Analysis of Smoking Hazard and Socioethnographic Factors Influencing Smoking in the American Youth

-Draft-

Sergio E. Betancourt (sergio.betancourt@mail.utoronto.ca)

2019-04-01

Contents

Introduction

Tobacco consumption (most notably cigarette use) and its effects are a major public health challenge, and understanding its exposure to the youth can assist the creation of better policies and educational programs that should, not only inform young people, but also control and prevent consumption.

In this report I study the uptake of cigarette smoking, as well as consumption of other tobacco products, in the 2014 American National Youth Tobacco Survey, a survey of American teenagers, aged 9 to 19 years old, which collected not only indicators on the usage of different tobacco products, but also demographic data.

The following are the approaches taken in this report:

1. **Bayesian Survival Analysis: Uptake Rate of Cigarette Smoking in the American Youth**
2. **ML Logistic Regression: Usage of tobacco product alternatives in the American Youth**

1 Bayesian Survival Analysis: Uptake Rate of Cigarette Smoking in the American Youth

1.1 Executive Summary

I investigate the **uptake** of cigarette smoking in a representative sample of American teenagers, surveyed in the 2014+ American National Youth Tobacco Survey.

I find greater variation in the rate of smoking uptake between school than between states. Moreover, we find that two comparable (with identical confounders and random geographical effects), non-smoking children do not have the same probability of trying cigarettes within the next month–such probability increases with age. Ultimately, we find that rural, white teenagers are more likely to try their first cigarette earlier than their urban counterparts, controlling for race and sex.

I focus my analysis on the following formal hypotheses:

1. State-by-state variation in the mean age children first try cigarettes is greater than among-school variation.
2. Smoking the first cigarette has a flat hazard function, i.e., two non-smoking chidren have the same probability of trying their first cigarette within the next month, irrespective of their age, but controlling for confounders and random effects.

I also extend my analysis to understand how urban vs rural settings may influence smoking uptake in our sample–specifically, in the white population.

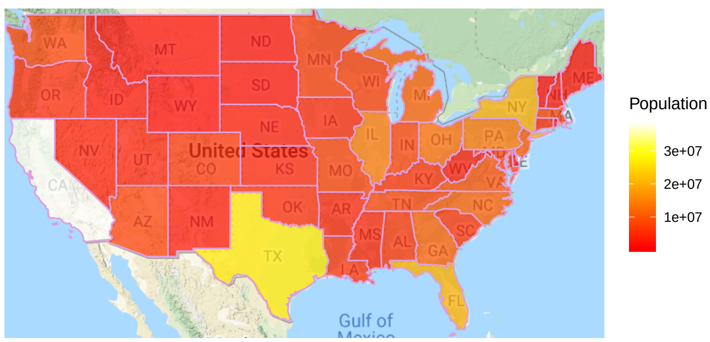


Figure 1: Map of the Contiguous United States with State Population (w. "ggplot2")

1.2 Methods

For all of the research questions at hand I consider the below hierarchical survival model, clustering by school and state.

$$Z_{ijk}|Y_{ijk}, A_{ijk}, U_i, V_{ij} = \min(Y_{ijk}, A_{ijk})$$
$$E_{ijk}|Y_{ijk}, A_{ijk}, U_i, V_{ij} = I(Y_{ijk} < A_{ijk})$$
$$Y_{ijk} \sim \text{Weibull}(\rho_{ijk}, \alpha)$$
$$\rho_{ijk} = \exp(-\eta_{ijk})$$
$$\eta_{ijk} = X_{ijk}\beta + U_i + V_{ij}$$
$$U_i \sim N(0, \sigma_U^2);\ V_{ij} \sim N(0, \sigma_V^2)$$

In the above:

- The dependent variable Y_{kij}, corresponds to the age when the kth child, from the jth school, of the ith state, tried cigarettes for the first time. This variable is left-censored, meaning that there are individuals who had not taken up smoking at the time of this survey, but may actually take it up in the future. To this end we define A_{ijk} to be the person's age when the survey was conducted.
- X_{ijk} corresponds to the covariates of rural vs. urban habitat, sex, and race, with an interaction between sex and race.

- U_i and V_{ij} are the state and school-level random effects, respectively.

