Three (Groups of) Blind Mice. Familial Clusters of Cataract Development in Irradiated Mice

Alyssa Allsop and Amira Burns
Department of Statistics, Colorado State University

June 23, 2022

Abstract

Astronauts traveling outside Earth's protective magnetic field are exposed to radiation with potentially serious side effects - one such effect is development of cataracts. Studying cataracts caused by radiation is complicated by genetic predisposition or resistance to cataract development. This study exposed mice bred in genetic families to specific types of radiation to examine familial clustering of cataracts. A mixed effects logistic regression model was fit to the data; the final model included parameters for sex, treatment group, sex by treatment group interaction, and a random intercept for family. Family was shown to be an important source of variation in predicting cataract development. Male sex was associated with differences in cataract development among treatment groups, while evidence showed little association between female sex and cataract development across among treatment groups. Males were more likely to develop severe cataracts than females across all three treatment groups.

Keywords: Hierarchical modeling, Logistic regression, Bayesian analysis

1 Introduction

Little is known about the effects of high atomic number and energy (HZE) radiation, a main component of space radiation to which exposure is unavoidable beyond Earth's magnetic field. In contrast, extensive research has shown that exposure to high doses of gamma radiation leads to acute radiation sickness. Adverse effects from radiation may also be attributable to other factors, including genetics; several different cancers have been observed to cluster in mice families (Chernyavskiy et al. 2017). This analysis examines potential genetic susceptibility to cataract development in the presence of radiation treatment.
The experiment included 1820 unique mice from 48 unique families, bred over several generations to create a genetically heterogeneous sample. Individuals were randomly assigned, with equal family weight, to one of three radiation treatment groups: HZE irradiation, gamma irradiation, and non-irradiated control. Mice in both irradiated groups were exposed to radiation at 7-12 weeks of age, and all mice were monitored weekly until 800 days of age. This analysis used a simplified version of the data that contained final observations for the 1169 mice from 47 unique families that survived at least 552 days. Family size ranged $[11, 48]$, with median family size $= 24$.

Due to small group sizes (Table 1), the response was converted from a discrete ordinal variable (levels $= [1, 2, 3, 4]$) matching ocular changes in the eye to a binary variable where score ≥ 2 indicated presence of cataracts. Both the main experimental factor, Treatment, and the single random effect, genetic Family, were central to the experimental design and of primary research interest; consequently, both variables were included in all models considered. Additional covariates under consideration were sex, coat color, weight in grams, body condition score (BCS), age in days, and presence of three cancers: myeloid leukemia, harderian tumors, and pre-T lymphoma.

	Cataract Score				
Treatment	1	2	3	4	Total
Unirradiated	438	45	11	2	496
Gamma	214	53	6	4	277
HZE	281	107	6	2	396

Table 1: Counts of score by treatment group

2 Summary Statistics

Exploratory data analysis focused on assessing cataracts status in terms of treatment group, and potential associations between covariates and the response. Tables and graphs (available in Appendix A.1) were created to investigate any empirical associations between coviarates and either treatment group or cataract status. Further examination of potential covariates revealed distinct patterns of cataracts status between male and female mice, as shown in Figure 1.

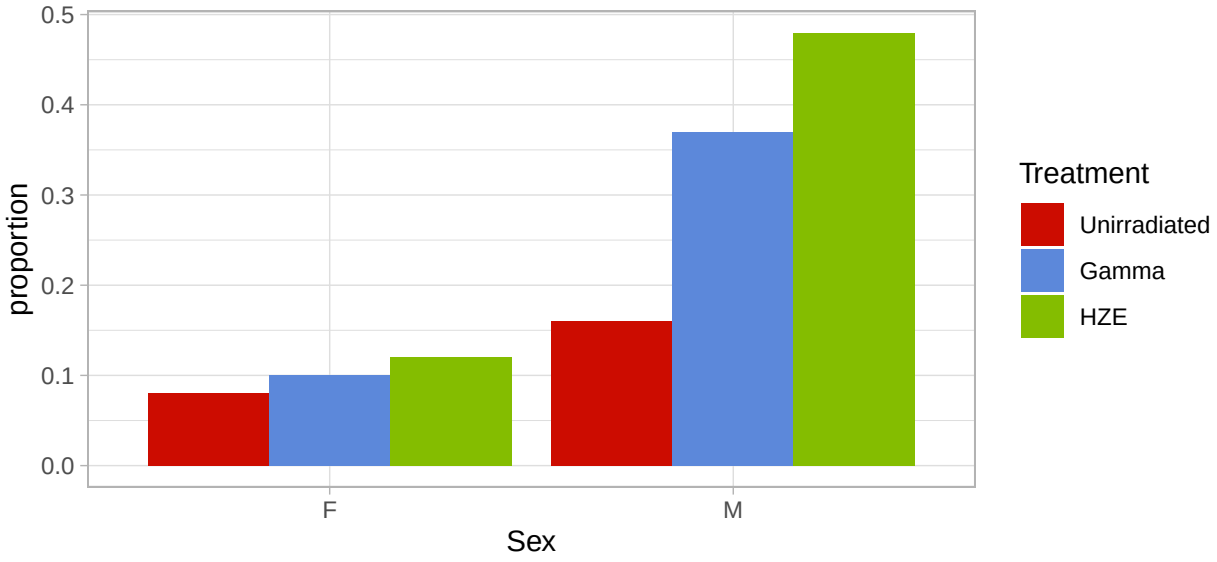

Figure 1: Sample proportions with cataracts by sex, treatment group

Empirical assessment of the effect of family across treatment group did not display apparent clustering where scores were higher across treatment groups in some families. Figure 2 did reveal some distinct patterns across families and treatment groups; for example, the Gamma treatment group appeared to have either the highest group score per family, or the lowest group score per family.

