Prevalence and Demographics of Smoking and E-cigarettes Use (2017)

By Mannat Sandhu

Abstract

Purpose: Analyze the prevalence and demographics of smokers and e-cigarette users.
Measures: Smoking and E-Cigarette use status were outcome variables. Individual-level demographic features (age, race/ethnicity, sex, education, income and employment status) were used as inputs.
Design: Data analysis of cross-sectional data from the 2017 Behavioral Risk Factor Surveillance System (BRFSS).
Analysis: Data visualization, multiple logistic regressions.
Results: Prevalence of current e-cigarette users decreased with increasing age, unlike smokers. The prevalence of current e-cigarette users and smokers is more in males than in females, highest among people who are out of work/unable to work, who have the highest education level as Grades 9-11, and decreases with increase in income. Fewer percentage of Black/African Americans and Native Americans are current e-cigarette users than Whites, but they are less prone to quit e-cigarette smoking.

Purpose

This study analyzes the prevalence and demographics of smokers and e-cigarette users, and compares and contrasts findings between these groups to identify meaningful differences.

Measurements

Two tobacco product types[1] were assessed in the analysis - cigarettes and e-cigarettes. For each of these, 4 categories were constructed - current daily user, current non-daily user, former user, never user.

E-Cigarette users -

- Current daily user - These are respondents who reported having used E-cigarettes in their lifetime and now use E-cigarettes every day.
- Current non-daily user - Respondents who reported having used E-cigarettes in their lifetime and now use E-cigarettes some days.
- Former user - Respondents who reported having used E-cigarettes in their lifetime and currently do not use E-cigarettes.
- Never user - Respondents who reported they had not used E-cigarettes in their lifetime.

[1] I have not assessed smokeless tobacco products, although the analysis for that can follow along similar lines.

Respondents who did not know if they had used E-cigarettes/refused to answer/or had missing responses were excluded from this analysis. The first two categories (Current daily user or Current non-daily user) are classified “Current E-Cigarette User” for the logistic regression analysis.

Cigarette users -

- Current daily smoker - Respondents who reported having smoked at least 100 cigarettes in their lifetime and now smoke every day.
- Current non-daily user - Respondents who reported having smoked at least 100 cigarettes in their lifetime and now smoke some days.
- Former user - Respondents who reported having smoked at least 100 cigarettes in their lifetime and currently do not smoke.
- Never user - Respondents who reported they had not smoked at least 100 cigarettes in their lifetime.

Respondents who did not know if they had smoked/refused to answer/or had missing responses were excluded from this analysis. The first two categories (Current daily smoker or Current non-daily smoker) are classified “Current Smoker” for the logistic regression analysis.

Design

I have used a two fold approach for this problem -

First, I visualize the prevalence of smokers and e-cigarette users based on various demographic features (age, race, sex, income, education, employment status). Here, I look at the current daily and non-daily users only, for smoking and e-cigarette use. The plot shows the actual values, with 95% CI intervals.

Next, I run multiple logistic regressions to identify meaningful differences between the demographic features between smokers and e-cigarette users. The 4 logistic regressions that I have run are -

1. Current Smokers Vs Never Smokers
2. Current Smokers Vs Former Smokers
3. Current E-Cigarette Users Vs Never E-Cigarette Users
4. Current E-Cigarette Users Vs Former E-Cigarette Users

Here the outcome variables are binary, and each of the input features are treated as factor variables, even in cases where there may be some natural ordering (e.g., in case of age buckets or income buckets). The results are shown in a table[2] at the end.

This treatment as a supervised learning classification problem allows us to study the effect of each of the demographic variables with respect to other buckets of the same feature, while controlling for other demographic variables.

The reason I have used this approach (similar to multinomial logistic regression) instead of an ordered probit like approach is because there is no inherent ordering of the 4 categories of the dependent variable that we have defined earlier.

[2] referred to as LR Table (i.e., logistic regression table) in the rest of the analysis.