For the issue at hand, which is to be modelled using Bayesian inference with the INLA package, I incorporate the following priors on the model parameters σ_V, σ_U, and α based on the experts' information:

$$\log(\alpha) \sim N(\log(1.5), 1); \quad \frac{1}{\sigma_U^2} \sim \text{PC}(.45, .1); \quad \frac{1}{\sigma_V^2} \sim \text{PC}(.2, .1)$$

The prior on $\log(\alpha)$ was chosen to be part of the normal family to allow for small and large α values. Expecting a flat hazard function which increases with age elicits a prior mean of log(1.5). Moreover, given the belief that the hazard function should increase at a quadratic or cubic rate with age, but not at a higher rate than this (with probability $< 10\%$), we let the standard deviation be 1. Thus we obtain the below quantiles on α:

Table 1: 10-50-90% percentiles of the prior distribution on α: log-Normal(log(1.5),1)

10%	Median	90%
0.416	1.5	5.403

The prior on the school-level random effect σ_V was chosen to communicate the belief that, in less than 10% of the cases, the healthiest schools in a given state should have an uptake rate greater than 50% that of the worst schools. This allows for a $\sigma_V \leq 0.2$, and we can use $P(\sigma_V \leq .2) = .99$, resulting in $\frac{1}{\sigma_V^2} \sim PC(.2, .1)$

Lastly, the prior on the state-level random effect σ_U was chosen to communicate a probable (less than 10%) difference between the healthiest and least healthy states of two or three times the smoking uptake rate, for which we can express the difference in the rate of smoking uptake between states i and j as $\exp\{U_i - U_j\}$. But noting how our hierarchical model already elicits $U_i \sim N(0, \sigma_U^2)$, which means we can think of the worst and best states being 2.6 standard deviations apart. Similarly to the prior on σ_V, we consider $P(\sigma_V \leq .45) = .99$, and thus we have $\frac{1}{\sigma_U^2} \sim \text{PC}(.45, .1)$.

1.3 Results

We first fitted the saturated model with all possible interactions, and then settled for keeping two two-way interactions between Race and Rural vs Urban setting, and Race and Sex. The saturated model diluted the effects of most covariates.

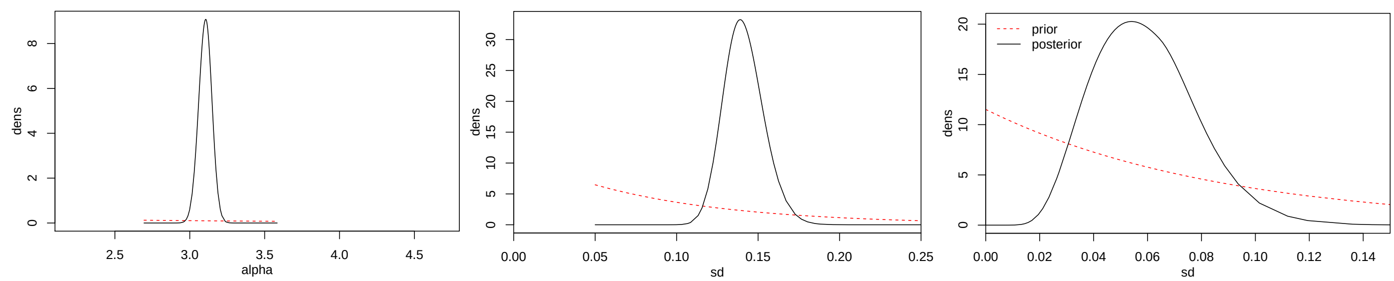
Table 2: Posterior mean and 2.5% and 97.5% percentiles for the relative rate (with respect to the baseline category: white, urban males) of smoking uptake by model coefficients

	mean	0.025quant	0.975quant
(Intercept)	0.545	0.515	0.576
RuralUrbanRural	1.132	1.064	1.205
Raceblack	0.985	0.932	1.040
Racehispanic	1.044	1.000	1.090
Raceasian	0.822	0.745	0.902
Racenative	1.106	0.952	1.270
Racepacific	1.108	0.882	1.352
SexF	0.952	0.926	0.978
RuralUrbanRural:Raceblack	0.945	0.885	1.009
RuralUrbanRural:Racehispanic	0.972	0.922	1.025
RuralUrbanRural:Raceasian	1.067	0.927	1.223
RuralUrbanRural:Racenative	1.011	0.863	1.191
RuralUrbanRural:Racepacific	1.149	0.884	1.510
Raceblack:SexF	0.987	0.934	1.043
Racehispanic:SexF	1.020	0.976	1.066
Raceasian:SexF	0.997	0.882	1.126
Racenative:SexF	0.959	0.824	1.112
Racepacific:SexF	0.839	0.609	1.114
SD for school	0.141	0.119	0.167
SD for state	0.059	0.027	0.102

