

# Smoking, Drinking and Drug use habits in 2015-16

**Vaibhav Mavi**

vm2241@nyu.edu

**So Hyun Chi**

sc9120@nyu.edu

## Abstract

Among the leading causes of avoidable fatalities in the US are drugs, alcohol, and tobacco. Additionally, studies show that these drugs are frequently consumed together: Smokers are significantly more likely to drink and consume drugs, and vice versa. The aim of this study is to find whether there is a relation between alcohol, cigarette and drug consumption. Through the experiments, it is found that there is a tendency to start consuming these substances at a very young age ( $< 20$ ). We also find that if people started consuming one or two of the three, they are more likely to start consuming the others.

## 1 Introduction

Smoking tobacco/cigarettes, the chronic use of alcohol, and drug usages are few of the major risk factors affecting health in the United States (Lushniak, 2014; Jha P, 2015; of Health and Services, 2004). Smoking has a direct impact on chronic drinking. Also, many interactions between tobacco smoke and drugs have been identified (Karila, 2013; eff, 2020) and the effects of use of these substances together with drinking (Koppiseti and Loka, 2011; Dunham, 2007; smo, 2016; Verplaetse TL, 2017). Various studies analyse the effects of tobacco, cigar, e-cigarettes, smokeless tobacco on health (Institute, 1998; Organization, 2007; Verplaetse TL, 2017). Alcohol and smoking roughly cause 3 million deaths in the United States annually, which includes 0.7 million deaths of women (of Health and Services, 2001) and 2.3 million deaths of men (for Disease Control and Prevention, 2013). More than 932,000 people have died since 1999 from a drug overdose. In 2020, 91,799 drug overdose deaths occurred in the United States (Lushniak, 2014).

### 1.1 Objectives

Smoking, drinking, and drug use are key public health concerns (Lushniak, 2014; Jha P, 2015; of Health and Services, 2004). According to the National Health and

Nutrition Examination Survey, there is a positive association between current smoking status and alcoholic drinking per day. Through this study, we want to answer questions like: If a person starts consuming one or more of such substances, how likely are they to consume the other substances. Additionally, if a person consumes multiple such substances, which ones did they start consuming first? Is there a correlation between the drinking, smoking or drug use behavior (such as frequency of use and current status)?

## 2 Methodology

### 2.1 Data

We intend to use the data from the National Health and Nutrition Examination Survey (NHANES), 2015-16<sup>1</sup>. NHANES is a program of studies designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations. Findings from this survey are used to determine the prevalence of major diseases and risk factors for diseases. Information is used to assess nutritional status and its association with health promotion and disease prevention. NHANES findings are also the basis for national standards for such measurements as height, weight, and blood pressure. From the survey, we are going to use three questionnaires:

- Alcohol use (ALQ\_I)<sup>2</sup>: A survey of 5735 males and females aged 18-150 years, 10 columns and 7 questions regarding alcohol use over past 12 months.
- Smoking: Cigarette use (SMQ\_I)<sup>3</sup>: A survey of 7001 males and females aged 18-150 years, 42 columns and 27 questions regarding smoking habits.

<sup>1</sup>[https://www.cdc.gov/nchs/nhanes/about\\_nhanes.htm](https://www.cdc.gov/nchs/nhanes/about_nhanes.htm)

<sup>2</sup>[https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/ALQ\\_I.htm](https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/ALQ_I.htm)

<sup>3</sup>[https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/SMQ\\_I.htm](https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/SMQ_I.htm)

- Drug use (DUQ\_I)<sup>4</sup>: A survey of 4843 males and females age 18-69 years, 44 columns and 37 questions regarding drug use and history.

Taking the intersection of the above datasets (inner-join on sequence number), we get the final dataset with 4843 rows and 92 columns.

## 2.2 Experiments

In the above mentioned datasets, we focus on the columns corresponding to the following questions:

- Have you ever smoked, ever consumed marijuana / methamphetamine / hashish / cocaine / heroine or ever had more than 4 drinks in a day?
- What was your age when you first consumed these substances and when you started consuming these regularly?
- How many times in last few days/months have you consumed these substances?