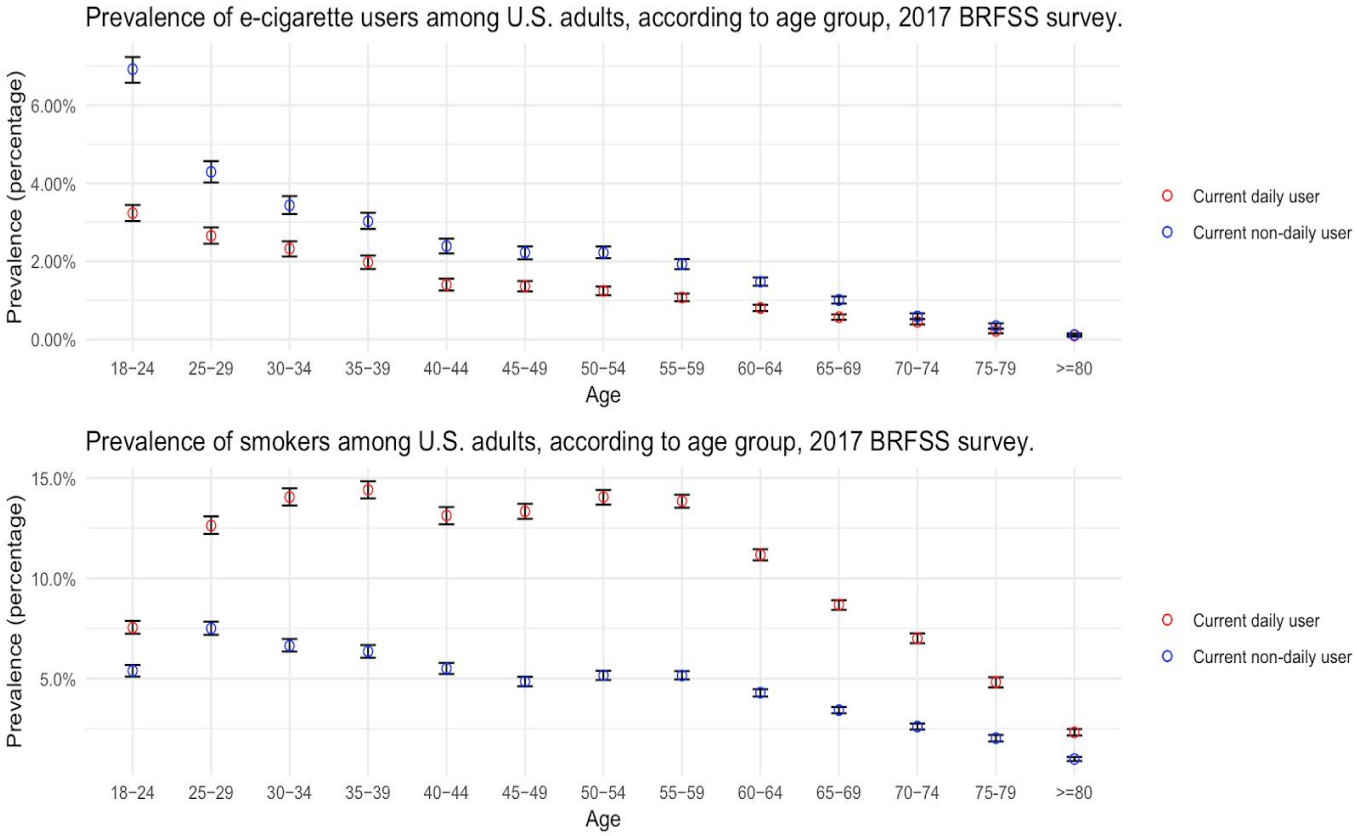
I use the odds of being a smoker/e-cigarette user (vs the odds of the other categories) based on each demographic feature to complement the qualitative analysis done by the visualizations.

The logistic regression analysis considers categories 1 and 2 together, with respect to categories 3 and 4 individually, whereas the visualization considers categories 1 and 2 alone independently.

In the end, I have chosen to analyze and visualize the state-specific prevalence of smokers and e-cigarette users, from the BRFSS data itself. I think this analysis is useful because it can highlight the state-level differences in tobacco policies.

