ, Becker AB, Azad MB, Moraes TJ, Lefebvre DL, Sears MR, Subbarao P, Takaro TK. Assessing secondhand and thirdhand tobacco smoke exposure in Canadian infants using questionnaires, biomarkers, and machine learning. *J Expo Sci Environ Epidemiol*. 2022;32(1):112–123. doi:10.1038/s41370-021-00350-4 [PubMed: 34175887]
59. Merianos AL, Jandarov RA, Mahabee-Gittens EM. Carcinogenic and tobacco smoke-derived particulate matter biomarker uptake and associated healthcare patterns among children. *Pediatr Res*. 2022;10.1038/s41390-022-02031-w. doi:10.1038/s41390-022-02031-w
60. Thomas JL, Guo H, Carmella SG, Balbo S, Han S, Davis A, Yoder A, Murphy SE, An LC, Ahluwalia JS, Hecht SS. Metabolites of a tobacco-specific lung carcinogen in children exposed to secondhand or thirdhand tobacco smoke in their homes. *Cancer Epidemiol Biomarkers Prev*. 2011;20(6):1213–1221. doi:10.1158/1055-9965.EPI-10-1027 [PubMed: 21467230]
61. Murawski A, Roth A, Schwedler G, Schmied-Tobies MIH, Rucic E, Pluym N, Scherer M, Scherer G, Conrad A, Kolossa-Gehring M. Polycyclic aromatic hydrocarbons (PAH) in urine of children and adolescents in Germany – human biomonitoring results of the German Environmental Survey 2014–2017 (GerES V). *Int J Hyg Environ Health*. 2020;226:113491. doi:10.1016/j.ijheh.2020.113491 [PubMed: 32092491]
62. Kuang H, Feng J, Li Z, Tan J, Zhu W, Lin S, Pang Q, Ye Y, Fan R. Volatile organic compounds from second-hand smoke may increase susceptibility of children through oxidative stress damage. *Environ Res*. 2021;207:112227. doi:10.1016/j.envres.2021.112227 [PubMed: 34666018]
63. Cornelius ME, Loretan CG, Wang TW, Jamal A, Homa DM. Tobacco product use among adults - United States, 2020. *MMWR Morb Mortal Wkly Rep*. 2022;71(11):397–405. doi:10.15585/mmwr.mm7111a1 [PubMed: 35298455]

