Project: The Effect of Religiosity on Smoking Patterns in Adolescents

Armaan Valvi

Introduction

While it is an established truism that smoking is injurious to health, in the United States, cigarette smoking continues to be the greatest cause of preventable death (Danaei et al., 2009; Mokdad, Marks, Stroup, & Gerberding 2004), with around 443,000 premature deaths every year. As of 2008, this accounts for nearly 20% of all deaths across the country, according to the Centers for Disease Control and Prevention. Despite its universally-known danger, deaths attributable to tobacco are predicted to reach a maximum of 8.3 million around the world and account for 10% of all global deaths. Curiously, in 2002, the Vatican released a law, signed by Pope John Paul II, that banned smoking in all closed places of work, all places accessible to the public and within all properties of the Holy See, thereby introducing a religious element to smoking habits. This event drives the motivation for this study; in this regard, this report aims to investigate whether how religious an adolescent is affects their smoking status (i.e., whether or not they are regular smokers). Thus, the research question of interest is: does the strength of the religious beliefs of the adolescent (henceforth called 'religiosity') decrease the likelihood of them being regular smokers? The treatment variable is religiosity and the outcome variable is smoking status, both coded as binary variables. This study involves running different logistic regression models on matched data, and controls for a number of factors, such as availability of cigarettes at home, whether or not the respondent's school conducted smoking-related instruction programs and gender.

Theory

Studies reveal, among adolescents, a general inverse association between religiosity and the odds of taking up smoking. A study conducted by Koenig (2011) finds that spirituality and religiousness were highly protective against smoking. In addition, a study conducted among 7th to 10th grade students by Wills, Yaeger and

Sandy (2003) reveals that higher ratings of religious importance were inversely correlated with tobacco use as well as mitigated the impact of stress on tobacco use. Similarly, Wallace et al. (2003) observe that among high school students, the perceived importance of religion in one's life was highly and inversely associated with tobacco consumption and that higher importance given to religion was positively correlated with a higher likelihood of abstaining from tobacco use altogether. However, while a majority of the literature on this topic points to an inverse relationship between religiosity and smoking, a notable exception is found in Nollen et al. (2005), who find that religious beliefs, practices and values were not associated with a reduction in smoking among young African American smokers. Furthermore, studies such as Garrusi and Nakhaee (2014), and Brown at al. (2014) highlight the limitations of using cross-sectional data in such studies as they do not account for a temporal change in smoking habits or their outlook on smoking over time due to religion. Thus, while there appears to be a general consensus on the negative relationship between religion and the likelihood of taking up smoking regularly among adolescents, some studies question the validity of the findings.

Data

The data used in this paper is collected from the first questionnaire of the National Longitudinal Study of Adolescent to Adult Health (1994-2008) (NLSAAH) conducted by the University of North Carolina-Chapel Hill. The treatment variable used in this study is religiosity, recoded as a binary variable (0 if atheist/not important, 1 if important). The outcome variable of interest is whether or not the respondent has smoked at least one cigarette every day in the last 30 days, recoded as a binary variable for smoking status (0 if no, 1 if yes). This study includes seven other variables recoded as binary variables. These are: i) whether or not the respondent received instruction on the dangers of smoking in school (0 if no, 1 if yes); ii) gender (0 if male, 1 if female); iii) whether or not the respondent has had trouble sleeping at least once a week in the past year (0 if no, 1 if yes) as an indicator of stress levels; iv) whether or not the respondent received instruction on stress management in school (0 if no, 1 if yes); v) whether or not the respondent's resident mother has ever smoked before (0 if no, 1 if yes); vi) whether or not the respondent's resident father has ever smoked before (0 if no, 1 if yes); and vii) whether cigarettes are readily available in the respondent's house (0 if no, 1 if yes).

The NLSAAH contained missing values for most variables because some respondents refused to answer certain questions or did not have an answer. Missing results were relatively very few in four variables (smoking-related instruction in school, having trouble sleeping, stress-management instruction in school and cigarette availability at home), and many in five (smoker status, religiosity, resident mother's smoking history, resident

father's smoking history and gender). Three variables alone (gender, mother's smoking history and father's smoking history) accounted for 4087 of the missing responses. Since the survey included just 'male' or 'female' as the options to choose from, the large number of missing values for the gender variable could perhaps be because many respondents did not identify with either category or did not feel comfortable revealing their gender. Similarly, many respondents did not have a resident mother or father, as a result of which many responses for the influence of resident mother/father's smoking history could not be included. Thus, this paper resorts to simple listwise deletion of data as imputing missing data for such variables would not be appropriate.