Analysis of E-cigarette users and smokers

Patterns across age groups

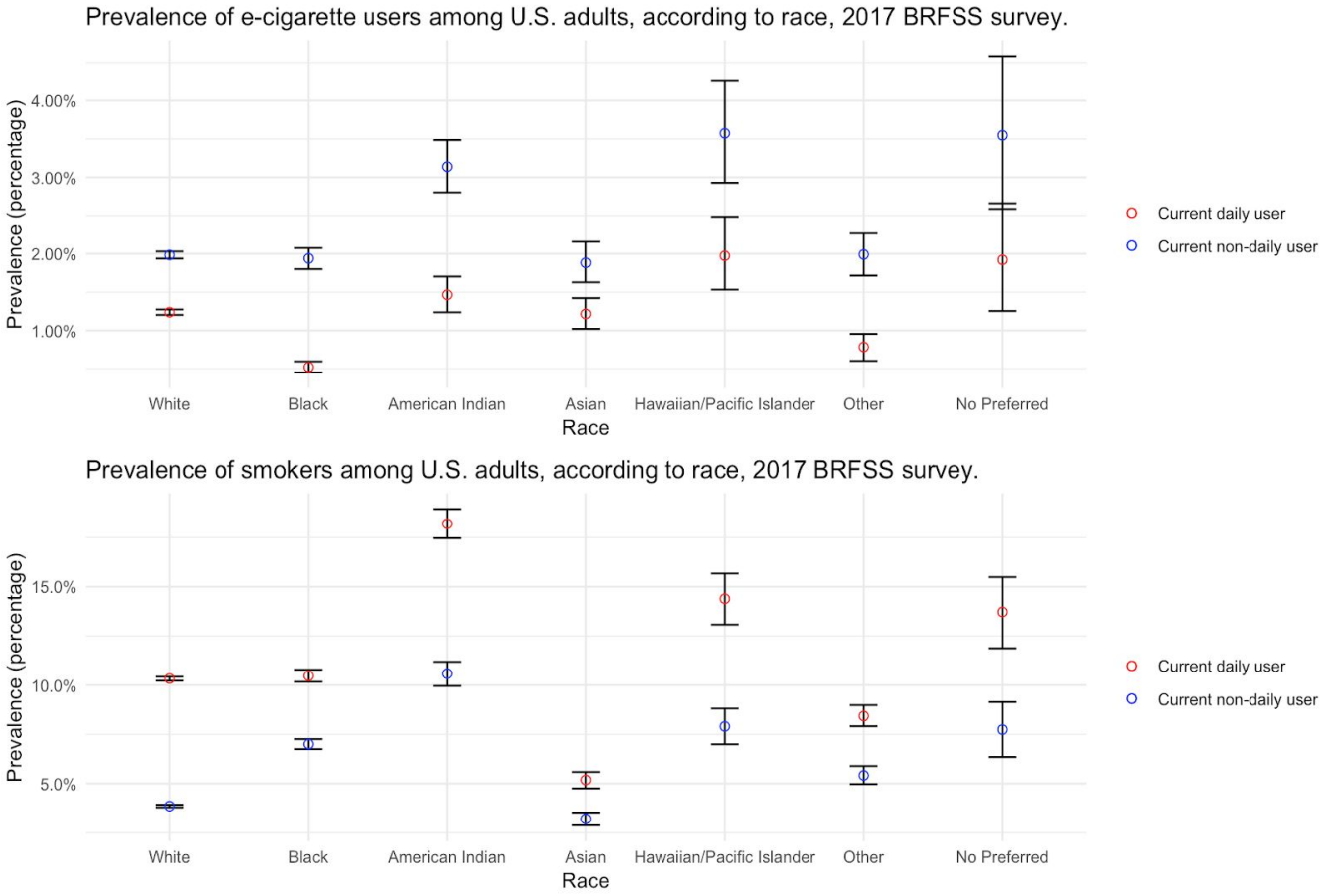


First, as shown in the following figure, the prevalence of current daily and non-daily e-cigarette users decreased with increasing age, with the prevalence of current e-cigarette as highest among those aged 18-24 years. This is in contrast to cigarette smokers, where there is no such pattern.

This can also be seen from the LR table, where the odds ratio of being a e-cigarette user as opposed to a never e-cigarette user decrease by 0.30 if you are in the 25-29 age group, as opposed to being in the 18-24 age group. They decrease even more, to 0.46 in the 30-34 age group and 0.54 in the 35-39 age group. A similar trend is followed by the e-cigarette users vs former e-cigarette users odds ratios.

Secondly, for all age groups, the number of current non-daily users are more than the current daily users. In the case of cigarette smokers, the current daily users are more than the current non-daily users across age groups.

Patterns across Race Groups



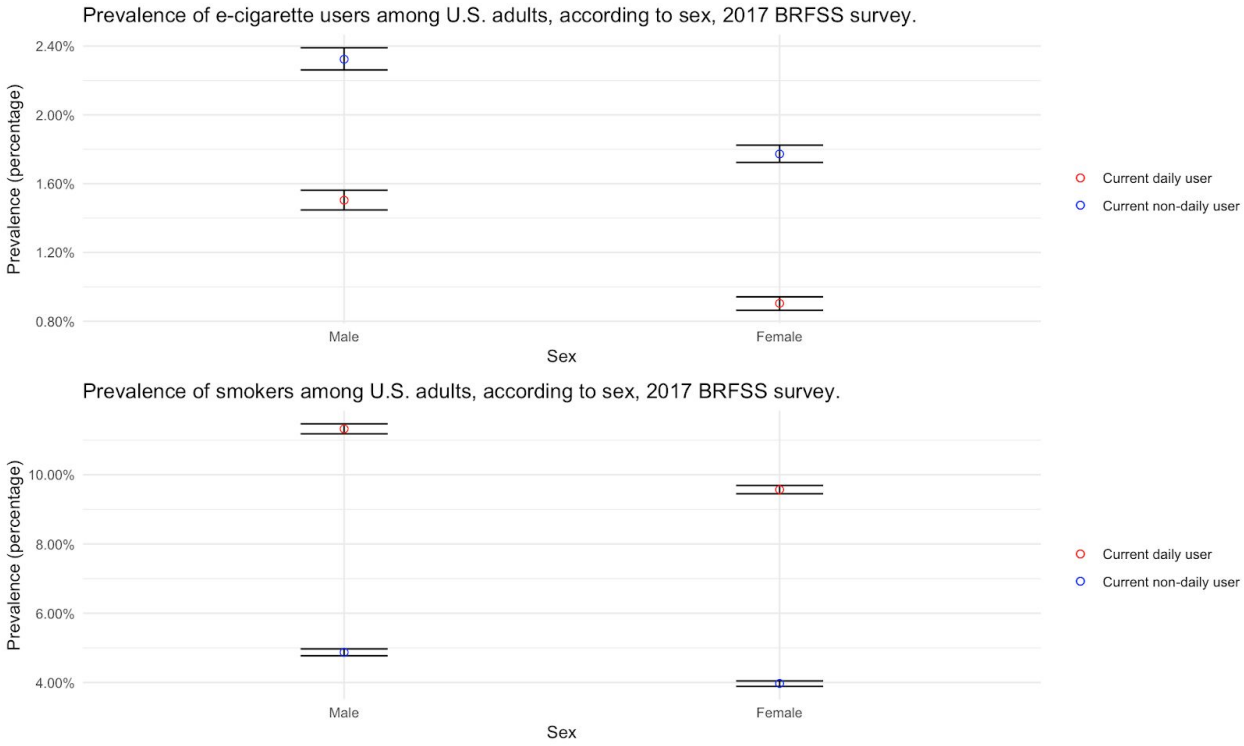
Here, we can see that there is a significant drop in the daily users of e-cigarettes for Black/African American and Other race categories. From the LR table comparison of current e-cigarette user vs a never e-cigarette user, being a Black/African American decreases your odds ratio of being a current e-cigarette user by 0.54 than if you are from White race category. However, being in a Black/African American race increases your odds ratio of being a current e-cigarette users vs a former e-cigarette users than being from White race category by 0.34. This is a similar trend as with American Indian race population with respect to White category.

