Lifestyle Factors Associated with Feeling Down or Depressed in American Adults Aged 25 - 65

Kristen Keller

TABLE OF CONTENTS

Introduction

While it may be interesting to know what lifestyle factors are associated with decreased happiness, there has been little research on this topic. In this study we use a Bayesian ordinal logistic regression model to investigate associations between explanatory variables and reports of feeling down or depressed. The explanatory variables examined in our study are age, gender, high school education, college education, marital status, income to poverty ratio, employment status, and children under 5. The outcome variable in our model reflects a subject's answer to the question "Over the last 2 weeks, how often have you been bothered by the following problem: feeling down, depressed, or hopeless?" A response of 1 indicates the subject did not report feeling down at all, a response of 2 indicates the subject reported feeling down several days, and a response of 3 indicates the subject felt down more than half of the time.

Data were taken from the 2013-2013 National Health and Nutrition Examination Survey (NHANES). This survey is administered every year to a sample of approximately 5,000 non-institutionalized United States citizens. There are a total of 5924 participants in the dataset, about 3830 of which are in our target age range of 25-65. Of the participants in the specified age range, 3201 have completely observed covariates and 629 have one or more missing covariate. Only participants with completely observed covariates were considered for our study. Note that the missing data was almost exclusively attributed to the income variable (n = 294) and the outcome variable (n = 381).

Model Assumptions and Prior Specification

The most important assumption for our model is the proportional odds assumption. To help elucidate this assumption, say we have an outcome variable Y_i that can take on j ordered levels 1 … J. Now say we create a new outcome variable, Z_i, such that Z_i is equal to 0 if Y_i is less than a cutoff, j^*, in 2… J and Z_i is equal to 1 if Y_i is greater than or equal to j^*. If the proportional odds assumption holds, the vector of regression coefficients governing the relationship between the covariates and the outcome variable Z_i will be the same for any choice of j*. Here there are only three levels the outcome variable can take on, so we just need the vector of logistic regression coefficients to be the same when the outcome variable categorized as $Z_i = 1$ if $Y_i = 2,3$ and 0 otherwise, and when $Z_i = 1$ if $Y_i = 3$ and 0 otherwise. There do not appear to be any major violations of the proportional odds assumption for our data (Figure 1).

After model assumptions were assessed, priors were specified for model parameters. Normal priors were used for all β coefficients and covariances between β coefficients were assumed to be 0. Assuming the proportional odds assumption holds, β coefficients from standard logistic regression models comparing patients with no sign of discontent to patients with any sign of discontent should be comparable to coefficients for our ordinal logistic model. This is because of the way our outcome variable was constructed, with 1 representing patients that did not feel down and 2 and 3 representing patients that did.

All prior information was elicited from previously published articles (Table 1). For most variables, confidence intervals for odds ratios were presented. Our β coefficients are log odds ratios, so when a confidence intervals for odds ratios were presented, the natural log of each endpoint was taken and the resulting values were specified as the 10^{th} and 90^{th} percentiles of a normal distribution. The mean was taken to be the midpoint between these two points and the mean and 10^{th} percentile were used to solve for precision.

Odds ratio confidence intervals were available for the covariates gender, marital status, college education, and employment. For the high school education variable, prevalence values were used to construct an odds ratio point estimate of 0.534. The interval (0.284,0.784) was used in place of a confidence interval. Age and income to poverty ratio were both continuous predictors, but information was readily available on similar indicators in categorical form. For both of these variables, the upper limit of the original confidence interval was used as the upper limit of our confidence interval and the lower limit of our confidence interval was specified to be much lower than that of the original confidence interval. The resulting interval for age was (0.3,1.84). For income to poverty ratio, the odds ratio for the lowest income group compared to highest was presented, but we were interested in an interval for the highest income group compared to lowest, so we took the reciprocal of each endpoint and proceeded as before to get an interval of (0.15,1.27). It was difficult to find information on the children variable so we used an interval of (0.3,1.2), which allows for a positive or negative association.