On the other hand, all missing values were included in the data as 0s for the treatment and outcome variables. In the case of 'religiosity', those who did not have a religion were included as those believed that religion is not important because these respondents were important to the treatment. However, since reverse-classifying missing NA values from the original data into specific responses is not possible, 11 additional missing values (where the respondent either refused to answer or did not know) were also included. Similarly, with the outcome variable, those who had never smoked a cigarette before were included as part of those who had smoked less than one cigarette every day in the last 30 days as they had to be counted as non-smokers. However, like with the religiosity variable, there were 4 additional missing values (where the respondent either refused to answer or did not know) that were converted to 0s. The limitations of this approach are addressed in the Discussion section. However, given the relatively minuscule number of unwanted missing values, including the legitimitate skips for these variables was preferred to imputing the missing data. Altogether after removing the unwanted missing data and adding the required 'legitimate skips', the sample size becomes 1083 (down from 6504). While this is a large reduction, a sample size of 1083 is still quite large. The degree of data missingness is depicted in Figure 3 (pre-matching) in the Appendix. The post-matched plot is seen in Figure 4, which shows the adjusted and unadjusted covariates of the data.

In order to minimize the effect of confounding variables and remove outliers, this paper matches the data, which involves finding pairwise data points that are similar in most respects except for the treatment variable. Thus, matching helps account for differences in the data that are not related to the treatment variable. By doing so, the aim is that only the difference in the treatment variable (religious or not) drives the differences observe in the outcome variable (regular smoker or not). However, after matching, the sample size fell from 1083 to 314, which is not sufficient to make causal claims that are generalizable. In addition, despite matching on all the observed covariates of interest, this paper does not assume that matching was done on all unobserved covariates as well. There are other unobserved factors that have not been accounted for. The table of summary statistics, Table 1, is presented below.

Table 1: Summary Statistics for the Variables (Matched)

Religiosity	# of Observations	% Who Smoked 1+ Cigarettes Daily	Std. Dev (Smoked 1+ Cigarettes Daily)	% With Smoking Education	% With Sleep Issues (1+ Times Weekly)	% With Stress Education	% Whose Mother has Smoked	% Whose Father has Smoked	% With Cigarette Access	% Who are Female
0	585	25.30	0.44	94.53	24.96	61.03	52.99	59.66	33.16	48.38
1	585	16.41	0.37	91.97	25.98	61.54	53.85	60.17	32.48	49.74

Methods

This paper adopts a multivariate logistic regression model for the data as the outcome variable is binary, thus such a model is an appropriate model for calculating the predicted probability of whether or not an adolescent is a regular smoker based on whether or not they are religious. This can be achieved from the logit function, which can be written as:

$$ln(\frac{p}{1-p}) = a_0 + a_1x_1 + a_2x_2 + ... + a_8x_8$$

$$p = \frac{exp(a_0 + a_1x_1 + a_2x_2 + ... + a_8x_8)}{1 + exp(a_0 + a_1x_1 + a_2x_2 + ... + a_8x_8)},$$

where p is the probability of being a regular smoker, and the $'a'$s are the coefficients for the nine covariates (including the constant term). The constant term, a_0, denotes the log (odds) value when all the covariate outcomes, x_1 to x_8, correspond to 0. The coefficients, a_1 to a_8, are the log-odds ratios, which tell you, on a log scale, to what extent the outcome of their corresponding covariate values affect the odds of being a regular smoker. Therefore, if a_1 is the coefficient for the treatment variable, receiving the treatment is associated with the odds of the dependent outcome of 1 changing by a multiplicative factor of a_1, holding all other variables constant.