In other words, with respect to White race category, there are fewer percentage of Black/African American and Native American people who are current e-cigarette users, but they are less prone to quit e-cigarette smoking.

In case of smokers, the American Indian category has a high increase in daily and non-daily smokers, whereas Asians have a low percentage of daily and non-daily users.

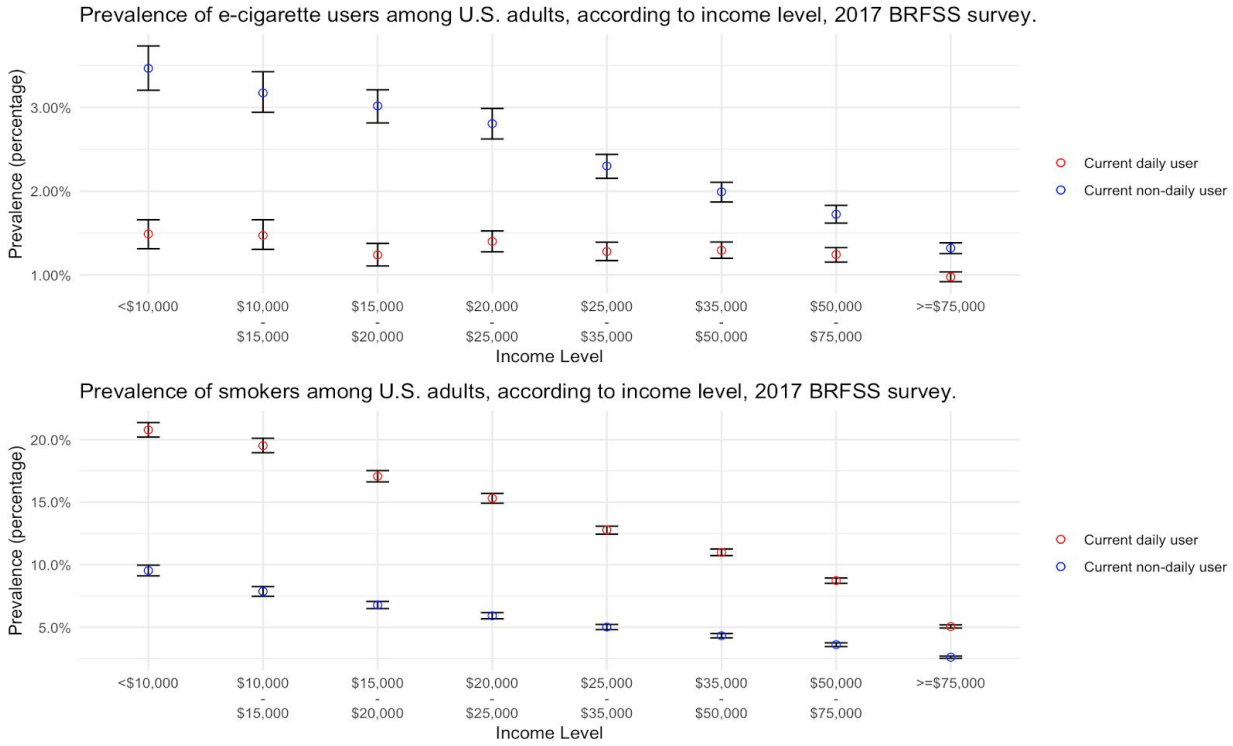
From the LR table, we can see that being an American Indian increases your odds ratio of being a current smoker than a never smoker by 0.32 more than being from a White race category, and 0.80 more than being a Asian race category.

Patterns across Sex



The prevalence of current e-cigarette users and smokers is more in males than in females. However, being a female increases your odds ratio of being a current e-cigarette users than a former e-cigarette user by 0.08 (statistically significant at 95% CI). That is, females are less likely to quit e-cigarette smoking than males, albeit by a slight margin.

