Llasmin Lopez
UC Davis

Exposure to E-cigarette Aerosol Project

Throughout the years, we've explored and acquired a substantial amount of information regarding the impact of tobacco use via cigarettes and exposure to smoke produced by cigarettes on health. We now see the rise of a new nicotine product, the electronic cigarette (e-cigarette). Many beliefs surrounding the e-cigarette suggest it may be a relatively healthier alternative to the traditional cigarette; however, a look at common ingredients found in e-cigarettes calls for further investigation of the exposure to e-cigarette aerosol.

Using data from the 2017-2018 National Health and Nutrition Examination Survey (NHANES), we focus on `demo_j`, `cot_j`, `smqrtu_j`, and `smqshs_j` which contain 9254, 7936, 6401, and 9254 observations respectively. Since all four datasets have information of the respondents by their sequence numbers in column '`seqn`',we merge the datasets by '`seqn`' to create the new data set '`new_data`'; which has a total of 9254 observations. If we further restrict our dataset to those who were 12 years old or older and were both interviewed and had a physical examination, we now have 6401 observations.

Table 1 provides sample size and percent of the population for each of the following six categories: Family Income, Gender, Race/ Ethnicity, Age, Used tobacco in the past five days, Exposed to e-cigarette aerosol within the last seven days.

Table 1. Characteristics of the NHANES population, 2017-2018 (n=6401)

Characteristic	Sample size/ frequency Percent
Family Income	n=5577
<1.00	1124
1.00-<2.00	20.15
2.00-<3.00	1603
3.00-<4.00	28.74
>=4.00	900
	16.14
	593
	10.63
	1357
	24.33
Gender	n=6401
Men	3115
Women	48.66
	3286
	51.34

Race/ ethnicity	n=6401	
Mexican American	907	14.17
Other Hispanic	581	9.08
Non-Hispanic White	2156	33.68
Non-Hispanic Black	1492	23.31
Non-Hispanic Asian	896	14.00
Other or multi-race	369	5.76

Age(years)	n=6401	
12-17	868	13.56
18-24	650	10.15
25-35	892	13.94
>35	3991	62.35
Used tobacco in the past 5 days	n=5952	
Yes	1211	20.35
No	4741	79.65
Exposed to e-cigarette aerosol Yes	n=6149	
	782	12.72
No	5367	87.28

In Table 2, we find the total quartiles of cotinine and then the quartiles of cotinine in people with and without exposure to e-cigarette aerosol. We tested Cotinine for normality and concluded Cotinine was not normally distributed. Then, to compare the two categories, Cotinine

and Exposure to e-cigarette aerosol, we use the nonparametric, Wilcoxon test.

Table 2. Cotinine level of adults with and without exposure to e-cigarette aerosol, NHANES 2017-2018

	Cotinine (ng/mL)		
	25th percentile	50th percentile (median) 75th percentile	p-value
Total	0.011	0.032 1.010	
Exposure to e-cigarette aerosol Yes No	0.0320 0.011	0.3685 147.5000 0.026 0.376	<0.01

Next, in Table 3 we found the percent of people exposed to e-cigarette aerosol by characteristics such as Age, Family Income, Gender, Race/ Ethnicity, and Used tobacco within the last five days. We used the Chi-squared test of association between each previously listed characteristic and exposure to e-cigarette aerosol.

Table 3. Percent of people exposed to e-cigarette aerosol by characteristic, NHANES 2017-2018

Characteristic	Percent exposed to e-cigarette aerosol p-value
Total	12.72 NA
Age(years) 12-17 18-24 25-35 >35	<0.0001 12.97 25.58 18.65 9.25

Family Income		0.1093
<1.00	14.49	
1.00-<2.00	12.33	
2.00-<3.00	14.75	
3.00-<4.00	12.13	
>=4.00	11.65	
Gender		<0.0001
Men	14.49	
Women	11.06	
Race/ ethnicity		<0.0001
Mexican American	9.82	
Other Hispanic	9.62	
Non-Hispanic White	15.47	
Non-Hispanic Black	12.00	
Non-Hispanic Asian	9.79	
Other or multi-race	18.31	
Used tobacco in the past 5 days		<0.0001
Yes	24.33	
No	10.16	

Taking a closer look at our tables, we notice in Table 1 that a majority of the study sample is older than 35 years old and that participants tend to lie either below a 2.0 or above a 4.0 in terms of family income scores. This should be noted as the target audience age for e-cigarettes is believed to be late teens/ early twenties and we did not expect to see a relatively high portion of the sample with a family income score above a 4.0. In Table 2, our Wilcoxon test returned a p-value below 0.01, indicating that there was a statistically significant difference in cotinine levels between those who were exposed to e-cigarette aerosols in the past seven days and those who were not exposed to e-cigarette aerosols in the past seven days. Finally, Table 3 indicates that family income level was not statistically significant on exposure to e-cigarette aerosol while the rest of the characteristics are statistically significant. This means we may be able to narrow our focus on motivating factors of e-cigarette aerosol exposure and exclude family income levels in future studies.