It is important to highlight the difference between the log (odds) and the log-odds ratio. The log (odds) is simply the natural logarithm of the fraction of the number of the successes divided by the number of failures, or the odds. On the other hand, the log-odds ratio is the natural logarithm of the ratio of two odds - the odds of the dependent variable outcome being 1 given that the covariate outcome is 1, divided by the odds of the dependent variable outcome being 1 given that the covariate outcome is 0, holding all other variables constant.

Assumptions for using the logistic regression method are that the degree of treatment are same for the

treatment group and degree of outcome is binary. Similarly, that the quality of smoking-related and stress-management instruction are equivalent for schools that provide them. Since this study deals with binary data, no marginal analysis can be performed.

Four different logistic regression models were used: a bivariate model, a model with two additional controls (gender and smoking-related education), a model with four controls (including trouble with sleep and stress-management education), and a full model with all eight controls (including mother's and father's smoking history and cigarette availability at home). Pairwise ANOVA tests comparing each of the four models in order to determine the optimal model were done, and the full model was found to be the best model (see the Results section for details). As the equation for the full model had the smallest p-value, the full model was used for the analysis. The equation for the full model is presented below:

$$smoker.status = (-0.573) * religiosity + (-0.177) * female + 0.008 * smoking.education$$

$$+0.595 * trouble.sleeping + 0.266 * stress.education + 0.219 * mother.smoke$$

$$+0.249 * father.smoke + 0.570 * cigarette.access + (-1.827).$$

Finally, a regression table (Table 2) for the four models, two plots (Figures 1 and 2) were included; Figure 1 plots the relationship only between smoking status and religiosity while Figure 2 also includes the statistically significant covariates as controls (see Figure 2 in the Results section).

Results

Table 2, depicts a regression table with four different logistic regression models, where the outcome variable represents the predicted probability that an adolescent is a regular smoker. With reference to the Method section above, the first logistic model is a bivariate model, the second a 3-factor model, the third a 5-factor model and the fourth a full model. The reason for choosing the full model as the model for analysis is illustrated in Tables 4, 5 and 6 (see Appendix), which present the results for the pairwise ANOVA comparisons for all four models. Based on the p-values of the tables, while model 2 is not significantly better than model 1 (with a p-value of 0.818), the p-value corresponding to the comparison between models 1 and 3, 0.01, suggests that model 3 has far less bias than model 1. However, the model with all controls, model 4, is the best representation of the data with a p-value of around 0 (9.459094e-08, to be precise), suggesting that it has the least bias with statistically significant control variables, despite having the highest deviance.

Using the full model equation in the Methods section, we can predict the probability of a respondent being

a regular smoker based on which covariates correspond to 0 and 1. As an example, if we assume that all the covariates correspond to 1, i.e. a religious female who has received stress and smoking related instruction in school, has trouble sleeping and access to cigarettes at home, and whose parents both have a history of smoking, then the predicted probability of her being a regular smoker is:

$$ln(\frac{p}{1-p}) = exp((-0.573)*1+(-0.177)*1+0.008*1+0.595*1+0.266*1+0.219*1+0.249*1+0.570*1+(-1.827))$$