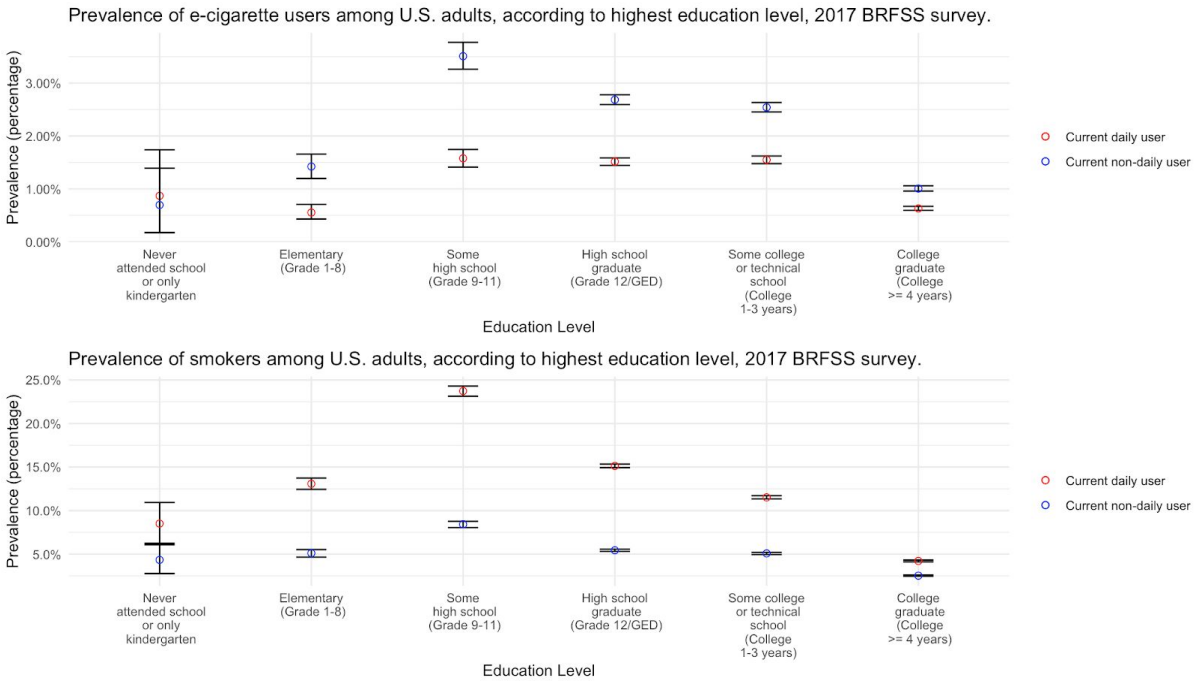
Patterns across Income



The prevalence of current e-cigarette users and smokers decreases with increase in income.

This is roughly supported by the LR table, with the notable exception that being in the $10,000-$15,000 income bracket increases your odds ratio of being a current e-cigarette user than a never e-cigarette user by 0.24 with respect to being from less than $10,000 income bracket.