SAS CODE:

```
/*Step 1*/
        proc contents data=sasuser.demo_j;
        run;
        proc contents data=sasuser.cot_j;
        run;
        proc contents data=sasuser.smqrtu_j;
        run;
        proc contents data=sasuser.smqshs_j;
        run;

/*Step 2: Merging*/
        /*since we are merging, we need to sort the data first*/
        proc sort data=sasuser.demo_j out=sorted_demo_j nodupkey dupout = extra_demo;
        by seqn;
        run;
        proc sort data=sasuser.cot_j out=sorted_cot_j nodupkey dupout = extra_cot;
        by seqn;
        run;
        proc sort data=sasuser.smqrtu_j out=sorted_smqrtu_j nodupkey dupout = extra_smqrtu;
        by seqn;
        run;
        proc sort data=sasuser.smqshs_j out=sorted_smqshs_j nodupkey dupout = extra_smqshs;
        by seqn;
        run;

        /*now we merge sorted data*/
        data new_data;
        merge sorted_demo_j sorted_cot_j sorted_smqrtu_j sorted_smqshs_j;
        by seqn;
        run;
        /*check observations in new data*/
        proc contents data=new_data;
        run;

/*Step 3*/
        /*new data with restrictions*/
        data ndata;
```

```
set new_data;
where ridstatr=2 and ridageyr>= 12;
run;
/*check observations in new restricted data*/
proc contents data=ndata;
run;

/*Step 4*/
/*preparing variables*/
proc format;
value famf 1='-<1.00'
           2='1.00-<2.00'
           3='2.00-<3.00'
           4='3.00-<4.00'
           5='>= 4.00';

value ref 1='Mexican American'
          2='Other Hispanic'
          3='White'
          4='Black'
          6='Asian'
          7='Other/Multi-race';

value gndrf 1='men'
            2='women';

value agef 1='12-17'
           2='18-24'
           3='25-35'
           4='>35';

value tobf 1='Used tobacco in last 5 days'
           2='Did not use tobacco in last 5 days';

value xpsdf 1='Exposed to e-cig in last 7 days'
            2='Was not exposed to e-cig in last 7 days';
run;
```

```
data tobacco_data;
set ndata;

if indfmpir=. then familyIL = .;
else if indfmpir <1.00 then familyIL = 1;
else if indfmpir >=1.00 and indfmpir <2.00 then familyIL = 2;
else if indfmpir >=2.00 and indfmpir <3.00 then familyIL = 3;
else if indfmpir >=3.00 and indfmpir <4.00 then familyIL = 4;
else familyIL = 5;

if ridageyr = . then age=.;
else if ridageyr in (12,13,14,15,16,17) then age =1; else if ridageyr
in (18,19,20,21,22,23,24) then age =2; else if ridageyr in
(25,26,27,28,29,30,31,32,33,34,35) then age =3; else if ridageyr
>35 then age =4;

if smdany in (7,9,.) then tobacco=.;
else if smdany=1 then tobacco=1;
else if smdany=2 then tobacco=2;

if smq940=1 then exposed=1;
else if smq940=2 then exposed=2;
else exposed=.;

          /*new label only for new variables*/
label familyIL = 'Family income level'
                    age = 'Age 12-17, 18-24, 25-35, or >35'
                    tobacco = 'Used tobacco in the past 5 days'
                    exposed = 'Exposed to e-cig in the past 7 days';

format familyIL famf. ridreth3 ref. riagendr gndrf.
       age agef. tobacco tobf. exposed xpsdf.;
run;

/*check new variables*/
proc freq data=tobacco_data;
tables indfmpir*familyIL/list missing;
run;
```

```
proc freq data=tobacco_data;
tables ridageyr*age/list missing;
run;
proc freq data=tobacco_data;
tables smdany*tobacco/list missing;
run;
proc freq data=tobacco_data;
tables smq940*exposed/list missing;
run; /*all looks good!*/

/*Step 5*/
/*table 1*/
proc freq data= tobacco_data;
tables familyIL*familyIL/chisq;
run;
proc freq data=tobacco_data;
tables riagendr*riagendr/chisq;
run;
proc freq data= tobacco_data;
tables ridreth3*ridreth3/chisq;
run;
proc freq data= tobacco_data;
tables age*age/chisq;
run;
proc freq data= tobacco_data;
tables tobacco*tobacco/chisq;
run;
proc freq data= tobacco_data;
tables exposed*exposed/chisq;
run;

/*table 2*/
proc univariate data=tobacco_data plot normal;
var lbxcot;
                                                    run; /*not normal, so need non-parametric test*/

proc npar1way data=tobacco_data wilcoxon;/*two categories so we use wilcoxon*/
class exposed;
```

```
var lbxcot;
run;

proc univariate data=tobacco_data; /*percentages*/
class exposed;
var lbxcot;
run;
/*table 3*/
proc freq data = tobacco_data;
table (age familyIL riagendr
ridreth3 tobacco)*exposed/chisq;
run;
```