Table 3: Posterior mean, standard deviation and percentiles of parameters $\alpha, 1/\sigma_V^2$ and $1/\sigma_U^2$.

	mean	sd	2.5%	median	97.5%
α	3.103	0.044	3.016	3.104	3.188
Precision for School ($1/\sigma_V^2$)	51.216	8.832	35.783	50.588	70.415
Precision for State ($1/\sigma_U^2$)	406.648	360.072	97.065	299.503	1355.947

Below are the diagrams contrasting our prior distributions on the model parameters α, σ_U, and σ_V, with their corresponding posterior densities:



(a) α

(b) SD for school (σ_V)

(c) SD for state (σ_U)

Figure 2: Densities of prior and posterior distributions pertaining to this hierarchical survival model

1.4 Discussion

With regards to the first research question, the fitted model provides evidence **against** the initial hypothesis: We estimate greater variation in smoking uptake between schools ($\hat{\sigma}_U = .142$) than between states ($\hat{\sigma}_U = .057$). We see the following in the above estimates for the posterior mean on the precision of these two, as well as the corresponding plots: $\sigma_V^2 < \sigma_U^2$.

- The above suggests a modification to the proposed tobacco control program outreach: We should seek to identify the schools with the highest smoking uptake rate and target them instead of tackling this issue at the state level.

With regards to the second research question, the initial hypothesis of a flat hazard function–meaning that two non-smoking children have the same probability of trying cigarettes within the next month, irrespective of their ages–can be condensed to $\alpha = 1$. Nonetheless, our posterior mean and plot of α provide evidence for a median value near 3.1, with a relatively tight 95% interval **(3.019,3.191)**. This provides strong evidence **against** a flat hazard function–namely, $\alpha = 3.1$ suggests a cubic increase in smoking uptake as age increases.

- The above suggests that children with identical (known) confounders and random effects, have different probabilities of trying cigarettes in the next month, where the elder is more likely to try them than the younger one. Therefore, it is wise to tailor tobacco control programs to age, perhaps with more reactive methods as the children enter high school, but more proactive ones with younger children.

Lastly, with regards to the difference in cigarette smoking uptake between white rural males and their urban counterparts, our model suggests a mean difference in the smoking uptake rate between **1.064** and **1.205** for rural vs urban with 95% credibility (seen in the table above). This suggests that rural white individuals are more likely to try the first cigarette earlier than their urban counterparts, controlling for race and sex.

- To address the above habitat difference in smoking uptake, we may consider addressing the white, rural population differently, perhaps with greater focus or more tailored messages to their lifestyle and values.

2 ML Logistic Regression: Usage of tobacco product alternatives in the American Youth

2.1 Executive Summary

Perceived to be healthier alternatives to smoking cigarettes, chewing tobacco products and smoking tobacco from a hookah or water pipe are still harmful to the health. I investigate the use of chewing tobacco products and hookah and waterpipes in a representative sample of American teenagers, surveyed in the 2014 American National Youth Tobacco Survey.

I find that Americans of European Descent are more likely, than Hispanics and Blacks, to chew tobacco products regularly, in both Urban and Rural settings. However, among whites, regular consumption of tobacco chew appears more likely in a rural setting.

I also find no significant evidence for differences in the likelihood of ever using a hookah or water pipe between the observed males and females, controlling for age, ethnicity, and other demographic characteristics.

Ultimately, I extend the scope of this study by quantifying variations in regular usage of tobacco chew with respect to changes in age, sex, and ethnic group. We find that, on average, individuals of the White and Pacific ethnicities exhibit the highest probability of consuming tobacco chew regularly, with Asians exhibiting the least. We also find that, compared to males, females exhibit lower likelihoods of consuming these products regularly across all races. Lastly, I find that increments in age decrease the likelihood for females to engage in these practices, while increase it for males.