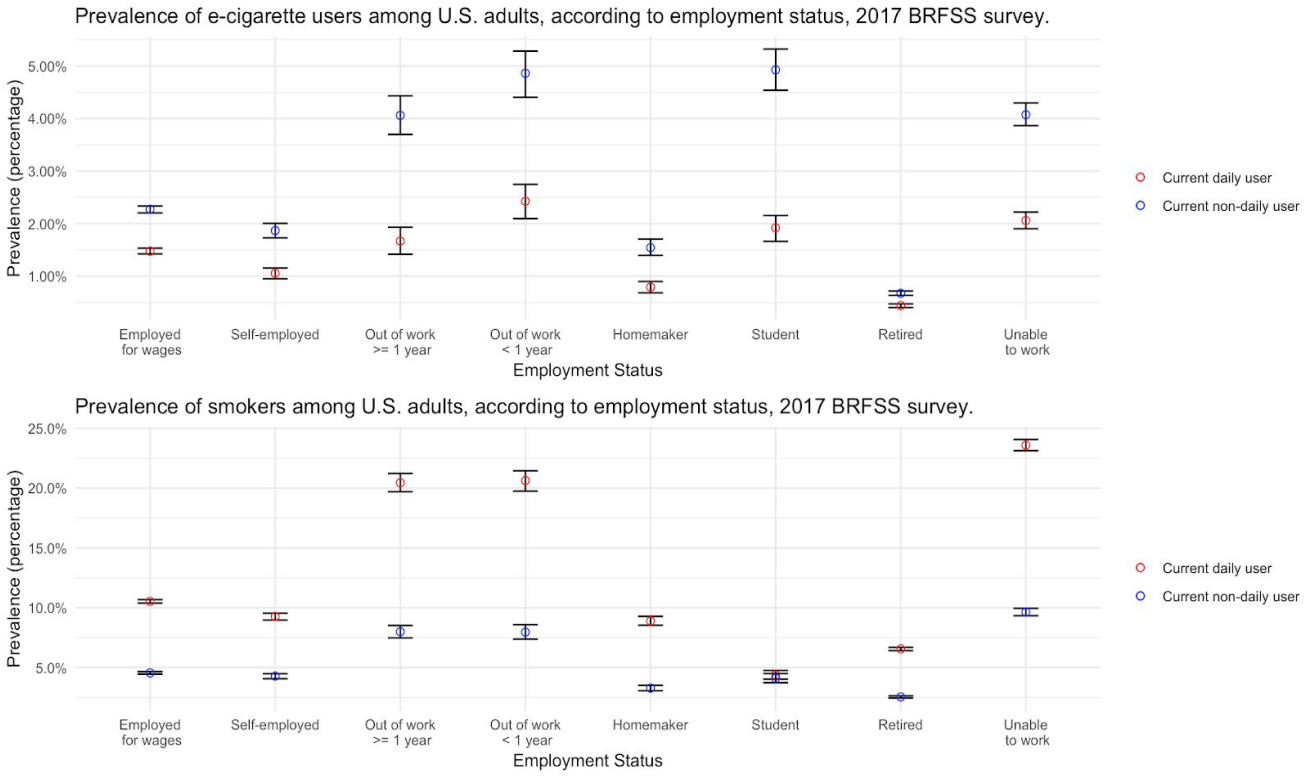
Patterns across Education



There are lots of patterns here. One noteworthy trend is that users whose highest education level is some high school (Grades 9 to 11), have the highest rate of current e-cigarette users and smokers.

From the LR table, we can confirm a corresponding trend, e.g., being in this category increases your odds ratio of being a current e-cigarette user than a never user by 1.16 as compared to if your highest education level was elementary, and by 0.24 as compared to if your highest education level was high school graduate or some college education.

Patterns across Employment Status



The prevalence of current e-cigarette users and smokers is highest among people who are out of work (both categories) and who are unable to work. This is supported by the LR table, where these coefficients have the highest values (i.e., more odds of being current users).

Logistic Regression Analysis

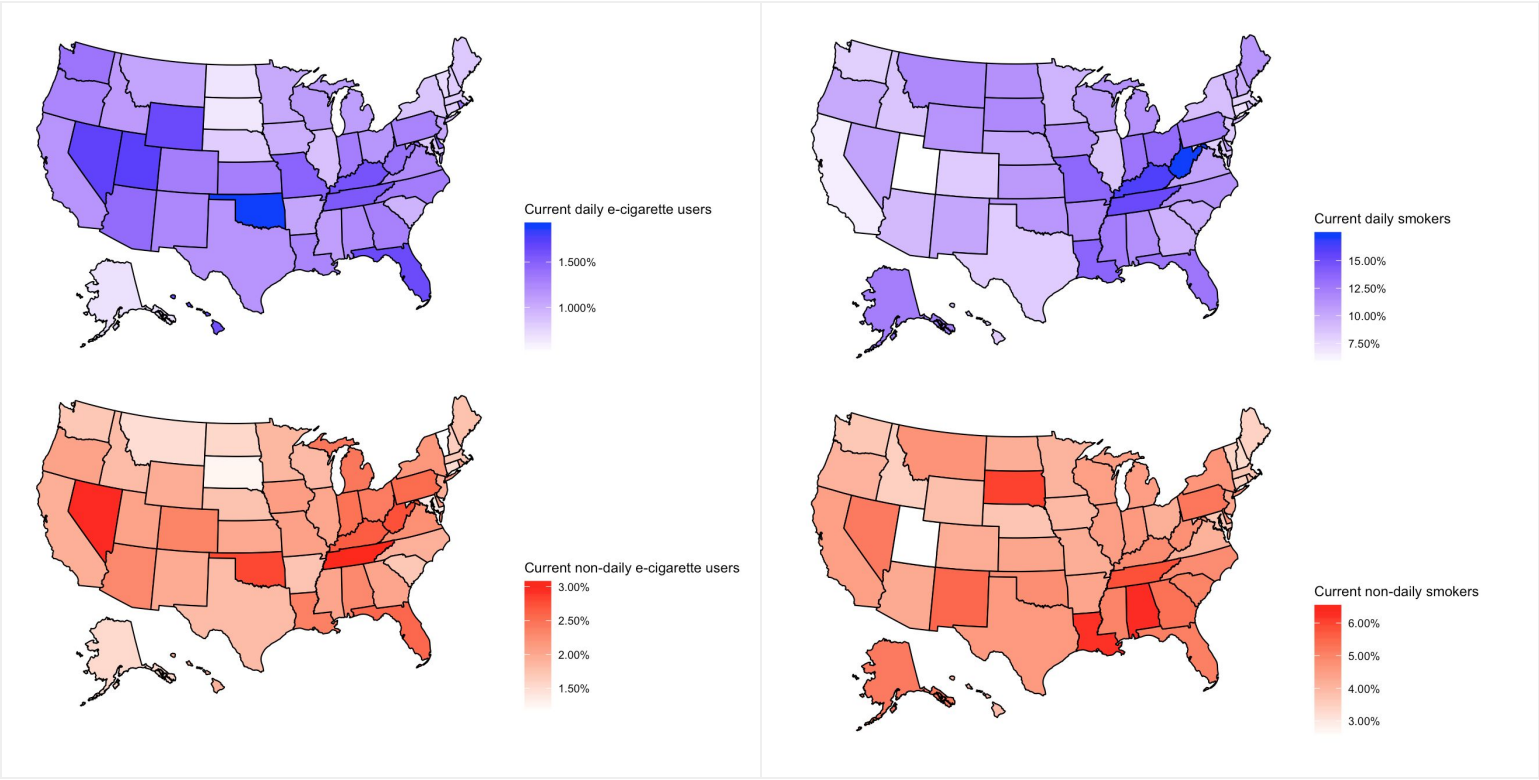
There are 4 logistic regression that have been used for the analysis. The regression coefficients are written in the form exp(β) with 95% CI for easy analysis.