**SYNOPSIS**

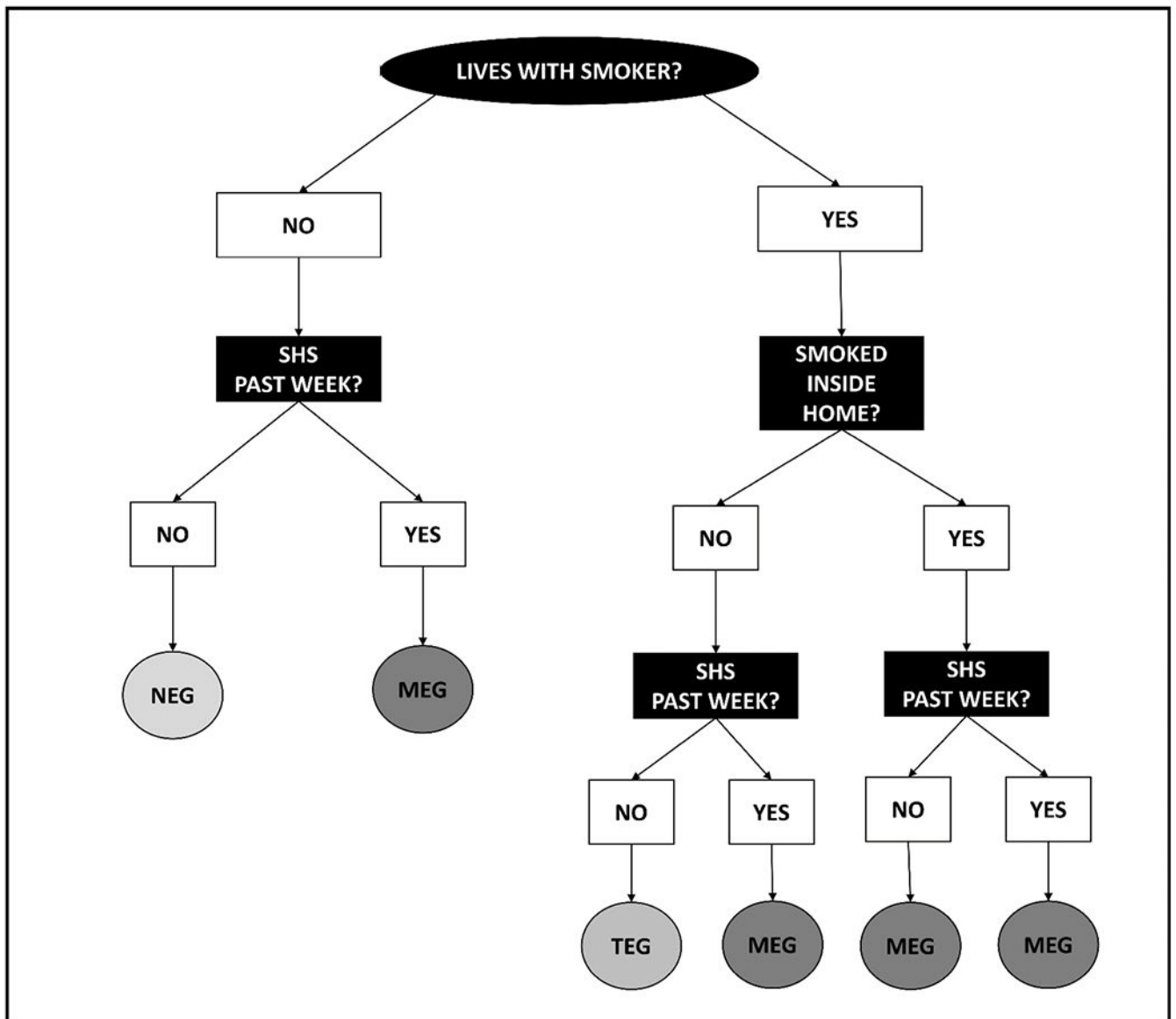
In absence of validated biomarkers for thirdhand tobacco smoke, machine learning methods identified NHANES biomarker and questionnaire data that distinguished children exposed to secondhand and thirdhand smoke with high accuracy.