I focus my analysis on the following formal questions:

- Regular use of chewing tobacco products is equally pervasive among Americans of European descent and Hispanic and African Americans.
- The likelihood of having ever been exposed to hookah or water pipe does not change with sex, controlling for age, ethnicity, and other demographic characteristics.

I also extend my analysis by modelling changes that a number of factors have in the likelihood of consuming tobacco chew products regularly in our sample.

2.2 Methods

The research questions above elicit logistic regression (Generalized linear models with binomial distribution) models. These models allow us to model and quantify the likelihood of a given binary outcome of interest given a number of covariates. It is important to recall that a binomial distribution may be characterized as a sum of bernoulli random variables. I formulate the below models as bernoulli given the inclusion of covariates, which allow us to 'profile' types of subject by covariates.

For the first research question I modeled the likelihood of chewing tobacco in the sample. I set up the Americans of European ancestry ethnicity as the baseline group, and compared it with Hispanic-Americans and African-Americans. This model includes Race and Urban status, as well as their interaction, as the explanatory variables, with the binary indicator of ever chewing tobacco products as the dependent variable. Consequently, the aggregation of these can be interpreted as incidence rates within the data.

Model 1

I have,

$$Y_i \sim \text{Bernoulli}(\mu_i),\ \text{where}$$
$$Y_i = \text{indicator for indidual i chewing tobacco regularly}$$
$$\text{and } \mu_i = \text{probability of subject chewing tobacco regularly given covariates } \mathrm{X}_i$$

which leads to the logistic regression model,

$$\log(\frac{\mu_i}{1-\mu_i}) = \beta_0 + \beta_1 \cdot \mathrm{I}_{\text{Race},i} + \beta_2 \cdot \mathrm{I}_{\text{RuralUrban},i} + \beta_3 \cdot \mathrm{I}_{\text{Race}\times\text{RuralUrban},i}$$

To address the second research question I fit a logistic regression model with the indicator for having ever used hookah or water pipe as the predictor, and sex, age, and race as indicators, as explanatory variables, without interactions between them.

Below is the model formulation:

Model 2

I have,

$$Y_i \sim \text{Bernoulli}(\mu_i),\ \text{where}$$
$$Y_i = \text{indicator for indidual i having ever smoked out of a hookah or water pipe}$$
$$\text{and } \mu_i = \text{probability of subject i ever using a hookah given covariates } \mathrm{X}_i$$

which leads to the logistic regression model,

$$\log(\frac{\mu_i}{1-\mu_i}) = \beta_0 + \beta_1 \cdot \mathrm{I}_{\text{Sex},i} + \beta_2 \cdot \mathrm{I}_{\text{Age},i} + \beta_3 \cdot \mathrm{I}_{\text{Race},i}$$

As an extension to this study, I seek to quantify how the practice of chewing tobacco changes with age, sex, and ethnicity. I fit a logistic regression model that uses an indicator of chewing tobacco regularly as the predicted variable, and age and sex, as well as their interaction, and race, as explanatory variables.

Model 3

I have,

$$Y_i \sim \text{Bernoulli}(\mu_i),\ \text{where}$$
$$Y_i = \text{indicator for indidual i chewing tobacco regularly}$$
$$\text{and } \mu_i = \text{probability that subject i chews tobacco regularly given covariates } \mathrm{X}_i$$

which leads to the logistic regression model,

$$\log(\frac{\mu_i}{1-\mu_i}) = \beta_0 + \beta_1 \cdot \mathrm{I}_{\text{Age},i} + \beta_2 \cdot \mathrm{I}_{\text{Sex},i} + \beta_3 \cdot \mathrm{I}_{\text{Race},i} + \beta_4 \cdot \mathrm{I}_{\text{Age}\times\text{Sex},i}$$

2.3 Results and Discussion

As it pertains to the first research question, **Model 1** suggests statistically significant differences (at the 5% level) for likelihood of ever chewing tobacco for blacks and hispanics when compared to whites, across both urban and rural areas. Below are the estimates of the odds and their 95% confidence intervals (Table 1), and estimated probabilities (Table 2):