$$p = 0.338.$$

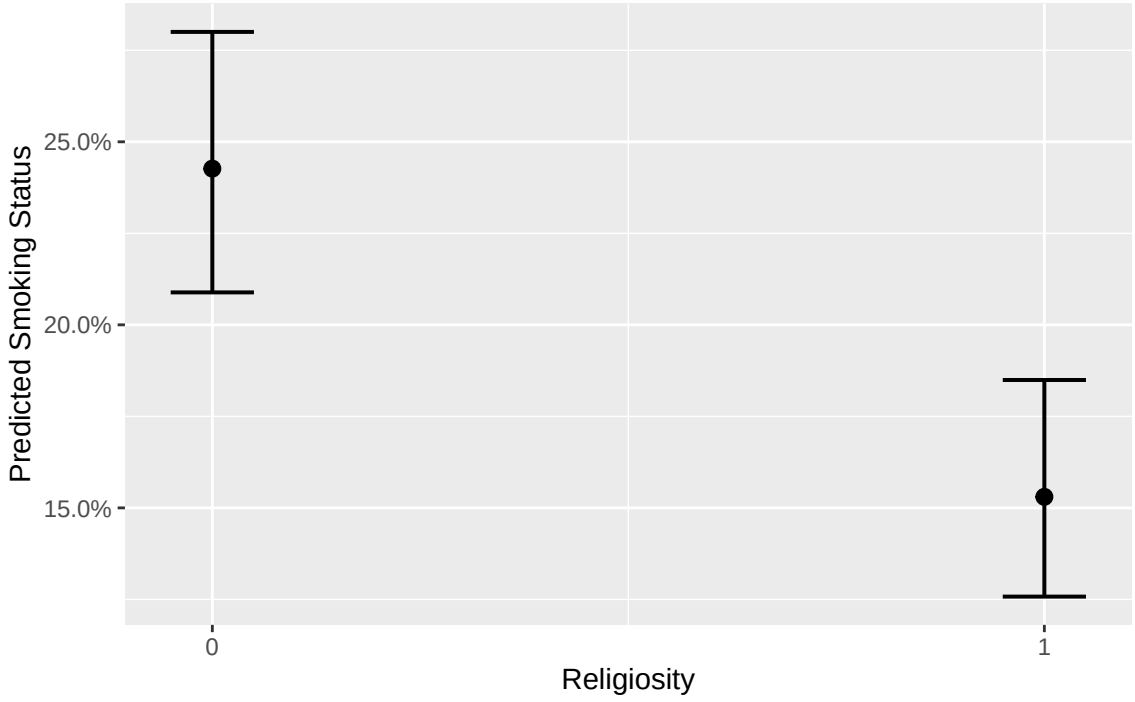
However, more than the predicted probability itself of a certain combination of outcomes, we are more interested in what effect the outcome of a covariate has on the odds of an adolescent being a regular smoker, which is given by the coefficients of the regression table. In all the models included, the coefficient of religiosity is negative and statistically significant at the 0.01 level, suggesting a strong negative association between religiosity and smoking status. Not only are these values statistically significant, they are great in magnitude, with log-odd ratio values between ranging -0.573 and -0.545. These values denote the log-odds ratio of the odds that a 'religious' adolescent is a regular smoker over the odds that an 'areligious' adolescent is a regular smoker, holding all other variables constant. The negative sign reveals that the odds than an 'areligious' adolescent is a regular smoker is greater than the odds that a 'religious' adolescent is one.