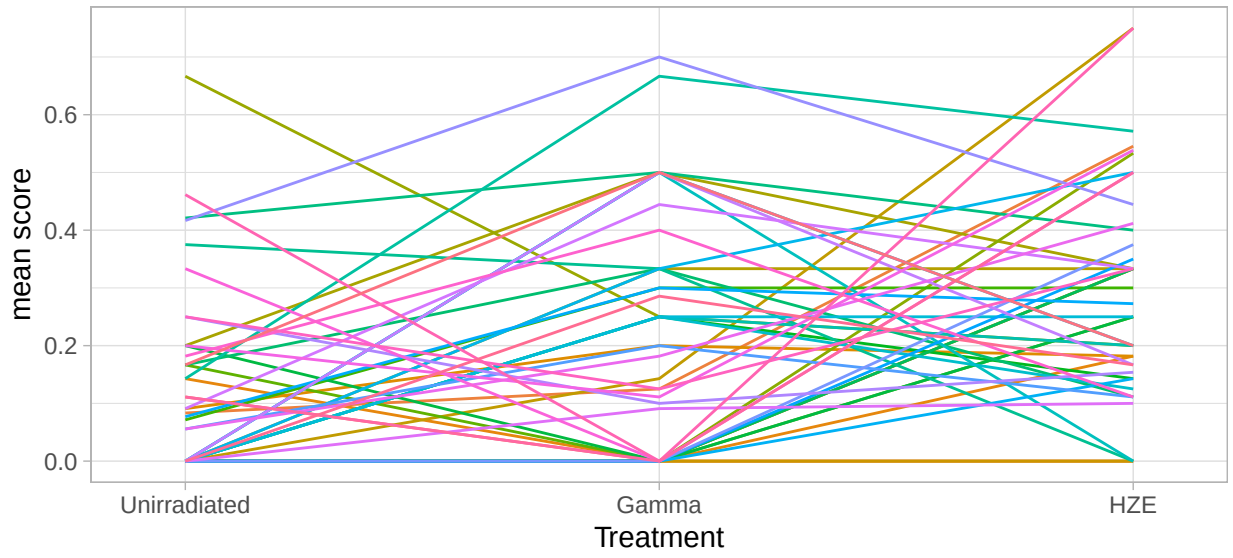

Figure 2: Mean family cataract score by treatment group

3 Statistical Methods

Hierarchical logistic regression was selected as the primary analytic method, due to the binary response and random effect of interest to the research question. Various models were fit and assessed using Aikike's Information Criterion (AIC); see Appendix A.2 for the full list of models examined. The final model can be written as:

$$\begin{aligned} Y_{ij} &\sim Bernoulli(p) \\ log(\frac{p}{1-p}) &= \beta_0 \ + \beta_1 * Gamma_i \ + \beta_2 * HZE_i \ + \beta_3 * M_i \ + \\ &\quad \beta_4 * Gamma_i * M_i \ + \beta_5 * HZE_i * M_i \ + \ [v_j + \epsilon_{ij}] \\ &\quad i = 1, ..., 1169 \ \text{ mice} \quad j = 1, ..., 47 \quad \text{families} \end{aligned} \tag{1}$$

where M_i is an indicator for males, v_j is the variance of the random intercept for Family j, and ϵ_{ij} is the residual variance. Model assumptions of Normally-distributed random effects and lack of overdispersion were verified in Appendix A.3.

A complementary Bayesian model with non-informative priors was fit to obtain probability distributions of the parameters from the final model, as well as test the estimates'

robustness to alternative approaches (Gelman et al. 2014). This model can be written as:

$$
\begin{aligned}
Y_{ij} \sim & Bernoulli(p) \\
log(\frac{p}{1-p}) = & \beta_0 * Control_i * F_i \ + \beta_1 * Gamma_i * F_i \ + \beta_2 * HZE_i * F_i + \\
& \beta_3 * Control_i * M_i \ + \beta_4 * Gamma * M \ + \beta_5 * HZE * M + \\
& v_j + \epsilon_{ij} \\
& i = 1, ..., 1169 \ \text{ mice} \quad j = 1, ..., 47 \quad \text{families, and} \\
v_j \sim & \ N(0, \tau) \\
& \text{with non-informative priors:} \\
\beta_0 \sim & \ N(0, 0.001) \quad \beta_1 \sim N(0, 0.001) \quad \beta_2 \sim N(0, 0.001) \\
\beta_3 \sim & \ N(0, 0.001) \quad \beta_4 \sim N(0, 0.001) \quad \beta_5 \sim N(0, 0.001) \\
v_i \sim & \ N(0, \sigma^2) \\
\tau \sim & \ Gamma(0.001, 0.001) \text{ where } \sigma^2 = 1/\tau
\end{aligned}
\tag{2}
$$

Model (2) was run through a Gibbs sampler Markov Chain Monte Carlo (MCMC) algorithm in the 'rjags' package (Plummer 2022) to approximate the posterior distributions of the estimates of all fixed effects and the variance of the random effect. 3 chains of 60000 total iterations with a 10000-iteration burn-in period; starting values were obtained from the estimates generated by Model (1), with noise added to avoid false convergence. Full model diagnostics may be accessed in Appendix A.4.

	GLMM Est	MCMC Mode	MCMC SD	HPD Lower	HPD Upper
b_0	-2.70	-2.68	0.27	-3.25	-2.20
b_1	0.21	0.09	0.38	-0.53	0.95
b_2	0.50	0.49	0.33	-0.14	1.15
b_3	0.89	0.90	0.30	0.31	1.50
b_4	0.97	1.09	0.46	0.08	1.88
b_5	1.21	1.27	0.41	0.40	2.00
sigma^2	0.40	0.32	0.18	0.15	0.81

Table 2: Final model parameter estimates on the log odds scale

Results from Model (1) and Model (2) were congruent, as shown in Table @ref(tab:est_tab). The 95% highest posterior density (HPD) credible intervals from Model (2) offered the advantage of concluding 95% confidence that estimates for each parameter lay within those bounds.

4 Limitations and Alternatives

Logistic regression assumes linearity of $log(\frac{p}{1-p})$ (Roback & Legler 2021). Limitations include lack of robust diagnostic assessment of GLMMs (Agresti 2013). Presence of competing risks in the experimental design obscure the presence of family clusters of cataracts. Simplifying the dataset by instituting a cutoff age attempts to address this limitation, but the issue persists even after excluding individuals deceased before Age = 552.

5 Results

The final probability for developing cataracts can be written as:

$$\begin{aligned} p_{ij} =& 0.063 + v_j + 0.55 * Gamma_i + 0.623 * HZE_i + \\ & 0.708 * Male_i + 0.726 * Gamma_i * Male_i + 0.770 * HZE_i * Male_i \end{aligned} \tag{3}$$

Comparisons of different groups across Sex and Treatment were assessed on Model (1) using the `emmeans` package (Lenth 2022). Figure 3 shows of the probability of developing cataracts for each combination of sex by treatment group. Differences between both sex and treatment group were clearly visible, notably the differences between treatment group within gender.

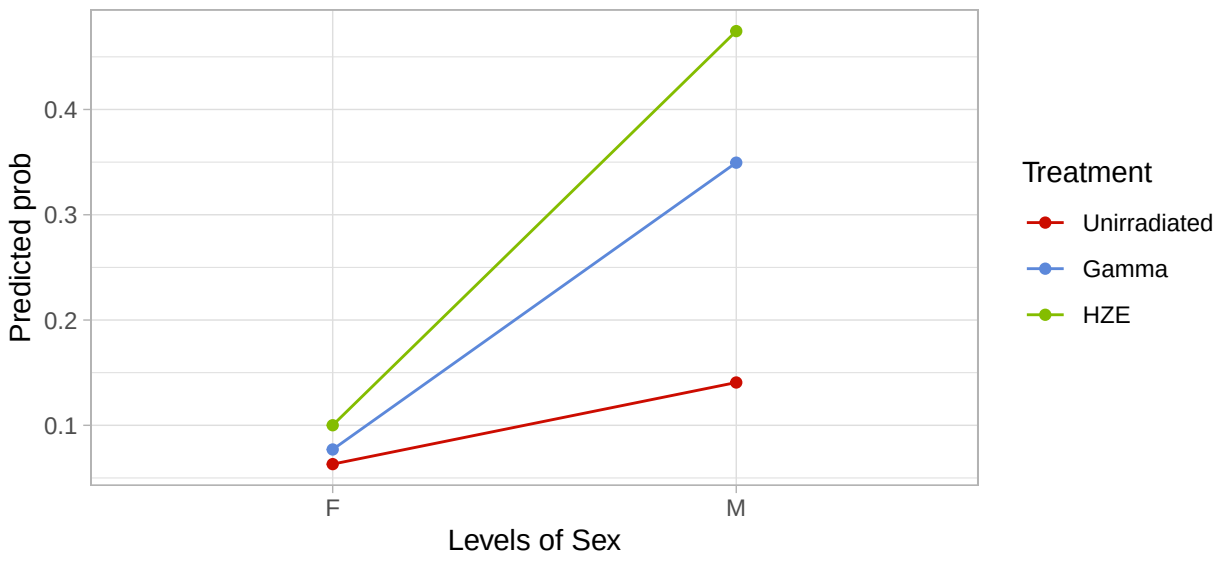

Figure 3: Predicted probability of cataracts by sex, treatment group

Figure 4 shows the relative risk of cataract development for all group comparisons in Model (1). Within females, all relative risks between treatment groups resulted in confidence intervals containing the possibility of no increased risk from exposure. Within males, conversely, all three comparisons between treatment groups showed strong evidence for increased risk with exposure. The largest difference was observed among Males between the HZE group and the control group.

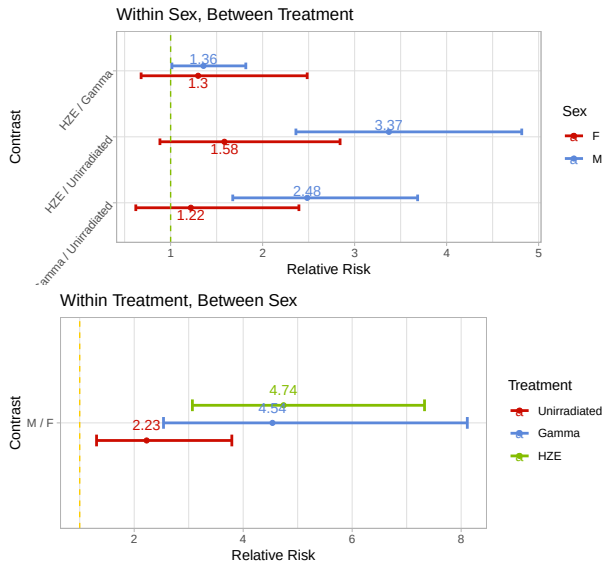

Figure 4: Relative risk of cataract development between all group combinations

Between sex, a greater relative risk of cataract development was observed for males than females within all treatment groups, although variability between treatment group remained salient. Males had a 4.74 and 4.54 times greater risk of developing severe cataracts than females in the HZE and Gamma groups respectively, compared to only 2.23 times greater risk within the control group.

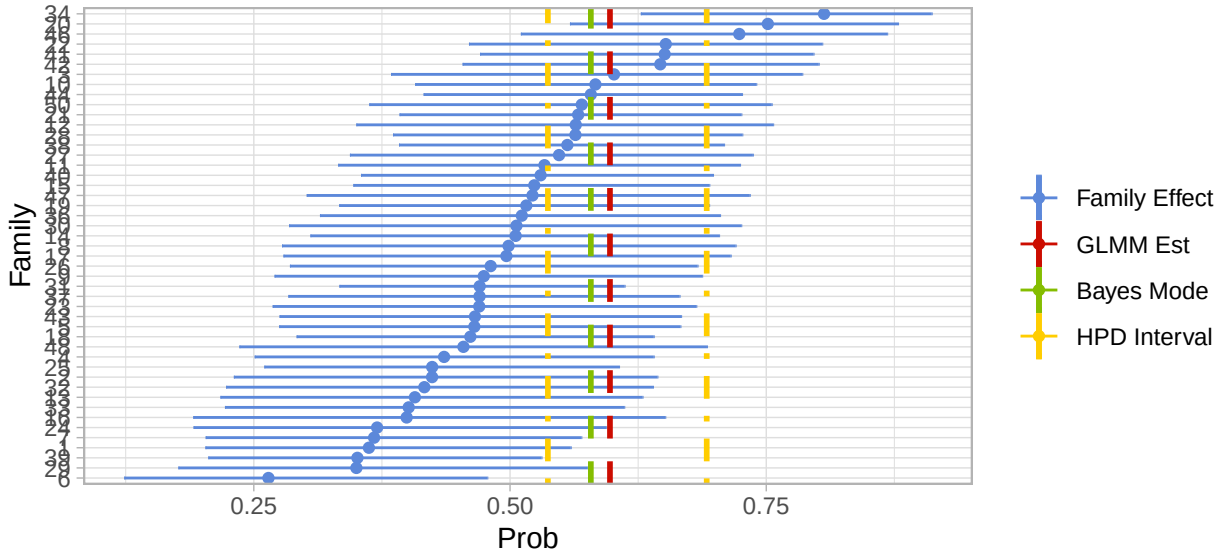

Figure 5: Estimated random probability of developing cataracts by family

Figure 5 visualizes the estimated variability of developing cataracts by family on the probability scale, along with confidence intervals. The vertical lines plot the GLMM estimated variance from Model (1) and the MCMC mode and HPD interval for σ^2 from Model (2). The dispersion on the tails of the probability scale indicate plausible genetic predisposition or resistance to cataract development.

6 Conclusions

Females faced lower probability of developing cataracts than males across all treatment groups, and the association of treatment group with cataract status was not apparent for female mice in this data set. Males, on the other hand, had a higher probability of developing cataracts than females across all treatment groups, and the differences of cataract development between treatment groups was of clinical importance. After accounting for sex and treatment group, additional random variability in the probability of cataract development can be attributed to Family; on the probability scale, 95% of the distribution of the variation due to family lies within the HPD interval (0.537, 0.692). Calculated random effects for the 47 families in the data showed variability above and below this range, providing strong evidence for genetic predisposition or resistance to cataract development.

7 Author's Statements

Both authors contributed to statistical analysis and writing for this report. Both authors read and approved the final report. Both authors contributed equally to introduction, exploratory data analysis, summary statistics, and results. Alyssa Allsop served as the primary statistician on fitting Frequentist models; Amira Burns served as the primary statistician on fitting the Bayesian model.

8 References

A Appendix

A.1 Additional Exploratory Data Analysis

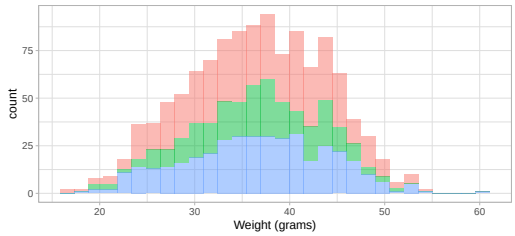

Figure 6: Histograms of contiguous covariates by treatment group

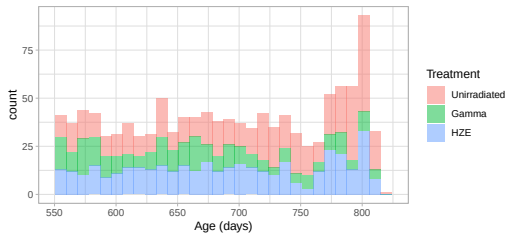

Figure 7: Histograms of contiguous covariates by treatment group

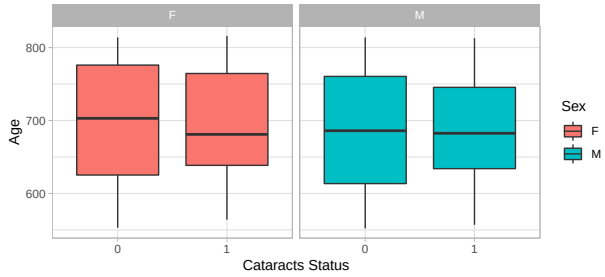

Figure 8: Boxplot of Age, cataracts status by sex and treatment

MyeloidLeukemia	HarderianTumor	PreTLymphoma	n
0	bilateral	0	39
0	none	0	954
0	none	1	6
0	unilateral	0	134
0	unilateral	1	1
1	none	0	30
1	none	1	1
1	unilateral	0	4

Table 3: Counts of cancer status

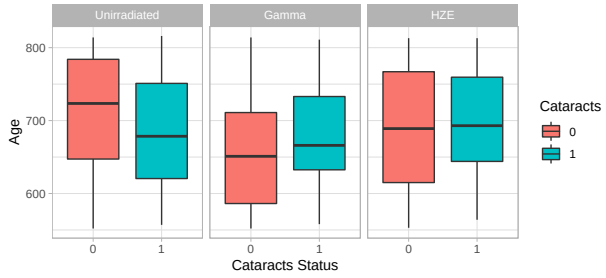

Figure 9: Boxplot of Age, cataracts status by sex and treatment

A.2 Model Selection

AIC was used as the primary model selection criteria because it can be used to compare non-nested models. This feature was of particular importance when assessing models with a random effect against a model with only fixed effects.

Model	AIC
Final Model	1017.98
Full Mixed Model	1038.9
Fixed Effects Model	1040.3
Mixed Model with no Interaction	1023.19
Base Model	1123.57

Table 4: Model selection via AIC comparision

A.3 GLM Model Diagnostics

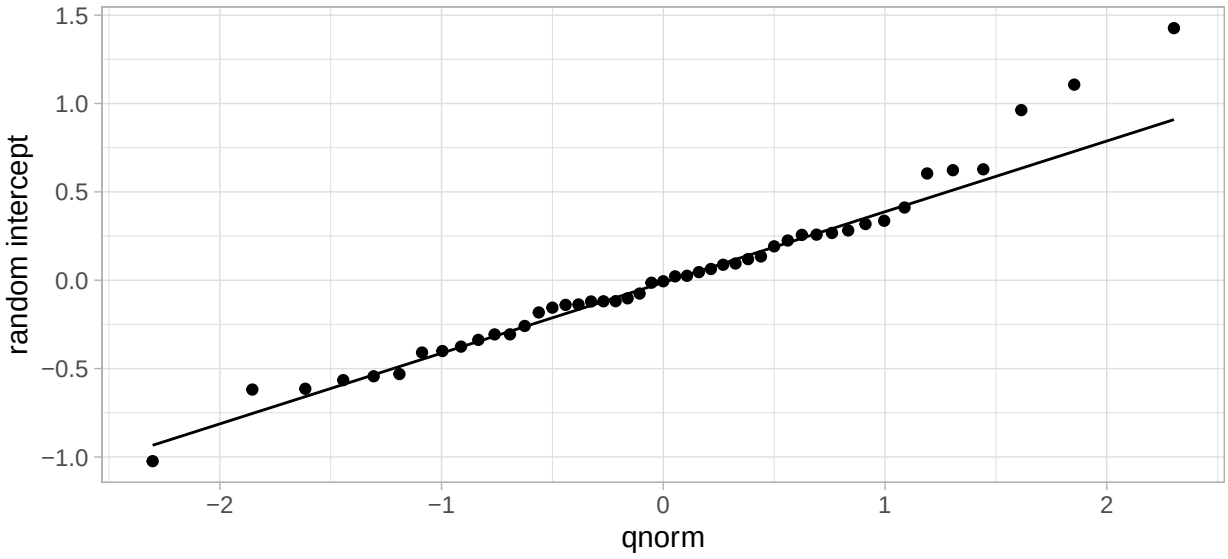

Figure 10: QQ-plot of random effect on intercept

Assumptions for fitting this model were: the random effects came from a normal distribution; the chosen link function was appropriate; and estimation of variance was not over-dispersed. These assumptions were verified. Quantiles of the random effect were compared with the quantiles of a normal distribution and looked approximately normal in Figure 10. The chosen logit link function was appropriate because the response used was binary. The ratio of the chi square statistic to the residual degrees of freedom was 0.89; a value ≤ 1 indicated no over-dispersion.

A.4 Bayesian Model Diagnostics

A.5 Source Code

```
knitr::opts_chunk$set(fig.height=3, fig.pos = 'H', fig.align = 'center',
                      echo=FALSE, warning=FALSE, message=FALSE)


library(readxl)
library(tidyverse)
library(DescTools)
library(lme4)
library(broom.mixed)
library(kableExtra)
library(xtable)
library(emmeans)
library(ggsci)
library(sjPlot)
library(coda)
library(rjags)
library(R2jags)
library(superdiag)
library(mcmcplots)
library(ggmcmc)
library(gridExtra)
cats <- read_excel(path = "GRSD.cataract.xlsx", sheet = "Sheet1")
# remove spaces from column and value names
names(cats) <- str_replace_all(names(cats), " ", "_")
cats <- cats %>%
  mutate(CoatColor = str_replace_all(coat_color, " ", "_"))


# turn categorical vars into factor
```

```r
cats <- cats %>%
  rename(Age = `age_(days)`,
         Weight = weight,
         Animal = animal) %>%
  mutate(Sex = as.factor(sex),
         CoatColor = as.factor(CoatColor),
         Family = as.factor(family), # should this stay a factor? Yes?
         BCS = as.ordered(BCS),
         Treatment = relevel(as.factor(groups), ref = "Unirradiated"),
         MyeloidLeukemia = as.factor(Myeloid_Leukemia),
         HarderianTumor = as.factor(Harderian_Tumor),
         PreTLymphoma = as.factor(PreT_Lymphoma),
         Score = as.ordered(Cataract_Score)) # ordinal cat; leave as numeric?

# select vars, add binary conversion for score
cats <- cats %>%
  select(c(Animal, Sex, Weight, CoatColor, Family, BCS, Age, Treatment,
           MyeloidLeukemia, HarderianTumor, PreTLymphoma, Score)) %>%
  mutate(Cataracts = ifelse(Score < 2, 0, 1))

# glance at the dataset
# str(cats)
gsc <- cats %>% group_by(Score) %>%
  count(Treatment) %>%
  pivot_wider(names_from = Score, values_from = n) %>%
  group_by(Treatment) %>%
  mutate(Total = sum(c(`1`, `2`, `3`, `4`)))

xtable(gsc, label = "tab:gtab", caption = "Counts of score by treatment group") %>%
  xtable2kable(booktabs = T,
```

```
                include.rownames = FALSE,
                table.placement = NULL) %>%
  add_header_above(c( " " = 1, "Cataract Score" = 4, " " = 1)) %>%
  kable_styling(full_width = F, position = "float_right")
# barplot of proportion of each group with cataracts
sex_trt <- cats %>%
  group_by(Treatment, Sex, Cataracts) %>%
  summarise(n = n()) %>%
  ungroup() %>%
  group_by(Treatment, Sex) %>%
  mutate(proportion = round(n/sum(n), digits = 2))

cats_grp <- sex_trt %>%
  filter(Cataracts == 1)

b1 <- ggplot(cats_grp, aes(x = Sex, y = proportion, fill = Treatment)) +
  geom_col(position = "dodge") +
  scale_fill_startrek() +
  theme_light()

b1
# Plot of average score of sex-group-family
gr_score <- cats %>%
  group_by(Treatment, Family) %>%
  summarize(mean_score = mean(Cataracts))

l1 <- ggplot(gr_score, aes(x = Treatment, y = mean_score, color = Family)) +
  geom_line(aes(group = Family)) +
  scale_x_discrete(expand = c(0, .2)) +
  theme_light() +
```

```
  theme(legend.position="none") +
  labs(y = "mean score")
l1
mod <- glmer(Cataracts ~ Treatment*Sex + (1|Family), data = cats, family = binomial)
# Specify the model
cat("model{
  for(i in 1:N){
    CAT[i] ~ dbern(p[i])     # Bernoulli-distributed response
    logit(p[i])<- b0+ b1*Gamma[i] + b2*HZE[i] + b3*Male[i] +
    b4*Male[i]*Gamma[i] + b5*Male[i]*HZE[i] + a[Family[i]]   # likelihood function
  }
  for(j in 1:nFam){
    a[j] ~ dnorm(0, tau)
  }
  b0 ~ dnorm(0.0, 1.0E-3)   # vaguely informative priors
  b1 ~ dnorm(0.0, 1.0E-3)
  b2 ~ dnorm(0.0, 1.0E-3)
  b3 ~ dnorm(0.0, 1.0E-3)
  b4 ~ dnorm(0.0, 1.0E-3)
  b5 ~ dnorm(0.0, 1.0E-3)
  tau ~ dgamma(1.0E-3,1.0E-3)
  sigma2 <- 1/tau     # convert precision 'tau' to variance 'sigma2'
}", file = "cat.jag")

# Prepare the data for JAGS
# break Treatment into dummy variables for each group
treatment <- model.matrix(~ Treatment - 1, cats)
sex <- model.matrix(~Sex -1, cats)
colnames(treatment) <- c("Unirradiated", "Gamma", "HZE")
cats <- data.frame(cats, treatment, sex)
```

```r
# format relevant data as a list
data <- list(CAT = cats$Cataracts, Gamma = cats$Gamma,
             HZE = cats$HZE, Male = cats$SexM, Family = cats$Family,
             nFam = length(unique(cats$Family)), N = nrow(cats))

nIter <- 60000
nChains <- 3
nThin <- 1
BurnIn <- 10000
nAdapt <- 1000
ests <- summary(mod)$coef[,1] # pull starting values from frequentist model
var <- as.numeric(as.data.frame(VarCorr(mod))$vcov)
inits <- list(list("tau" = var+0.2, "b0" = ests[1]+0.5, "b1" = ests[2]+0.5, "b2" = ests
                   "b3" = ests[4]+0.2, "b4" = ests[5]+0.2, "b5" = ests[6]+0.2),
              list("tau" = var-0.2, "b0" = ests[1]-0.5, "b1" = ests[2]-0.5, "b2" = ests
                   "b3" = ests[4]-0.2, "b4" = ests[5]-0.2, "b5" = ests[6]-0.2),
              list("tau" = var, "b0" = ests[1], "b1" = ests[2], "b2" = ests[3],
                   "b3" = ests[4], "b4" = ests[5], "b5" = ests[6]))

# -- Compile and run the model
params <- c("b0", "b1", "b2", "b3", "b4", "b5", "sigma2")
set.seed(556)
model.fit <- jags(data = data,
                  inits = inits,
                  parameters.to.save = params,
                  model.file = "cat.jag",
                  n.chains = nChains,
                  n.iter = nIter,
                  n.burnin = BurnIn,
```

```
                  n.thin = nThin)
mcmc.model <- as.mcmc(model.fit)
#summary(mcmc.model)

posts <- mcmc.model[[3]][,-7]
#posts <- exp(posts)/(1 + exp(posts))
phpds <- HPDinterval(posts)
posts <- data.frame(posts)

# Create table of posterior estimates
means <- apply(posts, 2, mean)
medians <- apply(posts, 2, median)
mode_fun <- function(x) {
  ux <- round(unique(x), digits = 3)
  return(ux[which.max(tabulate(match(x, ux)))])
}
modes <- apply(round(posts, 4), 2, mode_fun)
sds <- apply(posts, 2, sd)
glmm_probs <- exp(ests) / (1 + exp(ests))
p_var <- exp(var) / (1 + exp(var))
est_tab <- round(data.frame(c(ests, var), means, medians, modes, sds, phpds), digits = 3

rownames(est_tab) <- c("b_0", "b_1", "b_2", "b_3",
                       "b_4", "b_5", "sigma^2")
colnames(est_tab) <- c("GLMM Est", "MCMC Mean", "MCMC Median",
                       "MCMC Mode", "MCMC SD", "HPD Lower", "HPD Upper")
est_tab_show <- est_tab %>% select(-c(2:3))
xtable(est_tab_show, label = "tab:est_tab",
       caption = "Final model parameter estimates on the log odds scale") %>%
  xtable2kable(booktabs = T, include.rownames = TRUE,
```

```
              table.placement = NULL) %>%
  kable_styling(full_width = F, latex_options = "HOLD_position")
cats_emms <- emmeans(mod, ~ Treatment | Sex, infer = TRUE, type = "response")
emmip(cats_emms, Treatment ~ Sex) +
  theme_light() + scale_color_startrek()
em1 <- emmeans(mod, ~ Treatment|Sex)
em1log <- regrid(em1, "log")
rrs1 <- contrast(em1log, interaction = "revpairwise", type = "response")
rrs1 <- as.data.frame(confint(rrs1)) %>%
  rename(Contrast = Treatment_revpairwise, Lower = asymp.LCL, Upper = asymp.UCL)

rrplot1 <- ggplot(rrs1, aes(x = ratio, y = Contrast, xmin = Lower, xmax = Upper)) +
  geom_errorbarh(aes(height = 0.2, color = Sex),
                 position = position_dodge(0.3), lwd = 1) +
  geom_point(aes(color = Sex), position = position_dodge(0.3)) +
  geom_text(aes(label = round(ratio, 2), color = Sex),
            position = position_dodge(0.7)) +
  theme_light() +
  geom_vline(aes(xintercept = 1), color = "#84BD00FF", lty = 2) +
  theme(axis.text.y = element_text(angle=50)) +
  scale_color_startrek() +
  labs(x = "Relative Risk",
       title = "Within Sex, Between Treatment")

em2 <- emmeans(mod, ~ Sex|Treatment)
em2log <- regrid(em2, "log")
rrs2 <- contrast(em2log, interaction = "revpairwise", type = "response")
rrs2 <- as.data.frame(confint(rrs2)) %>%
  rename(Contrast = Sex_revpairwise, Lower = asymp.LCL, Upper = asymp.UCL)
```

```
rrplot2 <- ggplot(rrs2, aes(x = ratio, y = Contrast, xmin = Lower, xmax = Upper)) +
  geom_errorbarh(aes(height = 0.2, color = Treatment),
                 position = position_dodge(0.3), lwd = 1) +
  geom_point(aes(color = Treatment), position = position_dodge(0.3)) +
  geom_text(aes(label = round(ratio, 2), color = Treatment),
            position = position_dodge(0.3), vjust = -1) +
  theme_light() +
  geom_vline(aes(xintercept = 1), color = "#FFCD00FF", lty = 2) +
  scale_color_startrek() +
  labs(x = "Relative Risk", title = "Within Treatment, Between Sex")
par(mar = c(4, 4, .2, .2));
rrplot1
rrplot2
p_sigs <- est_tab[7,]
est <- as.numeric(p_sigs[1])
est <- exp(est) / (1 + exp(est))
psig <- as.numeric(p_sigs[4])
psig <- exp(psig) / (1 + exp(psig))
hpdl <- as.numeric(p_sigs[6])
hpdl <- exp(hpdl) / (1 + exp(hpdl))
hpdu <- as.numeric(p_sigs[7])
hpdu <- exp(hpdu) / (1 + exp(hpdu))
REs <- augment(ranef(mod,condVar = TRUE), ci.level = 0.95) %>%
  select(c(level, estimate, lb, ub)) %>%
  rename(Family = level) %>%
  mutate(Prob = exp(estimate)/(1+exp(estimate)),
         Lower = exp(lb)/(1+exp(lb)),
         Upper = exp(ub)/(1+exp(ub)))
colors <- c("Family Effect" = "#5C88DAFF", "GLMM Est" = "#CC0C00FF", "Bayes Mode" = "#8
            "HPD Interval" = "#FFCD00FF")
```

```
ggplot(REs, aes(x = Prob, y = Family, xmin = Lower, xmax = Upper)) +
  geom_errorbarh(aes(height = 0, color = "Family Effect")) +
  geom_point(aes(color = "Family Effect")) +
  geom_vline(aes(xintercept = est, color = "GLMM Est"), lwd = 1, lty = 2) +
  geom_vline(aes(xintercept = psig, color = "Bayes Mode"), lwd = 1, lty = 2) +
  geom_vline(aes(xintercept = hpdl, color = "HPD Interval"), lwd = 1, lty = 4) +
  geom_vline(aes(xintercept = hpdu, color = "HPD Interval"), lwd = 1, lty = 4) +
  theme_light() +
  scale_color_manual(values = colors) +
  labs(color = "")
wt <- ggplot(cats, aes(x = Weight, fill = Treatment)) +
  geom_histogram(alpha = 0.5) +
  theme_light() +
  theme(legend.position = "none") +
  labs(x = "Weight (grams)")

ag <- ggplot(cats, aes(x = Age, fill = Treatment)) +
  geom_histogram(alpha = 0.5) +
  theme_light() +
  labs(x = "Age (days)")

wt
ag
cancers <- cats %>%
  group_by(MyeloidLeukemia, HarderianTumor, PreTLymphoma) %>%
  count()

xtable(cancers, caption = "Counts of cancer status") %>%
  xtable2kable(booktabs = T, include.rownames = FALSE, table.placement = NULL) %>%
  kable_styling(full_width = F, latex_options = "hold_position")
```

```r
ggplot(cats, aes(x = as.factor(Cataracts), y = Age, fill = Sex)) +
  geom_boxplot() + facet_wrap(vars(Sex)) +
  theme_light() +
  labs(x = "Cataracts Status")

ggplot(cats, aes(x = as.factor(Cataracts), y = Age, fill = as.factor(Cataracts))) +
  geom_boxplot() + facet_wrap(vars(Treatment)) +
  theme_light() +
  labs(x = "Cataracts Status", fill = "Cataracts")

# mixed model binomial logistic regression
full_mod <- glmer(Cataracts ~ Treatment + Sex + Weight + CoatColor + BCS + HarderianTum
                    MyeloidLeukemia + PreTLymphoma + (1|Family), data = cats, family =
                  control=glmerControl(optimizer="bobyqa",optCtrl=list(maxfun=2e5)))
no_interaction <- glmer(Cataracts ~ Treatment + Sex + (1|Family), data = cats, family =
final_mod <- mod
simple_mod <- glmer(Cataracts ~ Treatment + (1|Family), data = cats, family = binomial)
fixed_mod <- glm(Cataracts ~ Treatment*Sex, data = cats, family = binomial)
aics <- c(round(AIC(final_mod),2), round(AIC(full_mod),2), round(AIC(fixed_mod),2), rou
models <- c("Final Model", "Full Mixed Model",
            "Fixed Effects Model", "Mixed Model with no Interaction", "Base Model")
dt <- cbind(models,aics)
colnames(dt) <- c("Model", "AIC")
xtable(dt, caption = "Model selection via AIC comparision") %>%
  xtable2kable(booktabs = T, include.rownames = FALSE, table.placement = NULL) %>%
  kable_styling(full_width = F, latex_options = "hold_position")
## assumption of normally distributed random effect
reff <- as.data.frame(ranef(mod)$Family) %>% rename(re = `(Intercept)`)

ggplot(reff, aes(sample = re)) +
```

```
  stat_qq() + stat_qq_line() +
  theme_light() +
  labs(x = "qnorm", y = "random intercept")
## assumption of no over-dispersion
rp <- residuals(final_mod, type = "pearson")
rat <- sum(rp^2)/df.residual(final_mod)
baysum <- summary(mcmc.model)
options(width=100)
```

References

Agresti, A. (2013), *Categorical Data Analysis*, third ed. edn, John Wiley & Sons.

Chernyavskiy, P., Edmondson, E. F., Weil, M. M. & Little, M. P. (2017), 'High-energy particle beam and gamma radiation exposure, familial relatedness and cancer in mice', *British journal of cancer* **117**(1), 41–50.
URL: *https://www.nature.com/articles/bjc2017141*

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. & Rubin, D. (2014), *Bayesian Data Analysis*, third ed. edn, CRC Press.

Lenth, R. V. (2022), *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.7.4-1.
URL: *https://github.com/rvlenth/emmeans*

Plummer, M. (2022), *rjags: Bayesian Graphical Models using MCMC*. R package version 4-13.
URL: *https://mcmc-jags.sourceforge.io*

Roback, P. & Legler, J. (2021), *Beyond Multiple Linear Regression. Applied Generalized Linear Models and Multilevel Models in R*, second ed. edn, CRC Press.
URL: *https://bookdown.org/roback/bookdown-BeyondMLR/*