Next prior distributions for the α coefficients were specified. α_1 is associated with the logit of the probability that a person with all 0 covariates will have outcome 1, so a normal prior was used. We chose 65% and 85% as the 5^{th} and 95^{th} percentiles for the number of people with outcome 1 and solved for the mean and precision as before. This translated to a prior with a mean of 1.17 and a precision of 8.70 on the logit scale. α_2 corresponds to the logit of the probability that a person with all 0 covariates will have outcome of 1 or 2, so α_2 must be greater than α_1. We set the prior for α_2 equal to α_1 plus a gamma variable. We expect about 90% of people to have outcome 1 or 2, so we set the mean of the prior for α_2 to 1.02, which is the difference between the logit of 0.90 and the mean of the α_1 prior. We set the variance of the prior for α_2 to be 0.15, which is about twice the variance of α_1.

Results and Convergence

There were no major convergence issues. A preliminary model was run for 10,000 iterations with 3 chains. Based on the autocorrelation plots, we decided to run our next model for 40,000 iterations. This appeared to be a reasonable number based on the time series plots.

The odds of feeling down with higher frequency are as follows. For a female compared to a male the odds are 1.51, for someone who is married compared to someone who is not the odds are 0.74, for someone who is unemployed compared to someone who is employed the odds are 2.17, and for someone who has a child under 5 compared to someone who does not the odds are 0.67. A one-year increase in age multiplies the odds by 1.01 and a one-unit increase in income to poverty ratio multiplies the odds by 0.81. There did not appear to be a significant relationship between education level and frequency of feeling down (Table 2, Figure 3).

Sensitivity Analysis

Although the β vector is the main parameter of interest in our analysis, misspecification of priors for α_1 and α_2 might lead to inaccurate posteriors for the β coefficients. It appears that when we specified our prior for α_2, we misestimated the number of people with an outcome of one or two (Figure 4). We conducted a sensitivity analysis in order to determine whether our α prior specification affected the study results. To allow for more uncertainty in the distribution of α_1 and α_2, the variances for the associated normal and gamma distributions were multiplied by 10. The only β for which a large difference was observed was the β associated with age, for which the distribution shifted closer to 0. The distributions for the α coefficients were both shifted to lower values, so we would expect more subjects to report feeling down under this model (Figure 5).

We ran an additional model where we made our α and β priors less informative by multiplying the variances by 10. As in the previous model, the variable age did not have a significant relationship with the outcome variable. Based on the sensitivity analysis, and the fact that age was just barely significant in the original model, we would conclude that age is not associated with the outcome. Other than that no major conclusions changed, though in the model with relaxed α and β priors the posterior distribution for college education was shifted much farther away from zero and the variable was close to being significant (Figure 6).

Conclusion

Based on our model, we conclude that gender, unemployment, marital status, income to poverty ratio, and young children are all associated with feeling down, whereas age and education level are not. Specifically, being male, being employed, being married, having young children, and having a high income to poverty ratio were all associated with fewer feelings of discontent. Age was not an important explanatory variable in our model, but has been shown to be associated with both happiness and depression in many other studies. In most of these studies age was recoded as a categorical variable representing age ranges, so it is possible age might have been significant if we had coded it as categorical.

When interpreting our results, we must keep in mind the sizable amount of missing data in our study, most of which was attributed to the outcome variable and the income to poverty variable. The missing data will likely bias our β coefficients and make it infeasible to make inference on the average number of people in each response category using the α variables. For example, the depression screener containing the outcome variable was not administered for participants that did not participate in the mobile examination center health evaluation portion of the study. It is possible that patients with more mental health issues and patients that were feeling depressed were less likely to go to the mobile examination center and therefore they were less likely to have a recorded outcome variable. Additionally, lower income was shown to be associated with lower levels of the outcome variable and it is likely that some people did not report their income because they have no income or inconsistent income. Based on this reasoning, we might have expected a large proportion of the subjects with missing data to report feeling down, though arguments can also be made for biases in the other direction.

References

Erickson, El-Gabalawy, Palitsky, Patten, Mackenzie, Stein, and Sareen. Educational Attainment as a Preventative Factor for Psychiatric Disorders: Findings From a Nationally Representative Longitudinal Study. Depress Anxiety (2016), 33: 1013–1022.