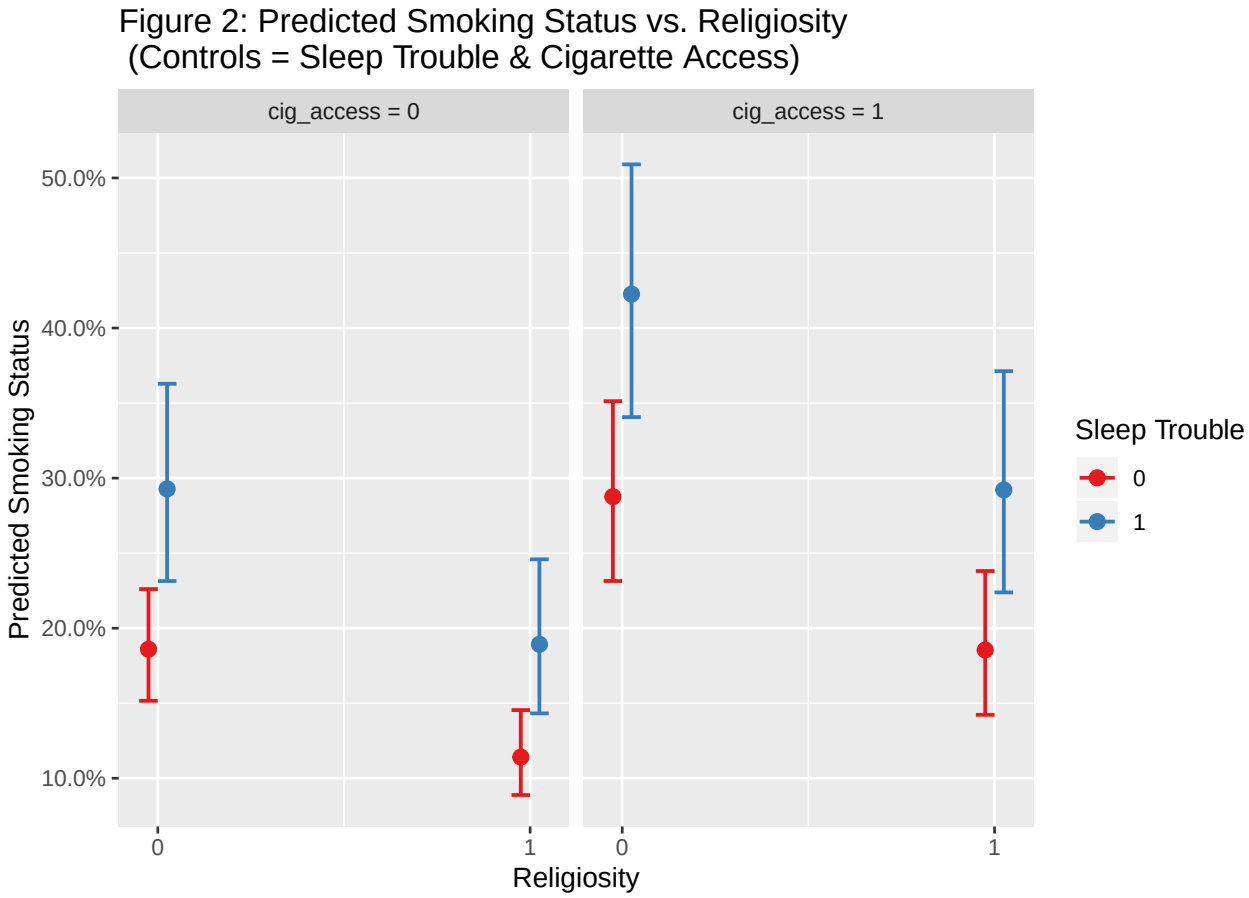
An interesting finding is that whether or not the respondent received smoking-related education in school had neither a practically significant nor a statistically significant impact on the odds of an adolescent being a regular smoker across all the models which had these covariates (models 2, 3 and 4). In models 3 and 4, we see sleep issues and stress-related education having a statistically significant effect (at the 0.01 and 0.05 levels respectively for model 3 and at the 0.01 and 0.1 levels respectively for model 4). The sleep issues variable was also practically significant with log-odd ratio values of 0.598 and 0.595, while stress-related education was less so, but still had high log-odd ratio values of 0.325 and 0.266 respectively. Although the parents' smoking histories variables (present only in model 4) are quite high in magnitude, they are not statistically significant at even the 0.1 level. Finally, access to cigarettes at home (also present only in model 4) proved to be very significant, both statistically and practically, with a log-odd ratio of 0.570. Thus, apart from religiosity, sleep issues (which in this study is chosen as the stress indicator) and access to cigarettes at home are the biggest factors associated with the odds of an adolescent being a regular smoker.

Table 2: Predicted Probability (Logged) of Respondent's Regular Smoker Status

	Dependent variable:			
	Predicted Probability of Being a Regular Smoker			
	Bivariate Model	3-Factor Model	5-Factor Model	Full Model
	(1)	(2)	(3)	(4)
Religiosity	−0.545***	−0.545***	−0.563***	−0.573***
	(0.147)	(0.147)	(0.148)	(0.150)
Female		−0.091	−0.149	−0.177
		(0.145)	(0.147)	(0.149)
Smoking-Related Education		−0.019	−0.076	0.008
		(0.292)	(0.299)	(0.305)
Trouble Sleeping			0.598***	0.595***
			(0.159)	(0.161)
Stress-Management Education			0.325**	0.266*
			(0.156)	(0.158)
Mother's Smoking History				0.219
				(0.162)
Father's Smoking History				0.249
				(0.167)
Cigarette Availability at Home				0.570***
				(0.162)
Constant	−1.083***	−1.021***	−1.309***	−1.827***
	(0.095)	(0.301)	(0.313)	(0.345)
Observations	1,170	1,170	1,170	1,170
Log Likelihood	−592.025	−591.824	−582.812	−569.113
Akaike Inf. Crit.	1,188.050	1,191.649	1,177.624	1,156.226
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 1: Predicted Smoking Status Based on Religiosity





The plots above visualize the results of the logistic model for the full model. Figure 1 depicts the predicted probability of an adolescent being a regular based solely on religiosity, while Figure 2 includes the statistically significant variables (at the 0.01 level) from the regression table, Table 2, as well. The two variables are trouble with sleep and cigarette access at home. From Figure 1, we see that the predicted probability is around 24% if the adolescent is areligious, compared to around 15.5% if they are religious, suggesting a noticeable change in the predicted probability based on religiosity.