Table 4: Exponentiated coefficient estimates (Odds) and their 95% Confidence Interval

	Estimate	Lower Bound	Upper Bound
(Intercept)	0.026	0.022	0.032
Raceblack	0.379	0.232	0.618
Racehispanic	0.731	0.540	0.989
RuralUrbanRural	3.162	2.558	3.908
Raceblack:RuralUrbanRural	0.476	0.246	0.920
Racehispanic:RuralUrbanRural	0.565	0.381	0.836

Table 5: Probability estimates of chewing tobacco products at least once in the past 30 days

	Urban	Rural
White	0.026	0.077
Black	0.010	0.005
Hispanic	0.019	0.011

The table above indicates that whites appear to have the highest likelihood for chewing tobacco, across both rural and urban domains, followed by Hispanics, and then Blacks. All of these estimates are statistically significant. It is interesting to note that although an average white individual in the sample does exhibit a higher likelihood for chewing tobacco if they are in a rural setting (caompared to an urban setting), blacks and Hispanics show the opposite within their ethnicities. An average individual of either of these two ethnicities appears to be more likely to chew tobacco in an urban setting than in a rural setting.

As it pertains to the second research question, **Model 2** does not present evidence against the null hypothesis of no difference in likehood of having ever used one of these devices for individuals of different sexes, controlling for age, ethnicity and other demographic traits.

Table 6: Exponentiated coefficient estimates (Odds) and their 95% Confidence Interval

	Estimate	Lower Bound	Upper Bound
(Intercept)	0.026	0.022	0.032
Raceblack	0.379	0.232	0.618
Racehispanic	0.731	0.540	0.989
RuralUrbanRural	3.162	2.558	3.908
Raceblack:RuralUrbanRural	0.476	0.246	0.920
Racehispanic:RuralUrbanRural	0.565	0.381	0.836

The above output does show significant differences in hookah consumption based on age and ethnicity.

With regards to the secondary research problem, **Model 3** estimates the odds of chewing tobacco products in the last 30 days, controlling for age (centered), ethnicity, sex, and interactions of age and sex, and age and race.

In Table 4 below I report the factors found to have a statistically significant effect on the likelihood of chewing tobacco products. Note that age has different impact based on sex and ethnicity, where increases in age increase the odds of chewing tobacco in our male sample, but decrease the odds in the females. However, the reduction in the likelihood of using these products occurs across the board when comparing males versus females.

Table 7: Exponentiated coefficient estimates (Odds) and their 95% Confidence Interval

	Estimate	Lower Bound	Upper Bound
(Intercept)	0.064	0.056	0.072
Age_Cent	1.607	1.523	1.696
Raceblack	0.277	0.196	0.393
Racehispanic	0.583	0.475	0.717
Raceasian	0.249	0.130	0.476
Racepacific	3.045	1.484	6.247
SexF	0.230	0.186	0.285
Age_Cent:Raceblack	0.731	0.628	0.851
Age_Cent:Racehispanic	0.738	0.672	0.811
Age_Cent:Raceasian	0.583	0.439	0.775
Age_Cent:SexF	0.671	0.607	0.741

The effect of age varies by race–a one year increment in age reduces the odds of smoking the most in the Asian population. It also reduces the odds on the Hispanic and the Black ethnicities, but increases that of whites.

In the below table I see how the estimated probability of chewing these products by ethnicity and sex, where individuals of the Pacific ethnicity appear with the highest likelihood, followed by whites. As mentioned above, the estimated probabilities are higher for males than females, across these races.

Table 8: Probability estimates of chewing tobacco products in the past 30 days

	White	Black	Hispanic	Asian	Pacific
Male	0.060	0.017	0.036	0.016	0.074
Female	0.014	0.004	0.008	0.004	0.018