Kandala, Saverio Stranges. Socioeconomic gradients and mental health: implications for public health. The British Journal of Psychiatry (2015) bjp.bp.114.147280

Margolis, R. & Myrskylä, M. Parental Well-being Surrounding First Birth as a Determinant of Further Parity Progression. Demography (2015) 52: 1147.

Miech, Eaton, Brennan. Mental health disparities across education and sex: a prospective analysis examining how they persist over the life course. J. Gerontol. B-Pyschol, 60 (2005), pp. S93–S98

Appendix 1: Model

$i = 1, \ldots, N$ subjects where $N = 3201$
$k = 1, \ldots, K$ covariates where $K = 8$
$j = 1, \ldots, J$ categories for the outcome variable where $J = 3$

$y_i \sim \text{Multinomial}(N, \pi_{i1}\ \pi_{i2}, \pi_{i3})$
$\mu_i = \beta_1 x_1 + \ldots + \beta_K x_K$

$\text{logit}(\theta_{i1}) = \alpha_1 - \mu_i$
$\text{logit}(\theta_{i2}) = \alpha_2 - \mu_i$
$\pi_{i1} = \theta_{i1}$
$\pi_{i2} = \theta_{i2} - \theta_{i1}$
$\pi_{i3} = 1 - \pi_{i1} - \pi_{i2}$

$\alpha_1 \sim \text{Normal}(a,b)$ where $a = 1.17$, $b = 8.70$
$\alpha_2 \sim \alpha_1 + \text{Gamma}(c,d)$ where $c = 0.15$, $d = 0.15$

$\beta_1 \sim \text{Normal}(e,f)$ where
$\beta_2 \sim \text{Normal}(g,h)$ where
$\beta_3 \sim \text{Normal}(i,j)$ where
$\beta_4 \sim \text{Normal}(k,l)$ where
$\beta_5 \sim \text{Normal}(m,n)$ where
$\beta_6 \sim \text{Normal}(o,p)$ where
$\beta_7 \sim \text{Normal}(q,r)$ where
$\beta_8 \sim \text{Normal}(s,t)$ where

β_1 – Age (continuous variable)
β_2 – Gender: Female (indicator variable)
β_3 – Education: High school degree (indicator variable)
β_4 – Education: College degree (indicator variable)
β_5 – Marital status: Married or living with partner (indicator variable)
β_6 – Income to poverty ratio (continuous variable)
β_7 – Employment: Unemployed (indicator variable)
β_8 – Children in household under 5 (indicator variable)

π_{ij} is the probability that subject i has outcome j
θ_{ij} is the probability the subject i has an outcome in $1, \ldots, j$
α_j is the probability of a subject with 0 for all covariates being in a group in $i \ldots j$

Appendix 2: Figures and tables

Table 1. Prior Information for β Coefficients

Variable	Outcome in paper	Data Source	Information given in paper	Corresponding prior	Citation
Age	Low mental well being	2010 and 2011 England Health Survey (n=13,983)	95% CI for middle aged compared to young: (1.35,1.84)	No(-0.30,2.00)	Kandala 2015
Gender	Low mental well being	2010 and 2011 England Health Survey (n=13,983)	95% CI for OR: (0.68,1.58)	No(0.04,9.24)	Kandala 2015
High School	Depression	Adults in Baltimore Maryland in 1981 surviving to 1993 (n=1,171)	7.23% for degree vs 12.74% for no degree	No(-0.75,6.37)	Miech 2005
College*	Depression	Wave 1 of National Epidemiologic Survey on Alcohol and Related Conditions in America (n=34,653)	99% CI for OR: (0.91,1.48)	No(0.15,27.22)	Erickson 2016
Marital Status**	Low mental well being	2010 and 2011 England Health Survey (n=13,983)	95% CI for OR: (0.62,0.97)	No(-0.25,32.79)	Kandala 2015
Income	Low mental well being	2010 and 2011 England Health Survey (n=13,983)	95% CI for lowest compared to highest: (0.79,1.78)	No(-0.83,1.44)	Kandala 2015
Employment ***	Low mental well being	2010 and 2011 England Health Survey (n=13,983)	95% CI for OR: (2.21,3.78)	No(1.06,22.80)	Kandala 2015
Children	Well Being and Life Satisfaction	German Socio-Economic Panel Study 1984-2010 (n=2301)	NA	No(-0.51,3.42)	Margolis 2015

* 99% confidence interval treated as 95% confidence interval
** = paper also includes an indicator for divorced/separated rather than just single or married
*** = original paper includes separate categories for unemployed seeking work and unemployed not seeking work, not seeking work is used here as only 267 of 2375 unemployed participants in our data report seeking work.

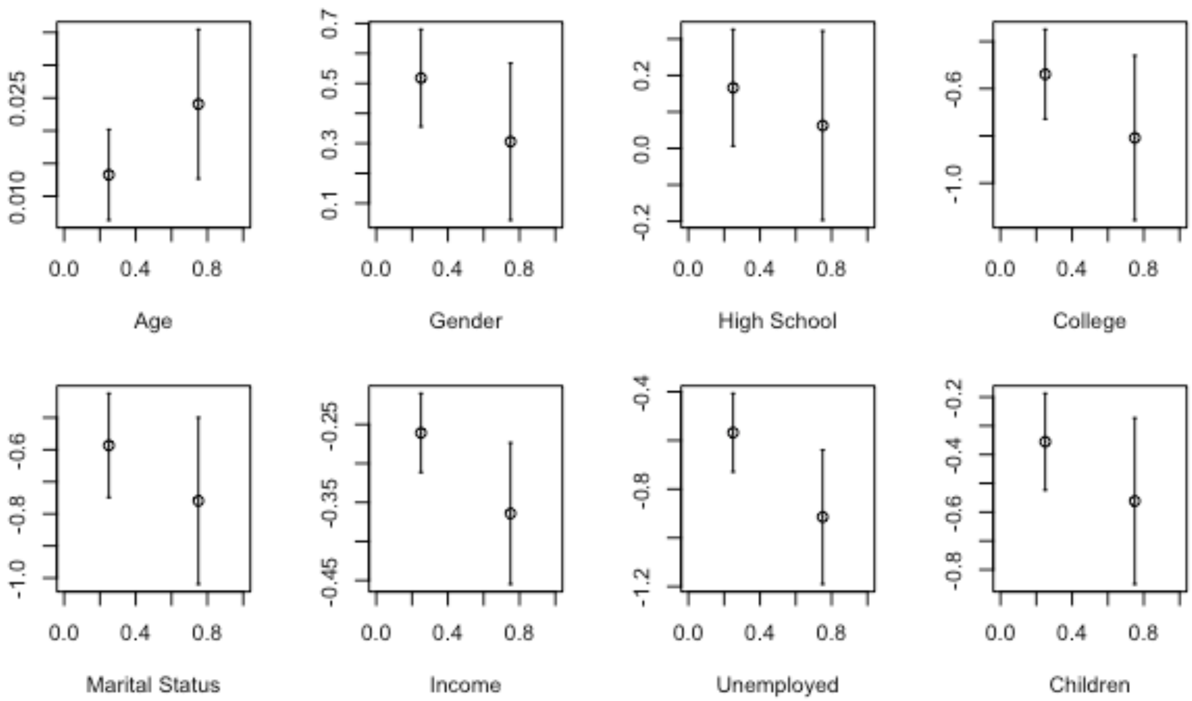

Figure 1. Assessment of proportional hazards assumption. Confidence intervals for coefficients of logistic regression models with outcome categorized as (Left) Y_1=1 if outcome level is 1 and 0 otherwise and (Right) Y_1=1 if outcome level is 1 or 2 and 0 otherwise

Table 2. Summary of posterior distributions for main parameters of interest.

	Mean (Logit)	SD (Logit)	95% CI Low (Logit)	95% CI High (Logit)	% Neg. (Logit)	% Pos. (Logit)	Mean (OR)	95% CI Low (OR)	95% CI High (OR)
Age**	0.007	0.003	0.000	0.014	0.018	0.982	1.007	1.000	1.014
Female*	0.409	0.082	0.247	0.563	0.000	1.000	1.505	1.28	1.756
High School	-0.036	0.096	-0.226	0.147	0.649	0.351	0.964	0.798	1.158
College	-0.064	0.114	-0.284	0.16	0.707	0.293	0.938	0.753	1.174
Married*	-0.303	0.08	-0.463	-0.146	1.000	0.000	0.738	0.63	0.864
Income*	-0.209	0.03	-0.267	-0.148	1.000	0.000	0.811	0.766	0.862
Unemployed*	0.777	0.083	0.613	0.935	0.000	1.000	2.176	1.846	2.547
Children*	-0.397	0.089	-0.572	-0.226	1.000	0.000	0.672	0.564	0.798
Alpha 1	1.028	0.177	0.689	1.372	0.000	1.000	2.795	1.992	3.944
Alpha 2	2.520	0.186	2.161	2.882	0.000	1.000	12.432	8.682	17.854

* Denotes significance at 95% level
** Lower bound for the 95% CI is rounded down actual value is greater than 0; significant at 95%

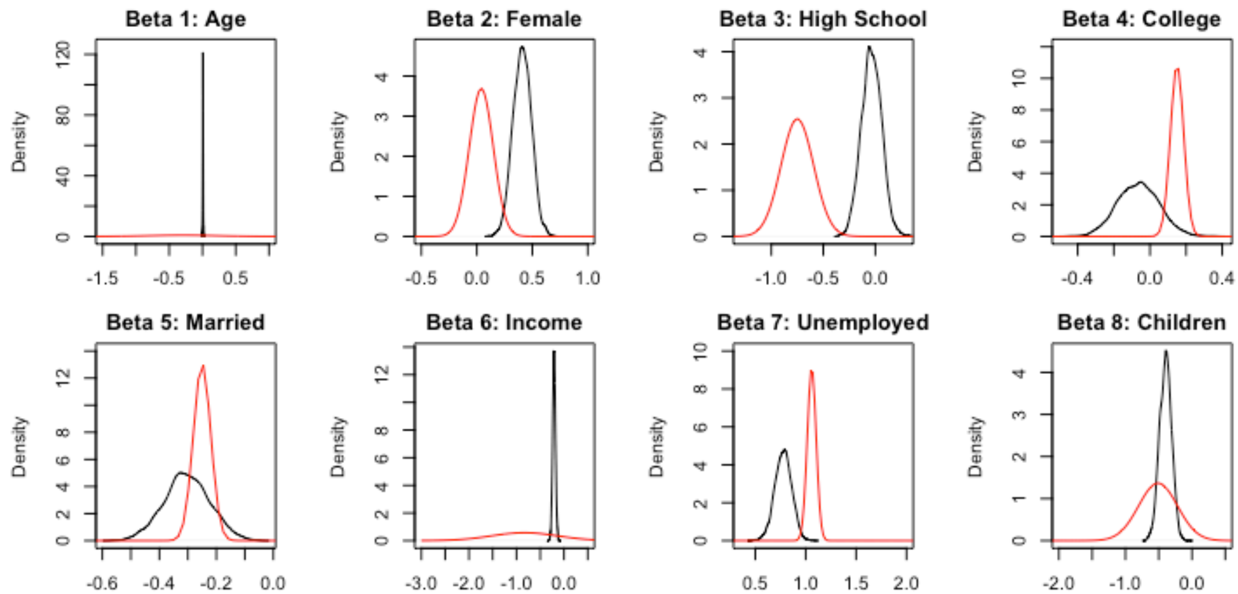

Figure 3. Prior and posterior distributions for β parameters on the logit scale. Red lines represent the prior distribution and black lines represent the posterior distribution.

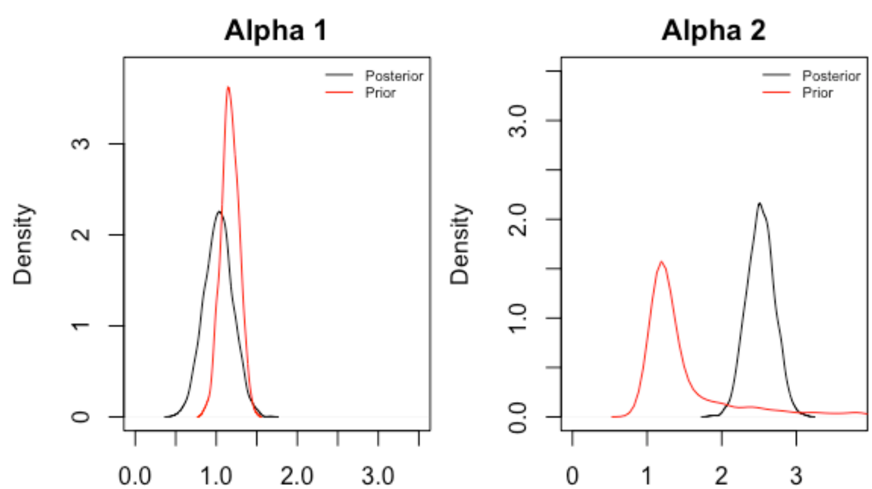

Figure 4. Prior and posterior distributions for α parameters on the logit scale.

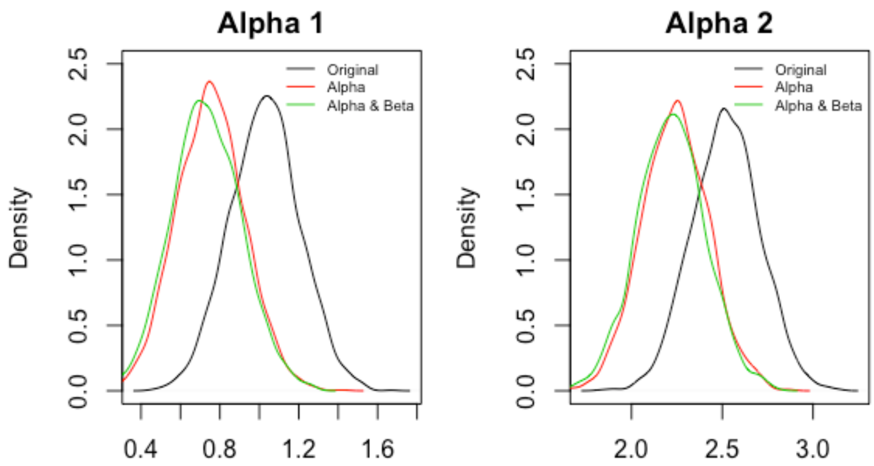

Figure 5. Posterior distributions for α parameters in sensitivity analysis.

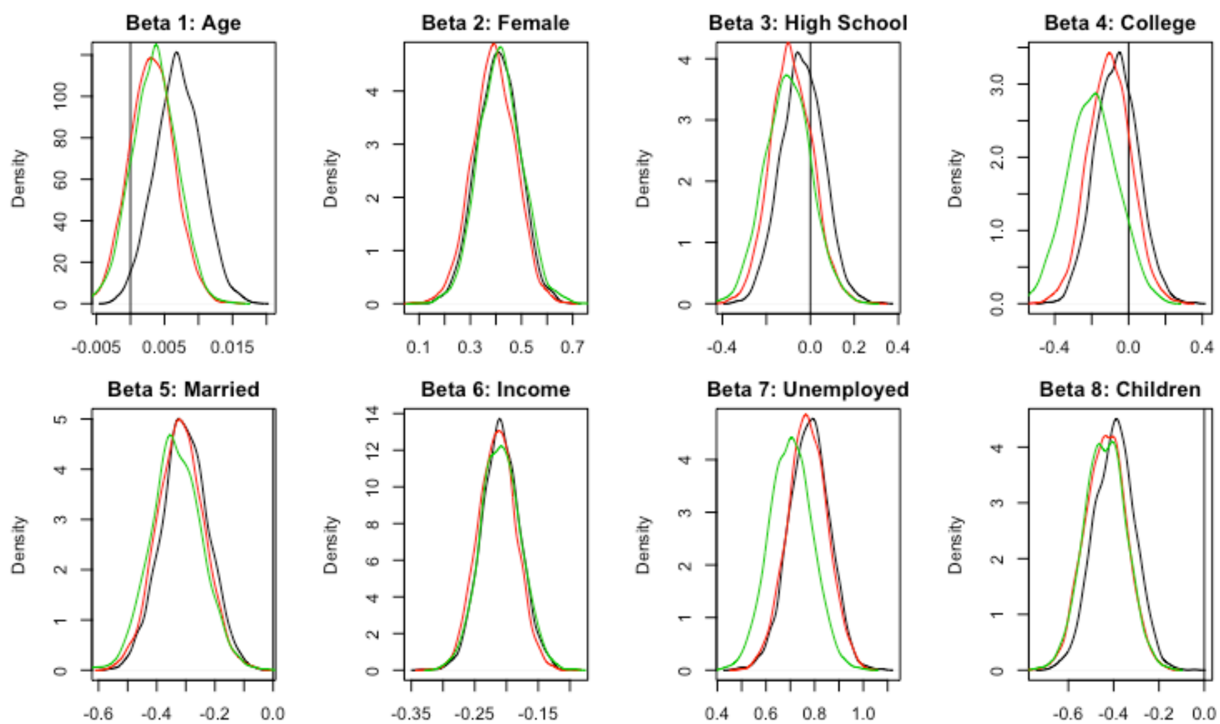

Figure 6. Posterior distributions for β parameters on the logit scale. The black line represents the original model. The red line represents the model for which α priors were relaxed. The green line represents the model for which α and β priors were relaxed.

Appendix 3: R Code for Original Model

```
### model
sink("model1.txt")
cat("
  model{

  for(i in 1:N){
    mu[i] <- beta[1]*x[i,1]+beta[2]*x[i,2]+beta[3]*x[i,3]+beta[4]*x[i,4]+beta[5]*x[i,5]+beta[6]*x[i,6]+
              beta[7]*x[i,7]+beta[8]*x[i,8]

    logit(theta[i,1]) <- alpha[1]-mu[i]
    logit(theta[i,2]) <- alpha[2]-mu[i]
    pie[i,1] <- theta[i,1]
    pie[i,2] <- theta[i,2] - theta[i,2-1]
    pie[i,3] <- 1 - theta[i,2]

    y[i] ~ dcat(pie[i,])
  }

  alpha0[1] ~ dnorm(a,b)
  a2 ~ dgamma(c,d)
  alpha0[2] <- alpha0[1] + a2
  alpha <- sort(alpha0)

  beta[1]~dnorm(e,f)
  beta[2]~dnorm(g,h)
  beta[3]~dnorm(i,j)
  beta[4]~dnorm(k,l)
  beta[5]~dnorm(m,n)
  beta[6]~dnorm(o,p)
  beta[7]~dnorm(q,r)
  beta[8]~dnorm(s,t)

  }
  ",fill = TRUE)
sink()

### data
dat<-read.csv("~/Documents/UCLA/CLASSES/234 bayes/project_final/data_jags")
x<-dat[1:3201,3:10]
y<-dat[1:3201,11]
a = 1.17; b = 8.70; c = 0.15; d = 0.15; e = -0.30; f = 2.00; g = 0.04; h = 9.24; i = -0.75; j = 6.37; k = 0.15; l = 27.22;
m = -0.25; n = 32.79; o = -0.83; p = 1.44; q = 1.06;  r = 22.80; s = -0.51; t = 3.42;
N=length(y)

### data, params, and inits
data1 <- list(x=x,y=y,a=a,b=b,c=c,d=d,e=e,f=f,g=g,h=h,i=i,j=j,k=k,l=l,m=m,n=n,o=o,p=p,q=q,r=r,s=s,t=t,N=N)
params1 <- c("alpha","beta","mu[1:10]","pie[1:10,1:3]")
inits1 <- rep(list(list(beta=rep(1,8))),3)
```

```
### run model
set.seed(242)
model1 <-
jags(model.file="model1.txt",parameters.to.save=params1,inits=inits1,data=data1,n.chains=3,n.iter=40000,n.burnin
=0,n.thin=1)
```