### 2.2.1 Visualizations

Following visualizations are performed to better understand the goal:

1. Histograms of age when people started consuming different types of drugs, cigarettes and alcohol, shown in Figure 1. This helps us to understand th...  
TODO
2. A bar graph comparing the numbers from the first graphs (age when they started consuming drugs and cigarettes). The data is split into age groups of [0, 20]; [21, 30]; [31, 40]; ... This helps us understand which age groups are more susceptible to these substances. For instance, whether the young people are more likely to smoke than to consume cocaine or vice versa. Since the number of people who smoke cigarettes are much higher than those who consume drugs, it is also informative to show the percentage of people in different age groups consuming that particular substance.
3. A Venn Diagram of whether people have ever consumed one or more of drugs, cigarettes and alcohol. This helps us understand the overlap between these habits in people.
4. A histogram of number of alcoholic drinks per day.

<sup>4</sup>[https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/DUQ\\_I.htm](https://wwwn.cdc.gov/Nchs/Nhanes/2015-2016/DUQ_I.htm)

5. A scatter plot of the number of cigarettes smoked per day in the past month and number of alcoholic drinks consumed per day in the past 12 months.

6. A scatter plot of age when people first smoke a cigarette, with the number of alcoholic drinks consumed to observe the correlation between smoking and drinking habits.

### 2.2.2 Predictive analysis

We plan to use the following predictive analytical methods in our project:

1. **Linear Regression** to understand the relation between the age when people started smoking or consuming drugs and the age at which they started consuming drugs. The hypothesis is that the people who started consuming drugs at a younger age might have bad smoking habits as well. Also, among the people who consume both cigarettes and drugs, which one did they start consuming first.
2. **Linear Regression** to understand the relationship between the number of joints or pipes of marijuana smoked and the average number of drinks consumed per day in the past 12 months. This will help us understand the association between marijuana and drinking habits of people.
3. **Logistic Regression** to understand how likely are the people who have ever consumed alcohol and cigarettes, to have ever consumed drugs. This will show whether people who consume and alcohol and cigarettes are more susceptible to drugs and if so, which one of the two is more likely to cause people to consume drugs.

## 3 Results

We include the results of each experiment in the figures below:

## 4 Discussion

### 4.1 Visualizations

From figure 1, we observe that people aged around 15-20 are more susceptible to consuming these substances which is surprising and not desirable. For some drugs like cocaine, the age is higher (around 20-25) but for drugs like marijuana, almost all of the data lies to the left of 20. This implies that most people who use marijuana started consuming it before the age of 20.

From figure 2, we observe that more number of people have used marijuana than have smoked cigarettes