Author Manuscript

Author Manuscript

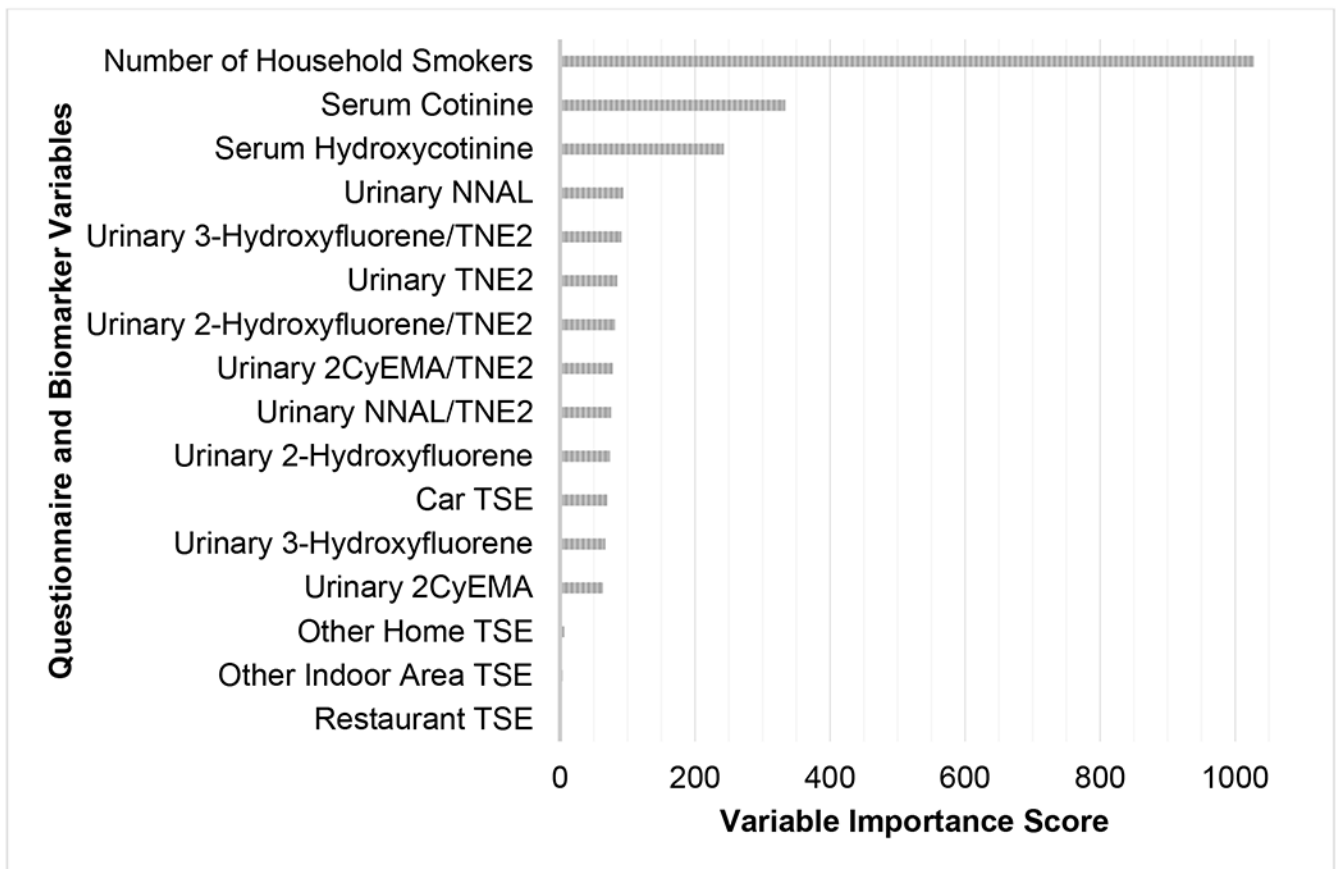
Author Manuscript

Author Manuscript

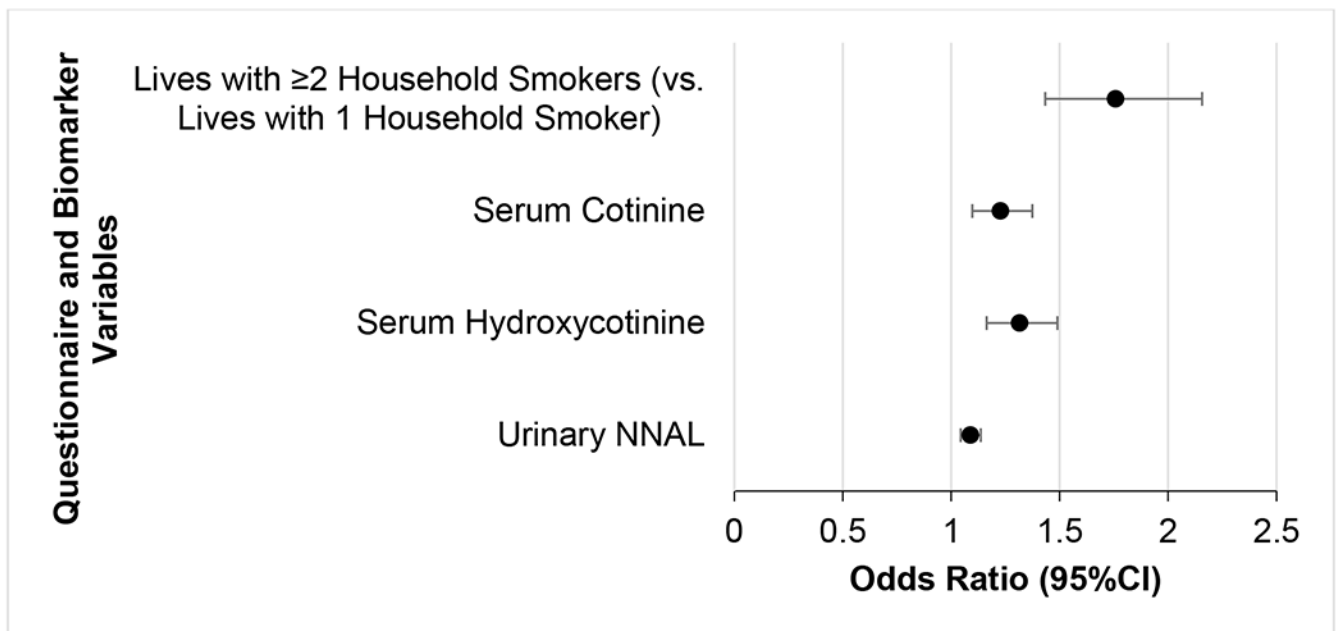


**Figure 1.**

Steps to preliminarily distinguish children into the three TSE groups using NHANES 2013-2016 questionnaire data. SHS in the past week refers to reported exposure patterns in locations other than the home (e.g., car). Additionally, serum cotinine levels were assessed to biochemically confirm NEG had no/minimal cotinine levels.