Figure 2 corroborates the findings in Figure 1 in that it is observed that religiosity has a significant impact on smoking status, while trouble with sleep (used as the 'stress' variable) and cigarette availability at home are both strongly positively correlated with smoking status. In the case when the adolescent does not have access to cigarettes, the difference in the impact of sleep issues when the adolescent is areligious on the predicted probability (given by the vertical difference between the plots at the 0 level in the 'cig_access = 0' panel) is around 11%, and when the adolescent is religious is around 7%. In the case when the adolescent has access to cigarettes, the difference when the adolescent is areligious is around 13%, and around 10.5% when the adolescent is religious. We can also examine the effect of cigarette availability as we did with that of sleep issues above. In the case when the adolescent has sleep issues, the difference in the impact

of cigarette availability when the adolescent is areligious (we compare the blue plots at the 0 level between the 'cig_access = 0' and 'cig_access = 1' panels), is around 13%, and around 10% when the adolescent is religious. In the case when the adolescent does not have sleep issues (the red plots), the difference when the adolescent is areligious is 10%, and around 7.5%. Thus we can see a reduced impact of religiosity once we include the two significant control variables between Figures 1 and 2, as the difference in heights between the plots is not as high in Figure 2 as it is in Figure 1.

Discussion

This study has scope for improvement in a number of ways. One major limitation of this study is that the leading cause of smoking is stress, but there was no explicit stress variable and thus this study used 'having trouble sleeping' as the stress equivalent. However, other reasons can also contribute to having trouble sleeping. Moreover, having trouble sleeping can also contribute to more stress - thus the cyclical nature of this problem is not addressed here.

Second, while one of the biggest causes is peer pressure (Urberg et al., 1990), this study again suffers from the same problem as it does with the stress variable - there is no explicit question asking whether the respondent had their first smoking experience due to peer influence. Thus this study is subject to a certain degree of omitted variable bias due to the non-inclusion of an appropriate 'peer pressure' variable.

Third, this paper makes the strong assumption that all religious affiliations are equally opposed to smoking and dedicate equivalent resources, time and effort in addressing the issue of smoking among adolescents.

Fourth, the 'smoking-related instruction in schools' binary control variable does not take into consideration the quality of instruction and the content of the programs, which could have a great impact on the respondent's decision to smoke or not. Similarly, the control binary variable for 'stress-management instruction in school' does not account for the quality and content of the programs, which could vary considerably and therefore could have varying impacts on the resondent's ability to manage stress, and thus their decision to smoke.

Fifth, for the treatment variable, 'religiosity', in order to include the 879 atheists, the missing NA values had to be converted to 0. However, for the variable, there were altogether 890 missing values (as 3 corresponded to 'refused' and 8 corresponded to 'don't know'). But since it was not possible to extract just the missing NA values that corresponded to no religion (i.e. , the 'legitimate skips'), the other 11 responses were included as well. But the inclusion of these 11 reponses could affect the accuracy of the results obtained (although not significantly enough, as a result of which they were included). The same problem holds for the outcome

variable, 'smoke_reg' where 4 respondents refused to answer the question.

Finally, although this paper uses a relatively large seven control variables, it could have included more variables from the data that could impact the respondent's decision to smoke or not, such as alcohol consumption and other indicators of stress beyond trouble with sleep, such as poor appetite. This would augment the external validity of the study and make the findings more generalizable to contexts outside of this study.

Thus possible extensions of this paper could look to include more control variables. Possible extensions may also choose a more precise treatment variable that takes into consideration the specificities of opposition to smoking for each individual religion, using interaction terms, as well as looking more into religious participation (how often they pray, attend religious services) rather than just religiosity.