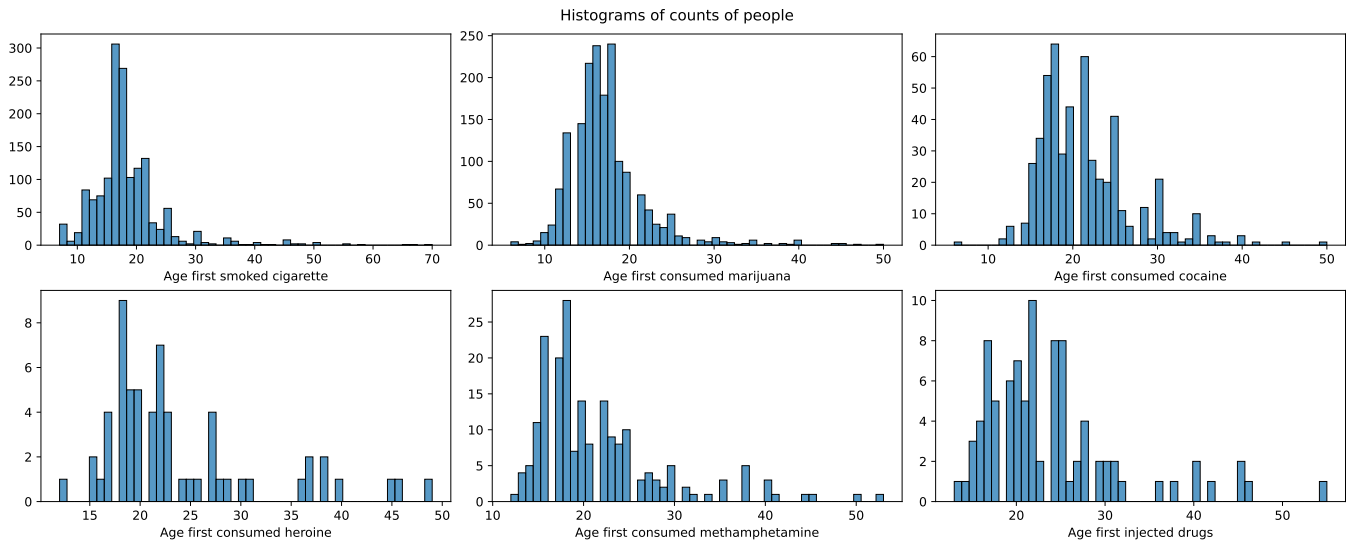


Figure 1: Histograms of age when people started consuming different substances.

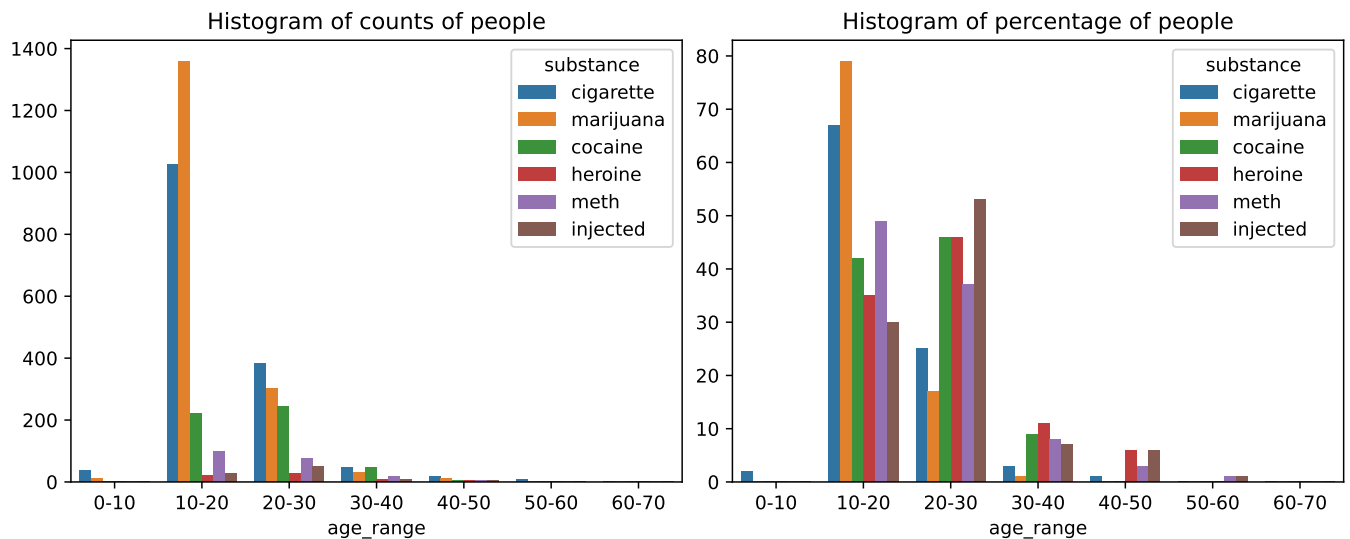


Figure 2: Comparison of data shown in Figure 1 for different age-groups. Left shows the number of people who started consuming these substances during that age range. Right shows the percentage of people using these substances who started consuming at that age.

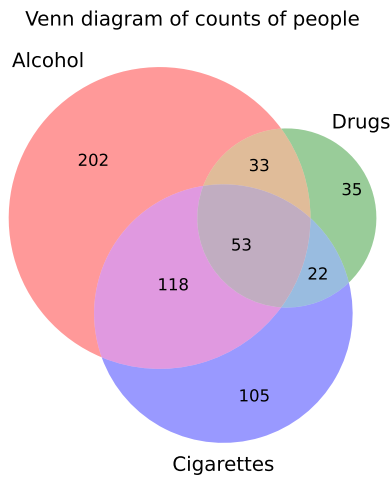


Figure 3: Venn diagram showing the number of people who consume one, two or all three of alcohol, drugs and cigarettes.