**Figure 2.** Variables by importance in the main random forest model predicting child TSE group membership using NHANES 2013-2016 data.



**Figure 3.**

Forest plot of logistic regression model results using the top four variables of importance in the main random forest model to distinguish between the two TSE groups of MEG versus TEG membership (reference category) using NHANES 2013-2016 data. The overall model prediction accuracy was 79%, and the accuracies of predicting MEG and TEG membership were 86% and 74%, respectively.

**Table 1.** Sociodemographic Characteristics Overall and based on Child TSE Group Membership, NHANES 2013-2016

Characteristic	Overall (N=4,485)	TSE Group <sup>a</sup>			p-value
		NEG (n=3,369)	TEG (n=701)	MEG (n=415)	
<b>Child Age, M (SE)</b>	10.5 (0.01)	10.5 (0.01)	10.4 (0.03)	10.6 (0.07)	0.698
<b>Child Sex</b>					0.826
Male	2,297 (51 [49-53])	1,724 (51 [49-54])	355 (51 [46-55])	218 (52 [46-58])	
Female	2,188 (49 [47-51])	1,645 (49 [47-51])	346 (49 [45-54])	197 (48 [42-54])	
<b>Child Race/Ethnicity</b>					<0.001
Non-Hispanic White	1,157 (51 [49-53])	826 (50 [49-53])	212 (53 [49-57])	119 (50 [44-56])	
Non-Hispanic Black	1,068 (14 [13-15])	700 (12 [11-13])	163 (13 [11-15])	205 (31 [26-36])	
Non-Hispanic Other/Multiracial	663 (10 [9-11])	523 (10 [9-11])	112 (12 [10-15])	28 (6 [3-9])	
Hispanic	1,597 (25 [24-27])	1,320 (28 [26-29])	214 (22 [19-25])	63 (13 [10-16])	
<b>Caregiver Age, M (SE)</b>	41.3 (0.03)	41.5 (0.04)	40.4 (0.19)	40.9 (0.43)	0.002
<b>Caregiver Sex</b>					<0.001
Male	2,032 (50 [48-52])	1,606 (52 [50-54])	289 (45 [40-50])	137 (42 [36-48])	
Female	2,453 (50 [48-52])	1,763 (48 [46-50])	412 (55 [51-60])	278 (58 [52-64])	
<b>Caregiver Education Level</b>					<0.001
High school graduate/equivalent	2,071 (40 [38-47])	1,472 (36 [34-38])	358 (47 [42-51])	241 (60 [54-66])	
Some college	1,386 (33 [31-35])	992 (31 [29-33])	239 (38 [34-43])	155 (36 [30-42])	
College graduate	917 (27 [26-29])	819 (33 [30-35])	83 (15 [11-18])	15 (4 [1-6])	
<b>FPL<sup>c</sup></b>					<0.001
<185%	2,496 (43 [42-45])	1,757 (40 [38-42])	405 (47 [43-52])	334 (72 [65-78])	
185-349%	900 (25 [23-27])	688 (25 [23-27])	156 (28 [24-33])	56 (19 [13-25])	
350%	804 (27 [25-28])	697 (30 [28-33])	92 (19 [14-23])	15 (7 [3-12])	
Unspecified	285 (5 [4-6])	227 (5 [4-6])	48 (6 [4-8])	10 (2 [1-3])	
<b>No. Household Members, M (SE)</b>	4.8 (0.01)	4.7 (0.01)	4.9 (0.01)	4.5 (0.01)	0.002