	Current Smokers Vs Never Smokers	**Current Smokers Vs Former Smokers**	**Current E-Cigarette Users Vs Never E-Cigarette Users**	**Current E-Cigarette Users Vs Former E-Cigarette Users**
Age				
18-24	1	1	1	1
25-29	2.17 (2.04, 2.31)	0.75 (0.69, 0.82)	0.70 (0.65, 0.76)	0.91 (0.83, 0.99)
30-34	2.69 (2.53, 2.86)	0.76 (0.70, 0.83)	0.54 (0.50, 0.59)	0.76 (0.70, 0.83)
35-39	3.08 (2.90, 3.28)	0.79 (0.72, 0.86)	0.46 (0.42, 0.50)	0.70 (0.65, 0.76)
40-44	2.65 (2.49, 2.82)	0.72 (0.66, 0.80)	0.33 (0.30, 0.36)	0.64 (0.59, 0.70)

45-49	2.39 (2.25, 2.54)	0.74 (0.68, 0.82)	0.27 (0.25, 0.30)	0.63 (0.58, 0.68)
50-54	2.39 (2.25, 2.53)	0.77 (0.70, 0.84)	0.24 (0.22, 0.26)	0.57 (0.53, 0.62)
55-59	2.32 (2.19, 2.46)	0.67 (0.61, 0.74)	0.18 (0.16, 0.20)	0.45 (0.41, 0.48)
60-64	1.86 (1.76, 1.98)	0.65 (0.59, 0.71)	0.13 (0.11, 0.14)	0.32 (0.29, 0.34)
65-69	1.52 (1.43, 1.63)	0.61 (0.54, 0.69)	0.09 (0.08, 0.10)	0.24 (0.22, 0.26)
70-74	1.20 (1.12, 1.29)	0.59 (0.51, 0.69)	0.06 (0.05, 0.06)	0.16 (0.15, 0.18)
75-79	0.75 (0.69, 0.81)	0.50 (0.41, 0.62)	0.03 (0.02, 0.03)	0.11 (0.10, 0.12)
Greater than or equal to 80	0.28 (0.25, 0.30)	0.36 (0.26, 0.48)	0.01 (0.01, 0.01)	0.05 (0.05, 0.06)
Race				
White	1	1	1	1
Black	0.65 (0.62, 0.67)	0.76 (0.70, 0.83)	0.46 (0.42, 0.49)	1.34 (1.28, 1.40)
American Indian	1.32 (1.25, 1.40)	0.91 (0.81, 1.02)	0.85 (0.76, 0.95)	1.35 (1.27, 1.44)
Asian	0.52 (0.48, 0.56)	1.06 (0.92, 1.22)	0.69 (0.61, 0.78)	1.00 (0.91, 1.10)
Hawaiian/Pacific Islander	1.05 (0.94, 1.16)	0.99 (0.82, 1.19)	0.94 (0.79, 1.13)	1.12 (0.99, 1.27)
Other	0.46 (0.43, 0.49)	0.89 (0.76, 1.03)	0.47 (0.41, 0.54)	0.84 (0.77, 0.91)
No Preferred	1.31 (1.11, 1.55)	0.87 (0.65, 1.16)	1.25 (0.95, 1.66)	1.16 (0.96, 1.40)
Sex				
Male	1	1	1	1
Female	0.69 (0.68, 0.71)	0.94 (0.90, 0.98)	0.76 (0.73, 0.79)	1.08 (1.06, 1.11)
Income				
Less than $10,000	1	1	1	1
$10,000-$15,000	1.11 (1.06, 1.18)	1.03 (0.93, 1.15)	1.24 (1.12, 1.37)	0.94 (0.88, 1.00)
$15,000-$20,000	0.98 (0.93, 1.03)	1.04 (0.94, 1.15)	1.12 (1.02, 1.24)	0.84 (0.79, 0.89)
$20,000-$25,000	0.92 (0.87, 0.96)	1.11 (1.00, 1.23)	1.17 (1.07, 1.29)	0.75 (0.71, 0.80)
$25,000-$35,000	0.78 (0.74, 0.82)	1.03 (0.93, 1.14)	1.04 (0.95, 1.15)	0.62 (0.59, 0.66)
$35,000-$50,000	0.66 (0.63, 0.69)	1.02 (0.93, 1.13)	0.96 (0.88, 1.06)	0.53 (0.50, 0.56)
$50,000-$75,000	0.52 (0.50, 0.55)	1.03 (0.93, 1.14)	0.88 (0.80, 0.96)	0.40 (0.38, 0.43)
Greater than or equal to $75,000	0.32 (0.31, 0.34)	1.08 (0.98, 1.19)	0.70 (0.64, 0.77)	0.27 (0.26, 0.29)
Education				
Never attended school or only	1		1	1