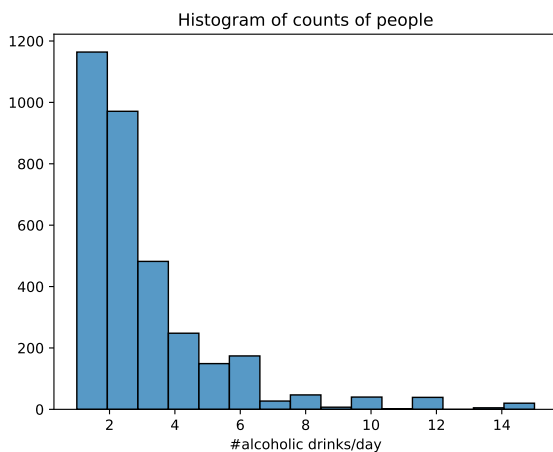


Figure 4: Histogram of number of alcoholic drinks consumed per day.

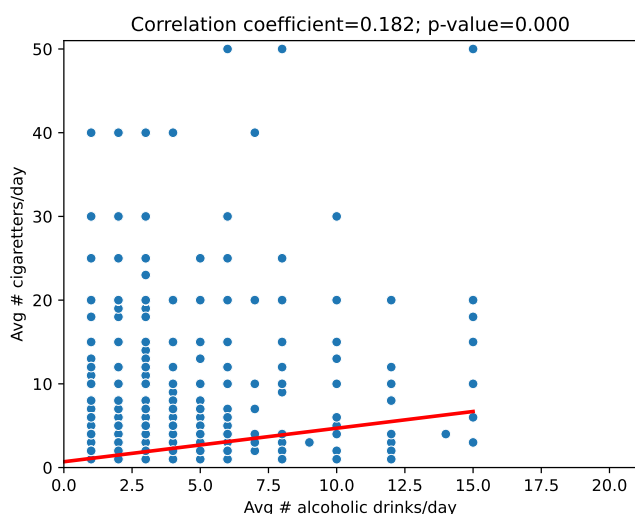


Figure 5: Scatter plot of average number of cigarettes smoked Vs average number of alcoholic drinks consumed per day.

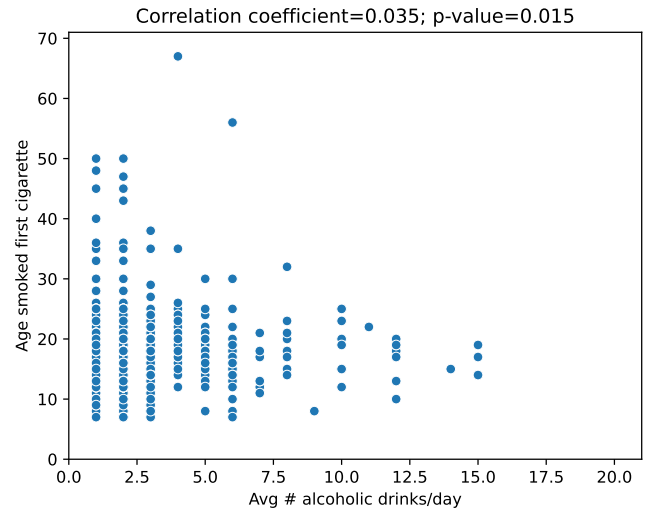


Figure 6: Scatter plot of age when first smoked a cigarette Vs average number of alcoholic drinks consumed per day.

before the age of 20. This is once again surprising and undesirable. Also a much higher number of people start consuming these substances at a younger age.

From figure 3, we can make multiple significant conclusions: a) Among the 171 people who consume both alcohol and cigarettes, 53( 31%) also end up consuming drugs. This is a significant indicator that people with smoking and drinking habits are likely to consume drugs as well. b) Conversely, among the 143 people who consume drugs, 86( 60%) of them consume alcohol, 75( 52%) consume cigarettes and 53( 37%) do both. This indicates a strong correlation between these habits.

From figure 4, we observe that most people consume 1 – 3 alcoholic drinks everyday. A significant number of people consume 3 alcoholic drinks everyday which is an unhealthy habit.

From figure 5 and looking at the Pearson coefficient and the p-value, we note that there is a moderately weak correlation between the number of cigarettes smoked everyday and number of alcoholic drinks consumed. This indicates that people who smoke more frequently are slightly more likely to drink more often.