Characteristic	TSE Group <sup>a</sup>			p-value
	Overall (N=4,485)	NEG (n=3,369)	TEG (n=701)	
<b>No. Child Household Members</b>				0.002
1 child	696 (17 [16-19])	507 (17 [15-18])	113 (19 [15-23])	76 (23 [17-29])
2 children	1,415 (35 [33-36])	1,117 (36 [34-38])	192 (29 [25-33])	106 (34 [28-41])
3 children	2,374 (48 [46-50])	1,745 (47 [45-50])	396 (52 [48-57])	233 (43 [37-49])
<b>Family Home Ownership Status</b>				<0.001
Owens Home/Being Bought	2,291 (61 [59-63])	1,823 (65 [63-67])	324 (51 [46-55])	144 (46.4 [40-53])
Rents Home/Other Arrangement	2,184 (39 [37-41])	1,536 (35 [34-37])	377 (49 [45-54])	271 (53.6 [47-60])
<b>No. Household Rooms, M (SE)</b>	6.7 (0.01)	6.9 (0.01)	6.4 (0.01)	6.0 (0.02)

Abbreviations: TSE, tobacco smoke exposure; NEG, no/minimal TSE group; TEG, predominant thirdhand smoke exposure group; MEG, mixed secondhand and thirdhand smoke exposure group; CI, confidence interval; FPL, federal poverty level.

<sup>a</sup>TSE group membership was determined using questionnaire data only. Serum cotinine levels were assessed to biochemically confirm NEG had no/minimal cotinine levels.

<sup>b</sup>n (% [95%CI]) refers to unweighted count and weighted column percent and 95% CIs unless noted otherwise.

<sup>c</sup>FPL is the ratio of family income to annual poverty thresholds that accounts for family size and geographic location to determine financial eligibility of federal programming.



**Table 2.**

TSE Patterns Overall and based on Child TSE Group Membership, NHANES 2013–2016

TSE Variable	Overall	TSE Group <sup>a</sup>		
		NEG	TEG	MEG
	n (% [95%CI]) <sup>b</sup>	n (% [95%CI]) <sup>b</sup>	n (% [95%CI]) <sup>b</sup>	n (% [95%CI]) <sup>b</sup>
<b>Number of Household Smokers</b>				
0 smokers	3,369 (76 [75-78])	3,369 (100)	-	-
1 smoker	709 (14 [13-15])	-	495 (68 [64-73])	214 (43 [38-49])
2 smokers	407 (10 [9-11])	-	206 (32 [28-36])	201 (57 [51-63])
<b>Car TSE</b>				
No	3,916 (88 [87-89])	3,187 (96 [95-97])	544 (73 [69-77])	185 (40 [35-47])
Yes	510 (12 [11-13])	134 (4 [3-5])	148 (27 [23-31])	228 (60 [54-66])
<b>Other Home TSE</b>				
No	4,143 (94 [93-95])	3,158 (95 [94-96])	647 (94 [92-96])	338 (83 [79-87])
Yes	272 (6 [5-7])	157 (5 [4-6])	42 (6 [4-8])	73 (17 [13-22])
<b>Restaurant TSE</b>				
No	4,399 (99 [99-100])	3,300 (99 [98-100])	687 (99 [98-100])	412 (100 [99-100])
Yes	25 (1 [0-1])	19 (1 [0-1])	5 (1 [0-1])	1 (0 [0-0])
<b>Other Indoor Area TSE</b>				
No	4,309 (97 [96-98])	3,258 (98 [97-99])	659 (95 [92-97])	392 (95 [92-97])
Yes	110 (3 [2-4])	63 (2 [1-3])	29 (5 [3-8])	18 (5 [3-8])

Abbreviations: TSE, tobacco smoke exposure; NEG, no/minimal TSE group; TEG, predominant thirdhand smoke exposure group; MEG, mixed secondhand and thirdhand smoke exposure group; CI, confidence interval.

<sup>a</sup>TSE group membership was determined using questionnaire data only. Serum cotinine levels were assessed to biochemically confirm NEG had no/minimal cotinine levels.

<sup>b</sup>n (% [95%CI]) refers to unweighted count and weighted column percent and 95%CIs.