Conclusion

This paper adopts a multivariate logistic regression framework to examine the effect of religiosity on the smoking habits of adolescents, and a strong negative correlation is observed between the two variables, based on the model with full controls. This is in keeping with Koenig (2011), Wills et al. (2003) and Wallace et al (2003). However, once the statistically significant control variables of stress (in the form of trouble with sleep) and cigarette availability at home are included, some explanatory power of religiosity is reduced (although it remains statistically significant). An interesting observation is that stress (or sleep issues) plays a role in houses where cigarettes are easily accessible as well as those where they are not (a potential interaction term may be interesting to isolate the effect of these control variables in order to get a better understanding of the true effect of each). Finally, this study does not aim to make causal claims and simply states a strong negative correlation between the treatment and outcome variables. These results assume significance in the context of future projections of smoking, which are said to be higher.

Bibliography

Brown, Qiana L., et al. "The influence of religious attendance on smoking." Substance use & misuse 49.11 (2014): 1392-1399.

Hlavac, Marek. "Stargazer: Well-Formatted Regression and Summary Statistics Tables." (2018).

Garrusi, Behshid, Nouzar Nakhaee. "Religion and smoking: a review of recent literature." The International Journal of Psychiatry in Medicine 43.3 (2012): 279-292.

Urberg, Kathryn A., Shiang-Jeou Shyu, and Jersey Liang. "Peer influence in adolescent cigarette smoking." Addictive behaviors 15.3 (1990): 247-255.

Koenig HG. Spirituality and health research: Methods, measurement, statistics, and resources. West Conshohocken, PA: Templeton Press; 2011.

Wallace JM, Jr, Brown TN, Bachman JG, Laveist TA. The influence of race and religion on abstinence from alcohol, cigarettes and marijuana among adolescents. Journal of Studies on Alcohol. 2003;64(6):843–848.

Harris, Kathleen Mullan, and Udry, J. Richard. National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2008. Carolina Population Center, University of North Carolina-Chapel Hill, Inter-university Consortium for Political and Social Research, 2018-08-06.

Nollen LN, Catley D, Davies G, Hall M, Ahluwalia JS. Religiosity, social support, and smoking cessation among urban African American smokers. Addictive Behaviors. 2005;30(6):1225–1229.

Appendix

Figure 3: Visualization of Missing Data (Before Matching)