3 Appendix: Code

```
library(MASS); library(lmtest); library(knitr); library(kableExtra); library(nleqslv); library(lme4); library
library(Pmisc); library(extrafont); library(VGAM); library(INLA); library(plyr); library(dplyr); library(MEMS
knitr::opts_chunk$set(fig.pos = 'H');

# VISUALIZATION
# Loading state coords and info for visualization
usstate <- read.csv("/Users/Balthazar/Desktop/Grad_School/Professional/Portfolio/Smoking/state-info.csv", hea
googkey <- readLines("/Users/Balthazar/Desktop/Grad_School/Professional/Portfolio/google_map_key.txt")

ggmap::register_google(key = googkey);

p0 <- ggmap(get_googlemap(center = c(lon = -98, lat = 42),
                    zoom = 3, scale = 4,
                    maptype ='terrain',
                    color = 'color'), maprange=T,extent = "normal") +
    labs(x = "", y = "") +
    scale_x_continuous(limits = c(-124.848974, -66.885444), expand = c(0, 0)) +
scale_y_continuous(limits = c(24.396308, 49.384358), expand = c(0, 0)) +
  theme(legend.position = "right",
        panel.background = element_blank(),
        axis.line = element_blank(),
        axis.text = element_blank(),
        axis.ticks = element_blank(),
        plot.margin = unit(c(0, 0, -1, -1), 'lines')) +
  xlab('') +
  ylab('')


if (requireNamespace("sf", quietly = TRUE)) {
  library(sf)
  data(us_states)
}

us_states <- as(us_states,'Spatial')
us_states@data$id = rownames(us_states@data)
us_states.points = fortify(us_states, region="id")
us_states.df = join(us_states.points, us_states@data, by = "id")

# Blank canvas with state divisions - boundary map "fifty_states"
p1 <- p0 + geom_polygon(data=us_states.df, aes(long,lat,group=group,fill=total_pop_15), alpha=.7,color="plum"

p1

# Aggregate counts by state (adding extra criteria) for visualization
#forVis

# Show total counts data
#ggplot() + geom_polygon(data=fifty_states, aes(x=long, y=lat, group = group),color="white", fill="grey92" )
#  geom_point(data=mapdata, aes(x=lon, y=lat, size = medals), color="black") +
#  scale_size(name="", range = c(2, 20)) +
#  guides(size=guide_legend("GABF medals 1985-2015")) +
#  theme_void()

# Show multiple data points
#ggplot() + geom_polygon(data=fifty_states, aes(x=long, y=lat, group = group),color="white", fill="grey92" )
#geom_point(data=mapdata, aes(x=long, y=lat, size = killed, color=year)) +
```

```r
#scale_size(name="", range = c(2, 15)) +
#guides(size=guide_legend("deaths")) +
#theme_void()
prior_a_mean <- log(1.5)
prior_a_sd <- 1
prioralpha <- t(exp(qnorm(c(0.1, .5, 0.9), mean=prior_a_mean, sd=prior_a_sd)))
colnames(prioralpha) <- c("10\\%", "Median", "90\\%")

knitr::kable(prioralpha, digits=3, escape=F, format="latex", booktab=T,linesep = "",caption="10-50-90\\% perc
  kable_styling(latex_options = "hold_position")
##### Part 3 #####

#Loading and prepping the dataset
load("/Users/Balthazar/Desktop/Grad_School/COURSEWORK/Spring 2019/App Stats II/data/smoke.RData")

forInla = smoke[, c("Age", "Age_first_tried_cigt_smkg",
"Sex", "Race", "state", "school", "RuralUrban")]
forVis = forInla
forInla = na.omit(forInla)
forInla = as.list(forInla)

forSurv = data.frame(time = (pmin(forInla$Age_first_tried_cigt_smkg,
forInla$Age) - 4)/10, event = forInla$Age_first_tried_cigt_smkg <=
forInla$Age)

# left censoring
forSurv[forInla$Age_first_tried_cigt_smkg == 8, "event"] = 2;
forInla$y = inla.surv(forSurv$time, forSurv$event);
# Customized model with new priors
#####################
# Priors:
# \log(\alpha) \sim N(\log(1.5),1)
# \frac{1}{\sigma_U^2} \sim \text{PC}(.45,.1)
# \frac{1}{\sigma_V^2} \sim \text{PC}(.2,.1)
prior_state <- .45