kindergarten				
Elementary (Grade 1-8)	1.62 (1.15, 2.28)	0.80 (0.34, 1.91)	0.94 (0.43, 2.04)	1.07 (0.70, 1.61)
Some high school (Grade 9-11)	3.74 (2.68, 5.24)	0.82 (0.35, 1.92)	2.10 (0.98, 4.52)	1.45 (0.96, 2.18)
High School Graduate (Grade 12/GED)	2.29 (1.64, 3.19)	0.88 (0.38, 2.07)	1.86 (0.87, 3.97)	1.13 (0.75, 1.70)
Some college or technical school (College 1-3 years)	1.89 (1.35, 2.64)	0.86 (0.37, 2.02)	1.86 (0.87, 3.97)	0.95 (0.63, 1.43)
College Graduate (College >= 4 years)	0.71 (0.51, 0.99)	0.68 (0.29, 1.60)	0.77 (0.36, 1.66)	0.62 (0.41, 0.93)
Employment Status				
Employed for wages	1	1	1	1
Self-employed	0.84 (0.81, 0.87)	1.05 (0.97, 1.13)	0.96 (0.89, 1.03)	0.86 (0.83, 0.90)
Out of work >= 1 year	1.49 (1.40, 1.59)	1.29 (1.15, 1.45)	1.69 (1.52, 1.88)	1.16 (1.08, 1.24)
Out of work < 1 year	1.49 (1.40, 1.59)	1.18 (1.06, 1.31)	1.61 (1.45, 1.78)	1.26 (1.17, 1.36)
Homemaker	0.70 (0.66, 0.74)	1.14 (1.02, 1.28)	0.76 (0.69, 0.84)	0.88 (0.83, 0.94)
Student	0.49 (0.45, 0.53)	0.97 (0.87, 1.07)	0.75 (0.68, 0.83)	0.72 (0.65, 0.80)
Retired	1.12 (1.08, 1.16)	1.18 (1.07, 1.30)	1.26 (1.16, 1.38)	0.89 (0.85, 0.93)
Unable to work	1.84 (1.77, 1.91)	1.34 (1.24, 1.44)	2.24 (2.09, 2.40)	1.11 (1.07, 1.16)

Patterns across States



I have plot the current daily and non-daily use of e-cigarette users and smokers across the US states. It is interesting to see that in certain states such as Oklahoma, Utah and Nevada, e-cigarette use is a lot more prevalent than smoking (relative to other states). In states like South Dakota and Alaska, smoking is relatively a lot more prevalent than e-cigarettes (relative to other states).

Limitations

As mentioned earlier, we have considered logistic regression analysis for categories 1 and 2 together, with respect to categories 3 and 4 individually, and the visualization considers categories 1 and 2 alone independently. Then, we have tried to analyze them together.

For the visualization, what we see for prevalence are the percentage values of users for different demographic variables. These are no causality tests performed here, e.g., the prevalence of current e-cigarette users and smokers being high among people who are out of work does not mean that if you are out of work, you will start/increase your smoking/e-cigarette use. It might very well be that there are unobserved variables and/or reversed causality that we have not considered that cause this. As an example, being out of work might lead to more free time, which increases tobacco use.

Similarly, for the logistic regression, we have made the assumption that the features used are independent of each other. This is strictly speaking, not true. Respondents who have a lower highest education level may have a lower income range. As a result of this, we might attribute the change in odds ratio to a lower income range, whereas it may have been caused by a lower highest education level.

Finally, there might be errors in the way the data is collected. BRFSS uses telephone surveys and the reliability and validity of the BRFSS is studied in numerous studies independently. However, BRFSS is usually regarded as a reliable and valid source of information on health related issues.

Conclusions

1. **Age:** The prevalence of current daily and non-daily e-cigarette users decreased with increasing age, unlike smokers.
2. **Race:** With respect to White race category, there are fewer percentage of Black/African Americans and Native Americans who are current e-cigarette users, but they are less prone to quit e-cigarette smoking.
3. **Sex:** The prevalence of current e-cigarette users and smokers is more in males than in females. However, females are less likely to quit e-cigarette smoking than males, albeit by a slight margin.
4. **Income:** The prevalence of current e-cigarette users and smokers decreases with increase in income.
5. **Education:** Users whose highest education level is some high school (Grades 9 to 11), have the highest rate of current e-cigarette users and smokers.
6. **Employment Status:** The prevalence of current e-cigarette users and smokers is highest among people who are out of work (both categories) and who are unable to work.

References

1. Social and Structural Determinants of Smoking Status and Quit Attempts Among Adults Living in 12 US States, 2015
2. Prevalence and Distribution of E-Cigarette Use Among U.S. Adults: Behavioral Risk Factor Surveillance System, 2016
3. State-Specific Patterns of Cigarette Smoking, Smokeless Tobacco Use, and E-Cigarette Use Among Adults — United States, 2016
4. Calculated Variables in the 2017 Behavioral Risk Factor Surveillance System Data File
5. LLCP 2017 Codebook Report, Behavioral Risk Factor Surveillance System