**Table 3.** Biomarker and Biomarker Ratio Levels Overall and based on Child TSE Group Membership, NHANES 2013-2016

Biomarker Variable	n (%) <sup>a</sup>	Overall GeoM (95%CI)	TSE Group Membership		
			NEG GeoM (95%CI)	TEG GeoM (95%CI)	MEG GeoM (95%CI)
<b>Serum Nicotine Metabolites</b>					
Serum Cotinine (ng/mL)	4,485 (100)	0.05 (0.04, 0.05)	0.03 (0.02, 0.03)	0.11 (0.10, 0.13)	0.93 (0.78, 1.11)
Serum Hydroxycotinine (ng/mL)	4,485 (100)	0.02 (0.02, 0.03)	0.02 (0.01, 0.02)	0.04 (0.04, 0.05)	0.29 (0.25, 0.34)
<b>Urinary Nicotine Metabolites</b>					
Urinary TNE2 (nmol/mL)	3,048 (68)	0.01 (0.01, 0.02)	0.01 (0.01, 0.02)	0.02 (0.02, 0.03)	0.15 (0.11, 0.20)
<b>Urinary TSNAs</b>					
NNAL (pg/mL)	2,636 (59)	1.45 (1.35, 1.55)	1.01 (0.95, 1.07)	2.17 (1.78, 2.65)	11.36 (9.18, 14.04)
NNAL/TNE2	3,910 (87)	164.87 (160.90, 168.93)	210.29 (203.66, 217.14)	107.14 (85.63, 134.06)	69.82 (64.77, 75.26)
<b>Urinary PAHs</b>					
2-Hydroxyfluorene (ng/L)	2,996 (67)	131.25 (126.28, 138.51)	129.59 (122.99, 136.54)	121.05 (106.70, 137.33)	177.60 (153.23, 205.83)
3-Hydroxyfluorene (ng/L)	2,996 (67)	62.87 (59.95, 65.93)	61.41 (58.19, 64.80)	56.58 (49.79, 64.29)	88.87 (76.23, 103.61)
2-Hydroxyfluorene/TNE2	3,048 (68)	16,034.02 (15,748.03, 16,235.20)	27,654.98 (27,198.56, 28,119.06)	5,619.66 (5,077.46, 6,219.76)	1,020.68 (878.98, 1185.23)
3-Hydroxyfluorene/TNE2	3,048 (68)	7,547.62 (7,413.08, 7,684.61)	13,048.17 (12,833.79, 13,266.14)	2,583.34 (2,347.40, 2,842.98)	490.87 (417.23, 577.51)
<b>Urinary VOCs</b>					
2CyEMA (ng/mL)	3,018 (67)	1.28 (1.22, 1.36)	1.10 (1.04, 1.17)	1.35 (1.17, 1.57)	3.69 (3.02, 4.51)
2CyEMA/TNE2	3,060 (68)	154.47 (152.06, 156.92)	240.61 (236.68, 244.62)	60.15 (55.05, 65.73)	20.13 (17.49, 23.16)

Abbreviations: TSE, tobacco smoke exposure; NEG, no/minimal TSE group; TEG, predominant thirdhand smoke exposure group; MEG, mixed secondhand and thirdhand smoke exposure group; GeoM, geometric mean; CI, confidence interval; TNE2, total nicotine equivalents 2; TSNAs, tobacco-specific nitrosamines; NNAL, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol; PAHs, polycyclic aromatic hydrocarbons; VOCs, volatile organic compounds; 2CyEMA, N-acetyl-S-(2-cyanoethyl)-L-cysteine.

<sup>a</sup> n (%) refers to the unweighted count and weighted row percent of the imputed value.

**Table 4.** Random Forest Model Accuracy for Predicting Child TSE Group Membership, NHANES 2013-2016

	Variable Selection for Each Prediction Model	No. Variables in Model	Model Prediction Accuracy (%)			
			TSE Group Membership			
			Overall	NEG	TEG	MEG
<b>Main Prediction Model</b>						
All Reported TSE and Biomarker Variables		16	95	100	88	71
<b>Sensitivity Prediction Models</b>						
All Reported TSE Variables		5	93	100	91	26
Reported TSE Variables Excluding Number of Household Smokers		4	76	95	1	55
All Biomarker Variables Excluding Ratios		7	75	94	9	52
Biomarker and Biomarker Ratio Variables Excluding Serum Cotinine, Serum Hydroxycotinine, Urinary TNE2, and Urinary NNAL		7	72	90	12	29
All Biomarker and Biomarker Ratio Variables		11	71	80	35	67
All Biomarker Ratio Variables		4	64	78	13	17

Abbreviations: TSE, tobacco smoke exposure; NEG, no/minimal TSE group; TEG, predominant thirdhand smoke exposure group; MEG, mixed secondhand and thirdhand smoke exposure group; TNE2, total nicotine equivalents 2; NNAL, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol.