Bayesian Approach to Modeling Melanocytic Nevus Development in Colorado

Final Project

Ryan Summers, M.S. Candidate[1]

2024-12-01

Table of Contents

[1] Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO

Abstract

Malignant melanoma risk is primarily associated with melanocytic nevi, with the majority of these nevi forming during childhood. This study identified significant associations between nevi counts and specific demographic and phenotypic factors. Children with blonde or brown hair had higher median nevi (mole) counts compared to those with red hair. Gender differences were also observed, with girls having 13% lower median nevi count than boys. Additionally, non-Hispanic children exhibited a 30.7% higher median nevi count compared to Hispanic children. No other comparisons showed statistically significant differences, as their 95% HDI overlapped zero.

Introduction

The incidence of malignant melanoma has nearly tripled among white individuals in the United States from 1975 to 2004, mirroring trends in Australia. Racial disparities are pronounced, with non-Hispanic white individuals experiencing an annual incidence rate of 25.1 per 100,000, compared to just 1.0 for black individuals and 4.5 for Hispanic white individuals (Crane et al., 2009).

Melanocytic nevi represent the strongest risk factor for melanoma, typically developing during childhood and influenced by factors like lighter skin, hair color, blue or green eyes, and sun exposure. Due to the close parallels between nevi risk factors and melanoma risk, researchers are increasingly studying nevi as potential markers of underlying causes (Pettijohn et al. 2009).

Hypothesis

This study aims to examine trends in nevus development over time and their associations with melanoma risk factors within the context of a sun protection intervention. We hypothesize that demographic factors (e.g., gender, ethnicity, and phenotypic characteristics such as eye and hair color) and waterside vacations may interact with the intervention to influence nevus development. Specifically, we expect that children with lighter skin tones, light-colored eyes, or hair may exhibit higher nevus counts over the study period.

Materials and Methods

The study was conducted between 1998 and 2007, focusing on improving sun protection practices in children born in the Denver/Boulder metropolitan area. From 2,148 births, 728 families (61.9%) were enrolled, with 472 participants ultimately completing the study.

Researchers conducted skin examinations to assess nevus counts when children were 3 and 4 years old. Parents provided demographic information and details about sun exposure through surveys (Crane et al., 2012).

Analysis Plan

Prior to analysis, participants with incomplete nevi count data from 2004 to 2008 were excluded, ensuring only children with full longitudinal data were included to minimize potential bias from missing information, which resulted in a total of 321 participants available for analysis. We will employ a Bayesian generalized non-linear multiple multilevel model with a negative binomial distribution to analyze nevus count data. The approach incorporates fixed effects for demographic variables, uses random effects (intercept only) to account for variability in mean mole counts for each respondent, and accounts for key covariates including waterside vacations, demographic characteristics, and baseline skin color. The analysis sets a significance level of 0.05, with results considered significant if the 95% highest density level (HDI) does not include zero. Analyses will be conducted using R (v4.4.1) and the brms package to perform Bayesian modeling that can assess the effect of nevus counts across different demographic and phenotypic subgroups. Priors were carefully chosen to incorporate prior knowledge while remaining weakly informative to allow the data to drive the results. Specifically:

- **Fixed effects coefficients (β):** A normal prior with mean 0 and standard deviation 10 was used (N(0,10)) to allow for reasonable variation in the regression coefficients without overly constraining their values.
- **Random effects standard deviations (SD):** A Cauchy prior with location 0 and scale 2 (Cauchy(0, 2))was applied to the standard deviations of the random intercepts, reflecting a belief that most group-level variation is moderate but allowing for larger variability.
- **Shape parameter for the negative binomial distribution:** An inverse gamma prior with shape parameters 0.4 and 0.3 (InvGamma(0.4, 0.3)) was specified to reflect weak prior knowledge about the dispersion in the count data.

Results

Table 1: Summary of Nevi (Mole) Counts Stratified by Gender and Ethnicity Over Study Period (2004 to 2008)

	Gender		Hispanic		
	Female	**Male**	**No**	**Yes**	**All**
	(N=173)	(N=148)	(N=286)	(N=35)	(N=321)
2004					
Mean (SD)	18.9 (12.7)	22.8 (13.0)	21.5 (13.3)	14.1 (7.09)	20.7 (12.9)
Median [Min, Max]	16.0 [1.00, 74.0]	21.0 [2.00, 68.0]	19.0 [1.00, 74.0]	12.0 [4.00, 31.0]	18.0 [1.00, 74.0]
2005					
Mean (SD)	25.1 (15.7)	30.3 (18.0)	28.9 (17.2)	16.2 (8.39)	27.5 (17.0)
Median [Min, Max]	23.0 [2.00, 98.0]	26.0 [2.00, 99.0]	25.0 [2.00, 99.0]	14.0 [3.00, 38.0]	24.0 [2.00, 99.0]
2006					
Mean (SD)	28.7 (18.0)	33.6 (19.4)	32.1 (19.1)	21.5 (12.2)	31.0 (18.8)
Median [Min, Max]	26.0 [3.00, 99.0]	32.0 [5.00, 118]	29.0 [3.00, 118]	21.0 [5.00, 60.0]	28.0 [3.00, 118]
2007					
Mean (SD)	38.3 (25.5)	44.8 (26.2)	43.2 (26.4)	26.0 (15.5)	41.3 (26.0)
Median [Min, Max]	35.0 [4.00, 190]	41.0 [2.00, 137]	39.0 [2.00, 190]	25.0 [5.00, 71.0]	37.0 [2.00, 190]
2008					
Mean (SD)	48.4 (28.8)	54.3 (32.0)	53.6 (30.6)	31.2 (19.1)	51.1 (30.4)
Median [Min, Max]	45.0 [4.00, 170]	47.5 [6.00, 187]	48.0 [4.00, 187]	32.0 [9.00, 103]	45.0 [4.00, 187]

Nevus counts showed a positively right skewed distribution; thus, median counts are presented (***Table 1***). Results reveal significant associations between certain demographic characteristics, waterside vacations taken during the study, and nevi counts.

Table 2: Contrast Analysis

Outcome: Median Nevus Counts

Contrast	Est. Ratio[1]	95% HDI[2]
Hair Color		
BLK / BLN	0.764	(0.44, 1.16)
BLK / BRN	0.784	(0.46, 1.17)
BLK / RED	1.193	(0.66, 1.85)
BLN / BRN	1.022	(0.89, 1.17)
BLN / RED	1.555	(1.20, 1.96)
BRN / RED	1.519	(1.17, 1.92)
Eye Color		
(BLU/GRN/COMB) / HAZ	1.020	(0.87, 1.18)
(BLU/GRN/COMB) / (LBRN/DBRN)	1.087	(0.91, 1.27)
HAZ / (LBRN/DBRN)	1.065	(0.88, 1.26)
Hispanic		
No / Yes	1.307	(1.03, 1.61)
Gender		
Female / Male	0.870	(0.77, 0.98)

[1]Point estimate displayed: Ratio of medians. Results are back-transformed from the log scale

[2]HPD interval probability: 0.95 w/ Tukey Adjustment

Note: Results are averaged over the levels of: hispanic, eyecolor, haircolor, gender

After adjusting for demographic, phenotypic, and number of waterside vacations taken, notable differences in nevi counts were found. Compared to children with red hair, those with blonde hair were expected to have a 55.5% higher median nevi count (rate ratio = 1.555, 95% HDI: [1.20, 1.96]). Similarly, children with brown hair demonstrated a 51.9% higher median nevi count relative to those with red hair (rate ratio = 1.519, 95% HDI: [1.17, 1.92]).

Gender also showed a significant variation in nevi counts. Girls were expected to have a 13% lower median nevi count compared to boys (rate ratio = 0.870, 95% HDI: [0.77, 0.98]). Ethnicity emerged as another distinguishing factor, with non-Hispanic children expected to have a 30.7% higher median nevi count compared to Hispanic children (rate ratio = 1.307, 95% HDI: [1.03, 1.61]). (***Table 2***).

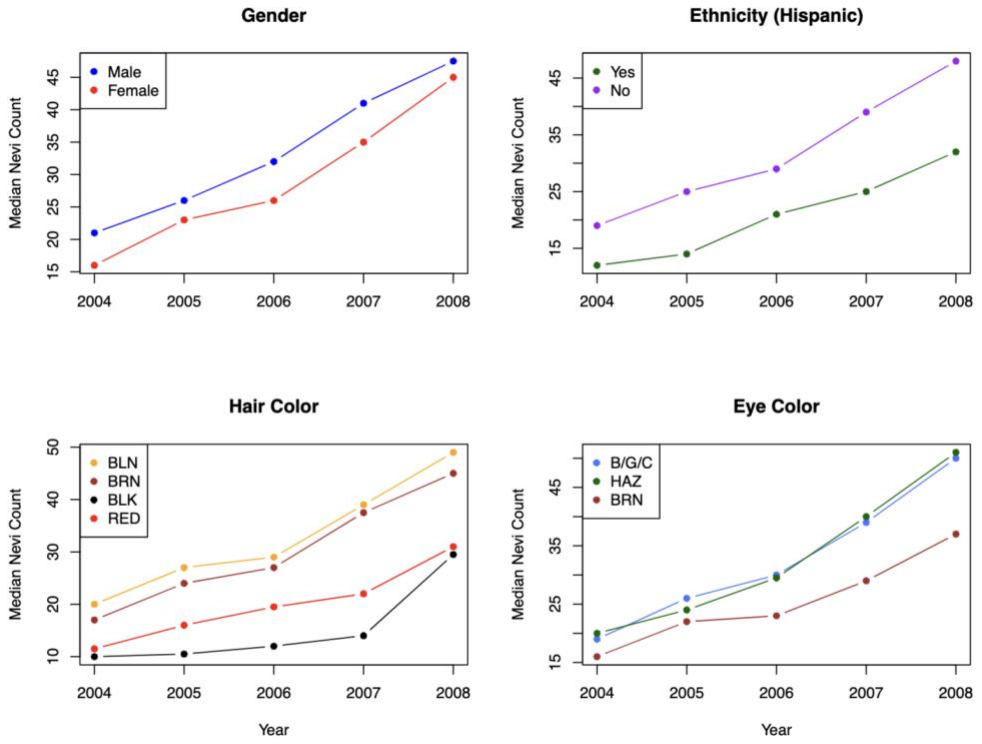

Figure 1: Nevi Count Trends (2004-2008)

Trend analysis revealed a significant increase in nevi counts year after year (***Figure 1***). For the children under study, for every additional year between 2004 and 2008, the expected nevi count increased by 25% on average ($exp(\hat{\beta}_{year}) = 1.25, 95\%\text{HDI: } [1.24, 1.26]$). Furthermore, it was noted that for every additional waterside vacation taken between 2005 and 2008, the expected nevi count increased by 10% on average ($exp(\hat{\beta}_{vacations})$=1.10, 95% HDI: [1.04, 1.17]) (***Supplementary Table 1***).

Lastly, a child's baseline skin color indicated that for every additional point on the color scale, there was a 6% decrease in expected nevi counts on average ($\exp(\hat{\beta}_{skin})$=0.94, 95% HDI: [0.90, 0.97]). Higher values of the continuous skin color scale indicate a darker skin color.

Conclusions

Our study provides insights into the demographic and environmental factors influencing mole development among children between 2004 and 2008. Hair color emerged as a significant predictor of mole count, with blonde and brown-haired children showing substantially higher mole counts compared to those with red hair. Gender differences were apparent, with girls demonstrating a 13% lower median mole count than boys.

A notable trend was the consistent year-over-year increase in mole counts, with a 25% average increase annually. The positive association with waterside vacations—a 10% increase in expected mole count per vacation—suggests potential environmental influences on mole development. Additionally, an inverse relationship was observed between skin color and mole counts, with a 6% decrease in expected mole count for each point increase on the skin color scale.

These findings highlight the need for increased sun protection awareness and the increased risk of developing skin-based cancers like melanoma. The observed variations in mole counts across different demographic groups underscore the importance of personalized skin health monitoring.

Limitations include the lack of explicit intervention and control group assignments and the narrow study timeframe. Model limitations were particularly evident in the inclusion of interaction terms. Convergence issues arose when attempting to model interactions between gender, hair, and eye color—areas of specific interest. Future work will explore different modeling techniques and run simulations testing various priors to achieve convergence.

References

Crane, Lori A., Nancy L. Asdigian, Anna E. Barón, Jenny Aalborg, Alfred C. Marcus, Stefan T. Mokrohisky, Tim E. Byers, Robert P. Dellavalle, and Joseph G. Morelli. 2012. "Mailed Intervention to Promote Sun Protection of Children: A Randomized Controlled Trial." *American Journal of Preventive Medicine* 43 (4): 183–202. https://doi.org/10.1016/j.amepre.2012.06.022.

Crane, Lori A., Stefan T. Mokrohisky, Robert P. Dellavalle, Nancy L. Asdigian, Jenny Aalborg, Tim E. Byers, Chan Zeng, Anna E. Barón, Joanna M. Burch, and Joseph G. Morelli. 2009. "Melanocytic Nevus Development in Colorado Children Born in 1998: A Longitudinal Study." *Archives of Dermatology* 145 (2): 148–56. https://doi.org/10.1001/archdermatol.2008.571.

Gabry, Jonah. n.d. "Graphical posterior predictive checks using the bayesplot package." https://cran.r-project.org/web/packages/bayesplot/vignettes/graphical-ppcs.

Pettijohn, Kelly J., Nancy L. Asdigian, Jenny Aalborg, Joseph G. Morelli, Stefan T. Mokrohisky, Robert P. Dellavalle, and Lori A. Crane. 2009. "Vacations to Waterside Locations Result in Nevus Development in Colorado Children." *Cancer Epidemiology, Biomarkers & Prevention* 18 (2): 454–63. https://doi.org/10.1158/1055-9965.EPI-08-0634.

Vehtari, Aki. n.d. "Prior Choice Recommendations." https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations.

(Crane et al. 2012) (Pettijohn et al. 2009) (Crane et al. 2009) (Vehtari, n.d.) (Gabry, n.d.)

Appendix A: Supplementary Tables

Supplementary Table 1: Bayesian Negative Binomial (NB) Regression Results

Trends in Nevus Development and Associations with Melanoma Risk Factors

Characteristic	exp(rate ratio) (95% CI)[1]
Year	1.25 (1.24 to 1.26)
Hispanic	
No	Ref
Yes	0.76 (0.61 to 0.95)
Gender	
Female	Ref
Male	1.15 (1.02 to 1.30)
Eye Color	
BLU/GRN/COMB	Ref
HAZ	0.98 (0.84 to 1.14)
L-BRN / D-BRN	0.92 (0.78 to 1.09)
Hair Color	
BLK	Ref
BLN	1.30 (0.81 to 2.07)
BRN	1.27 (0.80 to 2.01)
RED	0.84 (0.50 to 1.39)
Base Skin Color (Higher=Darker)	0.94 (0.90 to 0.97)
Vacations ('05-'08)	1.10 (1.04 to 1.17)

[1]CI = Confidence Interval

Table 3: Characteristics of Study Participants by Nevi Count

	Eye Color			Hair Color				
	BLU/GRN/COMB	HAZ	LBRN/DBRN	BLK	BLN	BRN	RED	All
	(N=147)	(N=86)	(N=88)	(N=6)	(N=161)	(N=132)	(N=22)	(N=321)
2004								
Mean (SD)	21.6 (13.8)	22.2 (13.1)	17.6 (10.7)	10.8 (5.95)	21.8 (12.4)	20.6 (13.5)	15.7 (13.0)	20.7 (12.9)
Median [Min, Max]	19.0 [1.00, 72.0]	20.0 [2.00, 67.0]	16.0 [3.00, 74.0]	10.0 [4.00, 22.0]	20.0 [1.00, 68.0]	17.0 [2.00, 74.0]	11.5 [2.00, 58.0]	18.0 [1.00, 74.0]
2005								
Mean (SD)	28.9 (16.6)	28.7 (18.7)	24.1 (15.5)	15.0 (10.6)	29.5 (17.0)	27.3 (17.3)	17.7 (11.0)	27.5 (17.0)
Median [Min, Max]	26.0 [2.00, 98.0]	24.0 [4.00, 99.0]	22.0 [2.00, 80.0]	10.5 [6.00, 32.0]	27.0 [2.00, 99.0]	24.0 [2.00, 98.0]	16.0 [2.00, 48.0]	24.0 [2.00, 99.0]
2006								
Mean (SD)	33.4 (19.8)	31.7 (19.0)	26.1 (15.8)	18.2 (13.4)	32.7 (18.3)	31.1 (19.8)	20.7 (11.9)	31.0 (18.8)
Median [Min, Max]	30.0 [3.00, 118]	29.5 [5.00, 99.0]	23.0 [5.00, 88.0]	12.0 [6.00, 36.0]	29.0 [3.00, 118]	27.0 [5.00, 99.0]	19.5 [5.00, 45.0]	28.0 [3.00, 118]
2007								
Mean (SD)	44.3 (27.7)	42.7 (25.5)	34.7 (22.4)	19.0 (13.3)	43.5 (24.8)	42.0 (27.7)	27.3 (20.6)	41.3 (26.0)
Median [Min, Max]	39.0 [2.00, 190]	40.0 [4.00, 126]	29.0 [5.00, 114]	14.0 [8.00, 44.0]	39.0 [8.00, 137]	37.5 [4.00, 190]	22.0 [2.00, 72.0]	37.0 [2.00, 190]
2008								
Mean (SD)	55.1 (31.4)	54.2 (31.1)	41.5 (25.8)	28.3 (14.6)	54.8 (30.6)	50.1 (30.7)	36.6 (22.3)	51.1 (30.4)
Median [Min, Max]	50.0 [6.00, 187]	51.0 [4.00, 174]	37.0 [9.00, 126]	29.5 [12.0, 49.0]	49.0 [8.00, 187]	45.0 [4.00, 170]	31.0 [6.00, 82.0]	45.0 [4.00, 187]

Eye Color

- BLU: Blue

- GRN: Green
- COMB: Combination of Blue and Green
- HAZ: Hazel
- LBRN: Light Brown
- DBRN: Dark Brown

Hair Color

- BLK: Black
- BLN: Blonde
- BRN: Brown
- RED: Red

Appendix B: Code

```
# ---
# output:
#   bookdown::pdf_document2:
#     toc: false
#     number_sections: false
# documentclass: article
# geometry: margin=0.25in
# classoption:
# - twocolumn
# bibliography: summers.bib
# ---
knitr::opts_chunk$set(number_sections = FALSE)
knitr::opts_chunk$set(fig.align='center')

library(knitr)
library(ggplot2)
library(tidyverse)
library(magrittr)
library(ggfortify)
library(olsrr)
library(kableExtra)
library(doBy)
library(psych)
library(MASS)
library(dplyr)
library(gmodels)
library(car)
library(broom)
library(tictoc)
library(parallel)
library(readxl)
library(lme4)
library(quantreg)
library(brms)
library(loo)
library(bayesplot)
library(table1)
library(gtsummary)
library(gt)

opts_chunk$set(tidy = F)
moles_df <- read_csv('Mole Development/Mole Count Data 2004-2008.csv')
# NAs per variable
round(colSums(is.na(moles_df)), 1) # counts
round(colSums(is.na(moles_df)) / nrow(moles_df) * 100, 1) # percentages

# number of complete cases (mole counts for each year)
nrow(moles_df[complete.cases(moles_df),])
```

```r
# percentage of complete cases to full dataset
sum(complete.cases(moles_df)) / nrow(moles_df)
# format columns
moles_df2 <- moles_df %>%
  # lowercase columns & rows
  setNames(tolower(names(.))) %>%
  mutate_if(is.character, tolower) %>%
  # remove white-space in columns
  setNames(gsub("\\s+", "_", names(.)))
# calculate number of waterside vacations between 2005 and 2007 (inclusive)
moles_df2 <- moles_df2 %>%
  mutate(num_vacs_bt_2005_2008 = number_vacs_birth_thru_2007-number_vacs_birt
h_thru_2005)

# change factor variable names
moles_df2 <- moles_df2 %>%
  mutate(
    gender = case_when(
      gender == 1 ~ "Female",
      gender == 2 ~ "Male"
    ),
    hispanic = case_when(
      hispanic == 0 ~ "No",
      hispanic == 1 ~ "Yes"
    ),
    eyecolor = case_when(
      eyecolor == 1 ~ "BLU/GRN/COMB",
      eyecolor == 2 ~ "LBRN/DBRN",
      eyecolor == 3 ~ "HAZ"
    ),
    haircolor = case_when(
      haircolor == 1 ~ "BLN",
      haircolor == 2 ~ "RED",
      haircolor == 3 ~ "BRN",
      haircolor == 4 ~ "BLK"
    )
  )
moles_df2 <- moles_df2 %>%
  filter(complete.cases(.)) %>%
  mutate(across(c(gender, hispanic, eyecolor, haircolor), as.factor))
label(moles_df2$gender) <- "Gender"
label(moles_df2$hispanic) <- "Hispanic"
label(moles_df2$eyecolor) <- "Eye Color"
label(moles_df2$haircolor) <- "Hair Color"
label(moles_df2$num_vacs_bt_2005_2008) <- "# of Vacations (between 2005 and 2
008)"
label(moles_df2$molecount2004) <- "Mole Count 2004"
label(moles_df2$molecount2005) <- "Mole Count 2005"
label(moles_df2$molecount2006) <- "Mole Count 2006"
```

```r
label(moles_df2$molecount2007) <- "Mole Count 2007"
label(moles_df2$molecount2008) <- "Mole Count 2008"

moles_df2 %>%
  summarise(totals = sum(molecount2004, molecount2005, molecount2006,
                         molecount2007, molecount2008))
moles_long <- moles_df2 %>%
  filter(complete.cases(.)) %>%
  pivot_longer(cols = c(molecount2004, molecount2005, molecount2006,
                        molecount2007, molecount2008),
               names_to = "year", values_to = "molecount") %>%
  mutate(year = ifelse(year == "molecount2004", 2004,
                       ifelse(year=='molecount2005', 2005,
                       ifelse(year=='molecount2006', 2006,
                       ifelse(year=='molecount2007', 2007,
                       ifelse(year=='molecount2008', 2008, NA)))))) %>%
  # lowercase columns & rows
  setNames(tolower(names(.))) %>%
  mutate_if(is.character, tolower) %>%
  # remove white-space in columns
  setNames(gsub("\\s+", "_", names(.)))
moles_strata1 <- c(
    split(moles_df2, ~gender),
    split(moles_df2, ~hispanic),
    list("All" = moles_df2)
  )

moles_strata2 <- c(
    split(moles_df2, ~eyecolor),
    split(moles_df2, ~haircolor),
    list("All" = moles_df2)
  )
moles_labels1 <- list(
    variables = list(
      molecount2004 = "2004", # names denote variables, values supply labels
      molecount2005 = "2005",
      molecount2006 = "2006",
      molecount2007 = "2007",
      molecount2008 = "2008"
    ),
    groups = list("Gender", "Hispanic") # this is a list of labels only
  )

moles_labels2 <- list(
    variables = list(
      molecount2004 = "2004",
      molecount2005 = "2005",
      molecount2006 = "2006",
      molecount2007 = "2007",
      molecount2008 = "2008"
```

```
    ),
    groups = list("Eye Color", "Hair Color", "")
  )

moles_groups1 <- c(2,2) # gender(2), hispanic(2),....
moles_groups2 <- c(3,4,1)

tb1.1 <- table1(moles_strata1, moles_labels1, groupspan = moles_groups1)
tb1.2 <- table1(moles_strata2, moles_labels2, groupspan = moles_groups2)
kable(tb1.1, "latex", booktabs = T, caption="Characteristics of Study Partici
pants by Mole Count") %>%
  add_header_above(c(" " = 1, "Gender" = 2, "Hispanic" = 2," " = 1)) %>%
  kable_styling(latex_options = c("scale_down", "HOLD_position")) %>%
  row_spec(0, bold = TRUE, color = "black") %>%
  column_spec(1, bold = TRUE)
# fit a Bayesian Poisson/Neg.Binomial model
bayes.model1 <- brm(molecount ~ year + hispanic + gender + eyecolor + haircol
or +
        baseskincolor + num_vacs_bt_2005_2008 + (1|respondent_code_number),
      data = moles_long,
      family = negbinomial(link = "log"), # NB to address over-dispersion
      prior = c(
        prior(normal(0, 10), class = "b"),  # prior for fixed effects
        prior(cauchy(0, 2), class = "sd"),   # prior for random effects
        prior(inv_gamma(0.4, 0.3), class = "shape") # prior for shape paramet
er
        ),
      iter=4000, chains=4, cores=8,
      save_pars = save_pars(all = TRUE),
      seed=42  # reproducibility
      )
summary(bayes.model1)
# Set up the plotting grid (2 rows, 2 columns)
par(mfrow = c(2, 2))

# Gender: Male vs. Female
plot_data <- aggregate(molecount ~ year + gender, moles_long, median)

# Plot the first group (e.g., Male)
plot(plot_data$year[plot_data$gender == "Male"],
     plot_data$molecount[plot_data$gender == "Male"],
     type = "b", col = "blue", pch = 16,
     main = "Gender",
     xlab = "", ylab = "Median Nevi Count", xlim = range(plot_data$year),
     ylim = range(plot_data$molecount))


# Overlay the second group (e.g., Female)
lines(plot_data$year[plot_data$gender == "Female"],
```

```
      plot_data$molecount[plot_data$gender == "Female"],
      type = "b", col = "red", pch = 16)
legend("topleft", legend = c("Male", "Female"),
       col = c("blue", "red"), pch = 16)

# Hispanic: Yes vs. No
plot_data <- aggregate(molecount ~ year + hispanic, moles_long, median)

# plot Hispanic
plot(plot_data$year[plot_data$hispanic == "Yes"],
     plot_data$molecount[plot_data$hispanic == "Yes"],
     type = "b", col = "darkgreen", pch = 16,
     main = "Ethnicity (Hispanic)",
     xlab = "", ylab = "Median Nevi Count", xlim = range(plot_data$year),
     ylim = range(plot_data$molecount))

# overlay the other groups
lines(plot_data$year[plot_data$hispanic == "No"],
      plot_data$molecount[plot_data$hispanic == "No"],
      type = "b", col = "purple", pch = 16)
legend("topleft", legend = c("Yes", "No"),
       col = c("darkgreen", "purple"), pch = 16)

# hair color
plot_data <- aggregate(molecount ~ year + haircolor, moles_long, median)

# plot hair color group
plot(plot_data$year[plot_data$haircolor == "BLN"],
     plot_data$molecount[plot_data$haircolor == "BLN"],
     type = "b", col = "orange", pch = 16,
     main = "Hair Color",
     xlab = "Year", ylab = "Median Nevi Count", xlim = range(plot_data$year),
     ylim = range(plot_data$molecount))

# overlay the other groups
lines(plot_data$year[plot_data$haircolor == "BRN"],
      plot_data$molecount[plot_data$haircolor == "BRN"],
      type = "b", col = "brown", pch = 16)
lines(plot_data$year[plot_data$haircolor == "BLK"],
      plot_data$molecount[plot_data$haircolor == "BLK"],
      type = "b", col = "black", pch = 16)
lines(plot_data$year[plot_data$haircolor == "RED"],
      plot_data$molecount[plot_data$haircolor == "RED"],
      type = "b", col = "red", pch = 16)
legend("topleft", legend = c("BLN", "BRN", "BLK", "RED"),
       col = c("orange", "brown", "black", "red"), pch = 16)

# eye color
plot_data <- aggregate(molecount ~ year + eyecolor, moles_long, median)
```

```r
# plot eye color groups
plot(plot_data$year[plot_data$eyecolor == "BLU/GRN/COMB"],
     plot_data$molecount[plot_data$eyecolor == "BLU/GRN/COMB"],
     type = "b", col = "royalblue1", pch = 16,
     main = "Eye Color",
     xlab = "Year", ylab = "Median Nevi Count", xlim = range(plot_data$year),
     ylim = range(plot_data$molecount))

# overaly the other groups
lines(plot_data$year[plot_data$eyecolor == "HAZ"],
      plot_data$molecount[plot_data$eyecolor == "HAZ"],
      type = "b", col = "darkgreen", pch = 16)
lines(plot_data$year[plot_data$eyecolor == "LBRN/DBRN"],
      plot_data$molecount[plot_data$eyecolor == "LBRN/DBRN"],
      type = "b", col = "brown", pch = 16)
legend("topleft", legend = c("B/G/C", "HAZ", "BRN"),
       col = c("royalblue1", "darkgreen", "brown"), pch = 16)

# reset plotting area
par(mfrow = c(1, 1))

library(gtsummary)
library(gt)
my_theme <-list("tbl_regression-str:ref_row_text" = "Ref")
set_gtsummary_theme(my_theme)
theme_gtsummary_journal("jama")


tbl_regression(bayes.model1, exponentiate=T,
               label = list(
                 year~"Year",
                 baseskincolor~"Base Skin Color (Higher=Darker)",
                 num_vacs_bt_2005_2008~"Vacations ('05-'08)")) %>%
  bold_labels() %>%
  italicize_levels() %>%
  as_gt() %>%
  tab_options(
    #table.width='65%',
     table.font.size = px(12)) %>%
  tab_header(
    title = md('Supplementary Table 1: Bayesian Negative Binomial (NB) Regres
sion Results'),
    subtitle=md('Trends in Nevus Development and Associations with Melanoma R
isk Factors'))

# pairwise (hypothesis) testing
library(emmeans)
```

```r
# unction to compute emmeans, pairs, and convert to a data frame
compute_pairs <- function(model, group_var) {
  emms <- emmeans(model, as.formula(paste("~", group_var)), adjust='Tukey')
  pairs <- pairs(emms, type = "response")
  as.data.frame(pairs)
}

# variables of interest
group_vars <- c("gender", "hispanic", "eyecolor", "haircolor")

# apply fucnction
emm_pairs <- do.call(rbind, lapply(group_vars, compute_pairs, model = bayes.m
odel1))

# create a table for comparisons
emm_pairs %>%
  gt() %>%
  tab_row_group(
    label = "Gender",
    rows = contrast %in% c("Female / Male")
  ) %>%
  tab_row_group(
    label = "Hispanic",
    rows = contrast %in% c("No / Yes")
  ) %>%
  tab_row_group(
    label = "Eye Color",
    rows = contrast %in% c("(BLU/GRN/COMB) / HAZ", "(BLU/GRN/COMB) / (LBRN/DB
RN)",
                           "HAZ / (LBRN/DBRN)")
  ) %>%
  tab_row_group(
    label = "Hair Color",
    rows = contrast %in% c("BLK / BLN", "BLK / BRN", "BLK / RED", "BLN / BRN"
, "BLN / RED",
                           "BRN / RED")
  ) %>%
  tab_style(
    style = list(
      cell_fill(color = "gray95"),
      cell_text(weight = "bold")),
    locations = list(
      cells_row_groups())) %>%

  tab_style(
    style=cell_fill(color="gray95"),
    locations=cells_title()
  ) %>%
```

```
  tab_style(
    style = cell_text(weight = "bold"),
    locations = cells_column_labels()) %>%

  cols_merge(