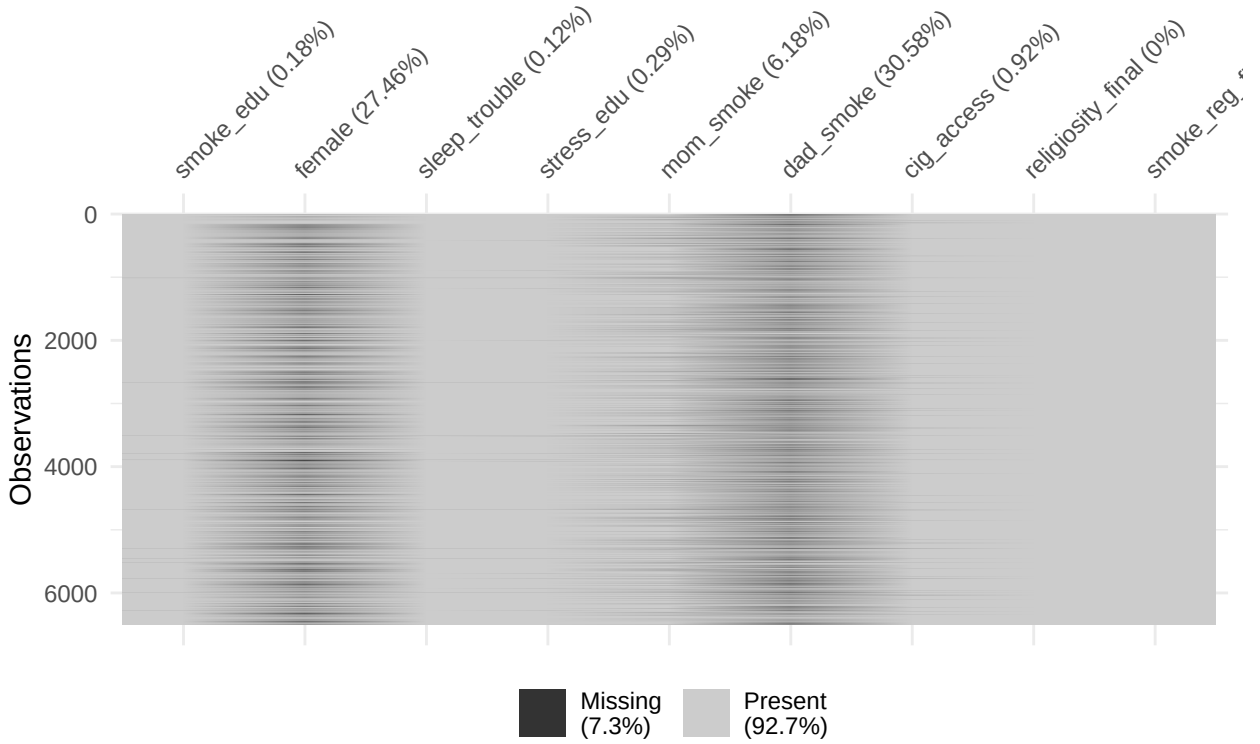
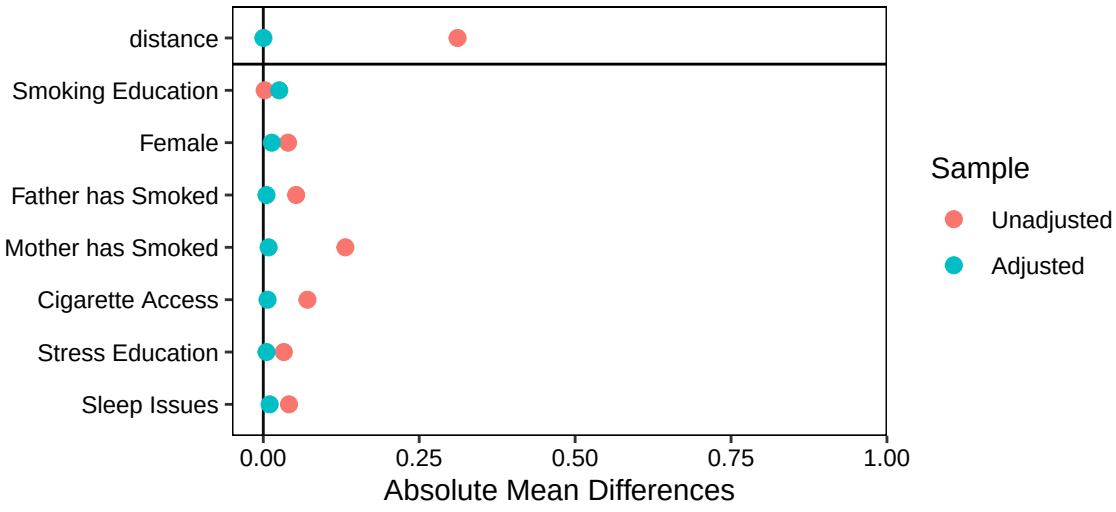


Figure 4: Covariate Balance



visuals-1.bb

Table 3: Summary Statistics (Before Matching)

Religiosity	# of Observations	% Who Smoked >= 1 Cigarette Daily	Std. Dev (Smoked >= 1 Cigarette Daily)	% With Smoking Education	% With Sleep Issues (>=1 Times Weekly)	% With Stress Education	% Whose Mother has Smoked	% With Cigarette Access	% Who are Female
0	593	25.46	0.44	93.59	24.96	61.05	59.19	33.73	48.23
1	2563	13.50	0.34	93.37	20.83	64.34	53.92	26.65	52.20

Table 4: ANOVA Comparison of Model 1 to Model 2

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1168	1184.050	NA	NA	NA
1166	1183.649	2	0.401	0.818

Table 5: ANOVA Comparison of Model 1 to Model 3

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1168	1184.050	NA	NA	NA
1164	1165.624	4	18.426	0.001

Table 6: ANOVA Comparison of Model 1 to Model 4

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1168	1184.050	NA	NA	NA
1161	1138.226	7	45.824	0