Looking at figure 6 and the Pearson coefficient and the p-value, we can conclude that there is a negligible correlation between when people started smoking and how many alcoholic drinks they consume. This indicates that early-age smoking does not imply either good or bad drinking habits.

## 4.2 Predictive methods

From the results of linear regression shown in figure 7, we see that the p-value is 0 and the coefficient for the

AIC = 2711.9242755463547

#### OLS Regression Results

Dep. Variable:	age_cig	R-squared:	0.070
Model:	OLS	Adj. R-squared:	0.068
Method:	Least Squares	F-statistic:	36.61
Date:	Wed, 30 Nov 2022	Prob (F-statistic):	2.88e-09
Time:	02:29:09	Log-Likelihood:	-1354.0
No. Observations:	489	AIC:	2712.
Df Residuals:	487	BIC:	2720.
Df Model:	1		
Covariance Type:	nonrobust		
	coef	std err	t P> t  [0.025 0.975]
const	13.5783	0.677	20.057 0.000 12.248 14.909
age_marijuana	0.2180	0.036	6.051 0.000 0.147 0.289
Omnibus:	160.944	Durbin-Watson:	2.028
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1283.760
Skew:	1.198	Prob(JB):	1.72e-279
Kurtosis:	10.567	Cond. No.	73.0

Figure 7: Linear regression stats of age when first consumed marijuana Vs age when first smoked a cigarette.

independent variable is 0.2. This means that people are more likely to start smoking cigarettes later than they start smoking marijuana. This result is also highlighted in the scatter plot in figure 8.

The results of linear regression shown in figure 9 have a very low p-value but the AIC value is extremely large which means that the model is not fit well enough to draw any significant conclusions. The reason might be that the dependent variable in this case is categorical and linear regression does not work well for such cases. Consequently, we use logistic regression for further analysis.

In the results shown in figure 10, we note that the p-values for both the dependent variables are lower than 0.05 which implies the result is statistically significant and the AIC score also implies that the model is reasonably fit. Thus, we look at the Odds Ratio and observe that the people who drink heavily (4-5 drinks/day), are 1.55 times more likely to consume drugs whereas people who have smoked at least 1 cigarette in life are 3 times less likely to consume drugs. This is in contrast to what we expect and what the previous analyses suggest. Possible reason for this might be that we are only considering whether people consumed these substances at least once and not if they do so regularly.

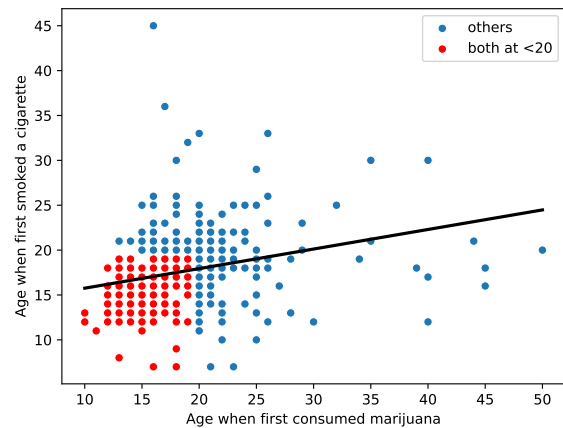


Figure 8: Scatter plot of age when first consumed marijuana Vs age when first smoked a cigarette. Points in red indicate those who started both the substances at an age of < 20

#### OLS Regression Results

Dep. Variable:	#joints/day	R-squared:	0.033
Model:	OLS	Adj. R-squared:	0.033
Method:	Least Squares	F-statistic:	102.1
Date:	Wed, 30 Nov 2022	Prob (F-statistic):	1.24e-23
Time:	03:11:52	Log-Likelihood:	-5366.4
No. Observations:	2968	AIC:	1.074e+04
Df Residuals:	2966	BIC:	1.075e+04
Df Model:	1		
Covariance Type:	nonrobust		
	coef	std err	t P> t  [0.025 0.975]
const	0.3174	0.042	7.550 0.000 0.235 0.400
#alcoholic drinks/day	0.1133	0.011	10.107 0.000 0.091 0.135
Omnibus:	1781.427	Durbin-Watson:	2.028
Prob(Omnibus):	0.000	Jarque-Bera (JB):	13650.850
Skew:	2.872	Prob(JB):	0.00
Kurtosis:	11.798	Cond. No.	6.07

Figure 9: Linear regression stats of number of joints smoked Vs alcoholic drinks per day.