#exp(2.6*prior_State)
prior_school <- .2
#exp(2.6*prior_School)

prior_a_mean <- log(1.5)
prior_a_sd <- 1
#exp(qnorm(c(0.1, 0.9), mean=prior_a_mean, sd=prior_a_sd))
######################

fitS2 = inla(y ~ RuralUrban*Race + Sex*Race +
               # School cluster
               f(school,model = "iid", hyper = list(prec = list(prior = "pc.prec",
param = c(prior_school, .1)))) +
  # State cluster
  f(state, model = "iid", hyper = list(prec = list(prior = "pc.prec", param = c(prior_state,
.1)))),
#fixed effects
control.family = list(variant = 1,
                      hyper = list(alpha = list(prior = "normal", param = c(prior_a_mean,
                                                                            prior_a_sd^(-2))))),
data = forInla, family = "weibullsurv")
# Summarizing posteriors
resTable1 <- exp(fitS2$summary.fixed[, c("mean", "0.025quant",
```

```
"0.975quant")]);
resTable2 <- Pmisc::priorPostSd(fitS2)$summary[,
c("mean", "0.025quant", "0.975quant")]
restable <- rbind(resTable1,resTable2)

knitr::kable(restable, digits=3, escape=F, format="latex", booktab=T,linesep = "", caption="Posterior mean an
  kable_styling(latex_options = "hold_position")
#Plotting priors and posteriors
#\alpha, sd state, and sd school
fitS2$priorPost = Pmisc::priorPost(fitS2)

par(mar = c(4,4,4,2) + 0.1);
par(mgp=c(2,1,0));

for (Dparam in fitS2$priorPost$parameters) {
  do.call(matplot, fitS2$priorPost[[Dparam]]$matplot)
}
fitS2$priorPost$legend$x = "topleft"
do.call(legend, fitS2$priorPost$legend)
post_summary <- data.frame(fitS2$summary.hyperpar)[,1:5]
rownames(post_summary) <- c("$\\alpha$",  "Precision for School $(1/\\sigma_V^2)$","Precision for State $(1/\
colnames(post_summary) <- c("mean", "sd", "2.5\\%","median","97.5\\%")
kable(post_summary, format="latex", escape=F, digits=3, booktab=T,linesep = "", caption="\\label{post} Poster
# Downloading datafile "smoke" from course website
dataDir = "/Users/Balthazar/Desktop/Grad_School/COURSEWORK/Spring 2019/App Stats II/data"
smokeFile = file.path(dataDir, "smoke.RData")
if (!file.exists(smokeFile)) {
  download.file("http://pbrown.ca/teaching/astwo/data/smoke.RData", smokeFile)
}
smokedat <- load(smokeFile)

# To view Data doc use:
# View(smokeFormats)

#for(D in c('chewing_tobacco_snuff_or', 'ever_tobacco_hookah_or_wa')){
#cat("- `", D, "`: ",
#as.character(smokeFormats[match(D, smokeFormats[,'colName']), 'label']),
#'\n\n', sep='')
#}

chew_mod <- glm(chewing_tobacco_snuff_or ~ Race*RuralUrban, family='binomial', data=smoke)
#round(summary(chew_mod)[['coefficients']],digits=3)

#confint(chew_mod)
# Odds and prob point estimates
chew_coeffs <- summary(chew_mod)[['coefficients']][c(1:3,7:9)]
#Transforming log-odds to odds
urban_likelihoods <- exp(c(chew_coeffs[1],chew_coeffs[1]+chew_coeffs[2],
                           chew_coeffs[1]+chew_coeffs[3]))
#Calculating probabilities
urban_probs <- round(urban_likelihoods/(1+urban_likelihoods),digits=3)

#Transforming log-odds to odds
rural_likelihoods <- exp(c(chew_coeffs[1]+chew_coeffs[4],   chew_coeffs[1]+
                             chew_coeffs[2]+chew_coeffs[5],chew_coeffs[1]+chew_coeffs[3]+chew_coeffs[6]))

#Calculating probabilities
rural_probs <- round(rural_likelihoods/(1+rural_likelihoods),digits=3)
```

```r
chew_mod1_results <- data.frame(t(matrix(c(urban_probs,rural_probs),nrow=2,ncol=3,byrow=T)),
                                row.names=c("White","Black","Hispanic"))

# Confidence intervals
theCiMat = Pmisc::ciMat(.95);