    columns=c(lower.HPD, upper.HPD),
    pattern="({1}, {2})"
  ) %>%

  # format rounding output
  fmt_number(
    columns = c(contrast, ratio),
    decimals = 3) %>%
  fmt_number(columns=c(lower.HPD, upper.HPD),
             decimals=2) %>%

    # modify column names
  cols_label(
    contrast ~ md('Contrast'),
    ratio ~ md("Est. Ratio"),
    lower.HPD ~ "95% HDI",
    upper.HPD ~ "UL.HPD"
  ) %>%

  cols_width(
    upper.HPD ~ px(150)) %>%

  # add a source note
  tab_source_note(
    source_note = md(c(
    '_Note:_ <br>Results are averaged over the levels of: hispanic, eyecolor,
haircolor, gender'))) %>%

  # add a header and subtitle
  tab_header(
    title = md('Supplementary Table 2: Contrast Analysis'),
    subtitle = md("Outcome: Median Mole Counts")) %>%

  tab_options(
    #table.width = '65%',
    table.font.size = px(12)) %>%

  # add a footnote (if needed)
  tab_footnote(
    footnote = md("Point estimate displayed: ratio of medians
Results are back-transformed from the log scale"),
locations=cells_column_labels(2)) %>%
  tab_footnote(
    footnote = md("HPD interval probability: 0.95 w/ Tukey Adjustment"),
```

```
    locations = cells_column_labels(c(3,4)))
kable(tb1.2, "latex", booktabs = T, caption="Characteristics of Study Partici
pants by Nevi Count") %>%
  add_header_above(c(" " = 1, "Eye Color" = 3, "Hair Color" = 4," " = 1),
                   bold=T) %>%
  kable_styling(latex_options = c("scale_down")) %>%
  row_spec(0, bold = TRUE, color = "black") %>%
  column_spec(1, bold = TRUE)

# exponentiate and summarize the posterior
summary(bayes.model1)$fixed %>%
  as.data.frame() %>%
  mutate(
    Exp_Beta = round(exp(Estimate),3),
    Exp_LL = round(exp(`l-95% CI`),3),
    Exp_UL = round(exp(`u-95% CI`),3)) %>%
  dplyr::select(Exp_Beta, Exp_LL, Exp_UL)
# cross-validation metric for predictive performance
(loo1 <- loo(bayes.model1, save_psis = TRUE)) #moment_match = T:too comp. exp
.
plot(loo1)

# marginal posterior predictive check
yrep <- posterior_predict(bayes.model1)
ppc_loo_pit_qq(
  y = moles_long$molecount,
  yrep = yrep,
  lw = weights(loo1$psis_object))


### Diagnostics code

# Convert model to an array of samples
posterior_samples <- as.array(bayes.model1)

# Trace plots for each parameter
mcmc_trace(posterior_samples,
                          pars = c("b_year", "b_genderMale", "b_hispanicYes",
                                   "b_num_vacs_bt_2005_2008", "b_baseskincolo
r"))

# Density plots for each parameter
mcmc_areas(posterior_samples,
                            pars = c("b_year", "b_genderMale", "b_hispanicYes
",
                                   "b_num_vacs_bt_2005_2008", "b_baseskincolo
r"))
```

```r
# Get the summary from brms that includes R-hat and ESS
model_summary <- summary(bayes.model1)
print(model_summary$fixed)

# Extracting ESS data
ess_data <- data.frame(
  Parameter = rownames(model_summary$fixed),
  ESS = model_summary$fixed[, "Est.Error"])

# Plotting using ggplot2
ggplot(ess_data, aes(x = Parameter, y = ESS)) +
  geom_bar(stat = "identity") +
  labs(title = "Effective Sample Size for Each Parameter",
       x = "Parameter",
       y = "ESS")
# posterior predictive check
  # compaires the observed outcome variable (molecounts) to simulated dataset
s
  # (molecounts^rep) from the posterior predictive distribution
pp_check(bayes.model1)
# model estimates of dispersion
dispersion <- function(x) {var(x)/mean(x)}
ppc_stat(y = moles_long$molecount,
         yrep = posterior_predict(bayes.model1, draws = 1000), stat='dispersi
on')
lagsar <- as.matrix(bayes.model1, pars = "b_hispanicYes")
estimates <- quantile(lagsar, probs = c(0.25, 0.5, 0.75))
mcmc_hist(lagsar) +
  vline_at(estimates, linetype = 2, size = 1) +
  ggtitle("posterior median and 50% central interval")
sessionInfo()

sessionInfo()
```