AIC = 2391.7676389662142

	OddsRatio	P> z	2.5%	97.5%
Intercept	1.231183	5.722397e-01	0.598221	2.533864
ever_smoked[T.True]	0.392986	9.569970e-21	0.323047	0.478068
ever_had_5_drinks	1.559316	1.670920e-02	1.083718	2.243635

Figure 10: Logistic regression stats of 'Ever consumed drugs Ever smoked a cigarette + Ever had 5 drinks/day'.



## 5 Conclusion

Based on the above analysis, we can make the following conclusions about the smoking, drugs and drinking habits of people:

- People who consume both alcohol and cigarettes are very likely to consume some kind of drugs. Also, people who consume drugs are very likely to consume cigarettes and alcohol.
- People are most likely to start consuming marijuana and cigarettes when they are 10-20 years of age and most likely to consume or inject illegal substances during the ages of 20-30.
- People in their teens are more likely to consume marijuana than to smoke cigarettes, while the trend is the other way around for the other age groups. Also, people who consume both are more likely to first consume marijuana.
- Among people who consume alcohol, close to half have at least 3 alcoholic drinks everyday which indicates that drinking is likely to become an unhealthy habit.
- People who consume alcohol in large amounts (4-5 drinks/day) are more likely to consume marijuana than those who don't.

These results clearly indicate that smoking, drinking and drug use are very correlated in a lot of ways, and consuming one or more of these makes people more likely to consume the other substances. Consequently, since all of these substance use can be detrimental to a person's health, it is best advised to avoid ever using any of these substances.

## References

2016. [Smoking cessation in people with alcohol and other drug problems](#). *Australian Journal for General Practitioners*, 45(8):569–573.
2020. [How alcohol increases the effects of cigarettes](#).
- Centers for Disease Control and Prevention. 2013. [Quickstats: Number of deaths from 10 leading causes—national vital statistics system, united states](#). *Morbidity and Mortality Weekly Report*.
- Will Dunham. 2007. [Smoking linked to teen alcohol, drug use: U.s. study](#).
- U.S. Department of Health and Human Services. 2001. [Women and smoking: A report of the surgeon general](#). Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Coordinating Center for Health Promotion, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.
- U.S. Department of Health and Human Services. 2004. [The health consequences of smoking: A report of the surgeon general](#). Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.
- National Cancer Institute. 1998. [Cigars: Health effects and trends](#). *Smoking and Tobacco Control Monograph No. 9*. Bethesda (MD): U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute.
- Landsman V Rostrom B Thun M Anderson RN McAfee T Peto R Jha P, Ramasundarahettige C. 2015. [21st century hazards of smoking and benefits of cessation in the united states](#). *New England Journal of Medicine*, 2013;368(4):341–50 [accessed 2015 Aug 17].
- Petit A. Zarmidini R. Coscas S. Lowenstein W. Reynaud M. Karila, L. 2013. [Consommation de tabac et trouble lié à l'usage de substances illicites: que devrions-nous faire ? \[tobacco use and illicit substance use disorders: what should we have to do?\]](#). *Presse medicale (Paris, France : 1983)*.
- V.S. Koppiseti and Nikhil Loka. 2011. Influence of alcohol and smoking on drug action: A step for better utilization of drugs. *Journal of Chemical and Pharmaceutical Research*, 3:242–248.
- Jonathan M.;Pechacek Terry F.;Norman Leslie A.;Taylor Peter A. Lushniak, Boris D.;Samet. 2014. [The health consequences of smoking—50 years of progress : a report of the surgeon general](#).
- World Health Organization. 2007. [Smokeless tobacco and some tobacco-specific n-nitrosamines](#). *International Agency for Research on Cancer Monographs on the Evaluation of Carcinogenic Risks to Humans Vol. 89*. Lyon, (France): World Health Organization, International Agency for Research on Cancer.
- McKee SA Verplaetse TL. 2017. [An overview of alcohol and tobacco/nicotine interactions in the human laboratory](#). *Am J Drug Alcohol Abuse*.