parTable = summary(chew_mod)$coeff[,rownames(theCiMat)] %*% theCiMat;
parTable = exp(parTable[c(1:3,7:9),]);


# Reported table of estimates odds
knitr::kable(parTable, "latex", caption= "Exponentiated coefficient estimates (Odds) and their 95\\% Confiden
             booktabs =T, digits=3, col.names =  c("Estimate","Lower Bound","Upper Bound")) %>%
  kable_styling(latex_options = "hold_position") #%>%
  #add_header_above(c("Urban"=1, "Rural"=2))


# Reported table of estimated probabilities
knitr::kable(chew_mod1_results, "latex", caption= "Probability estimates of  chewing tobacco products at leas
             booktabs =T, digits=3, col.names =  c("Urban","Rural")) %>%
  kable_styling(latex_options = "hold_position") #%>%
  #add_header_above(c("Urban"=1, "Rural"=2))
#Controlling for age, ethnicity, and other dem chars, there is no statistical evidence for difference in the

#Using no interaction terms

#Centering Age as Age2
#Average age in data set
avg_age <- mean(na.omit(smoke$Age))
Age_Cent <- smoke$Age - avg_age
smoke <- cbind(smoke,Age_Cent)

#reference is male. female is treatment
hookah_mod <- glm(ever_tobacco_hookah_or_wa ~ Sex+Age_Cent+Race, family='binomial', data=smoke)

repp <- exp(summary(hookah_mod)$coefficients);

# Reported table of estimates odds
knitr::kable(parTable, "latex", caption= "Exponentiated coefficient estimates (Odds) and their 95\\% Confiden
             booktabs =T, digits=3, col.names =  c("Estimate","Lower Bound","Upper Bound")) %>%
  kable_styling(latex_options = "hold_position") #%>%
  #add_header_above(c("Urban"=1, "Rural"=2))
chew_quant_mod0 <- glm(chewing_tobacco_snuff_or ~ Age_Cent*Sex*Race, family='binomial', data=smoke)

chew_quant_mod1 <- glm(chewing_tobacco_snuff_or ~ Age_Cent*Race + Age_Cent*Sex, family='binomial',
                       data=smoke)

#Choosing simpler model chew_quant_mod1
#lmtest::lrtest(chew_quant_mod0,chew_quant_mod1)

##Model Coefficients
# Confidence intervals
theCiMat = Pmisc::ciMat(.95)
parTable2 = summary(chew_quant_mod1)$coeff[,rownames(theCiMat)] %*% theCiMat
chew_q_coeffs <- parTable2[,1]
expparTable2 = exp(parTable2[c(1:5,7:11,14),])


# Reported table of estimates odds
```

```r
knitr::kable(expparTable2, "latex", caption= "Exponentiated coefficient estimates (Odds) and their 95\\% Conf
             booktabs =T, digits=3, col.names =  c("Estimate","Lower Bound","Upper Bound")) %>%
  kable_styling(latex_options = "hold_position") #%>%
#white, black, hispanic, asian

odds_male <- as.numeric(c(chew_q_coeffs[1],
               chew_q_coeffs[1]+chew_q_coeffs[3],
               chew_q_coeffs[1]+chew_q_coeffs[4],
               chew_q_coeffs[1]+chew_q_coeffs[5],
               chew_q_coeffs[1]+chew_q_coeffs[6]))
odds_female <- as.numeric(c(chew_q_coeffs[1],
               chew_q_coeffs[1]+chew_q_coeffs[3],
               chew_q_coeffs[1]+chew_q_coeffs[4],
               chew_q_coeffs[1]+chew_q_coeffs[5],
               chew_q_coeffs[1]+chew_q_coeffs[6])) + as.numeric(chew_q_coeffs[8]);
odds_male <- exp(odds_male)
odds_female <- exp(odds_female)
probs_male <- odds_male/(1+odds_male)
probs_female <- odds_female/(1+odds_female)

quant_df <- as.data.frame(rbind(probs_male,probs_female),row.names=c("Male","Female"))

# Reporting table of estimates probabilities
knitr::kable(quant_df, "latex", caption= "Probability estimates of chewing tobacco products in the past 30 da
             booktabs =T, digits=3, col.names =  c("White","Black","Hispanic","Asian","Pacific")) %>%
  kable_styling(latex_options = "hold_position")
```