Posterior Summary

Jonathan Ma

2025-11-15

Table of Contents

1 Posterior Summary

This section summarizes the estimated posterior distributions of the model parameters, including fixed and random effects. We compute point estimates, credible intervals, and derive odds ratios and intra-class correlation metrics to assess variance partitioning.

1.1 Summary

The following shows the first few rows of the full posterior summary for all parameters, including fixed effects, group-level effects, and auxiliary parameters.

```
posterior_summary <- posterior_summary(brm_model2)

## Loading required package: rstan

## Loading required package: StanHeaders

## rstan (Version 2.21.8, GitRev: 2e1f913d3ca3)

## For execution on a local, multicore CPU with excess RAM we recommend calli
ng
## options(mc.cores = parallel::detectCores()).
## To avoid recompilation of unchanged Stan programs, we recommend calling
## rstan_options(auto_write = TRUE)

##
## Attaching package: 'rstan'

## The following object is masked from 'package:tidyr':
##
##     extract

head(posterior_summary)

##                    Estimate Est.Error       Q2.5      Q97.5
## b_Intercept     -4.32250000  1.362928 -7.174176 -1.752077
## b_age01M04years -0.33931969  1.886682 -4.046930  3.323349
## b_age05M09years -1.16955713  1.915065 -4.824186  2.580967
## b_age10M14years  0.08948779  1.882779 -3.473392  3.905725
## b_age15M19years  0.07176377  1.904186 -3.653903  3.866177
## b_age20M24years -1.27660843  1.826520 -4.950319  2.311217
```

1.2 Fixed Effects

Below we visualize the fixed effect posterior estimates and 95% credible intervals. Coefficients are on the log-odds scale. Positive values indicate increased odds of late-stage diagnosis.

```
fe <- fixef(brm_model2, probs = c(0.025, 0.975))
head(fe)
```
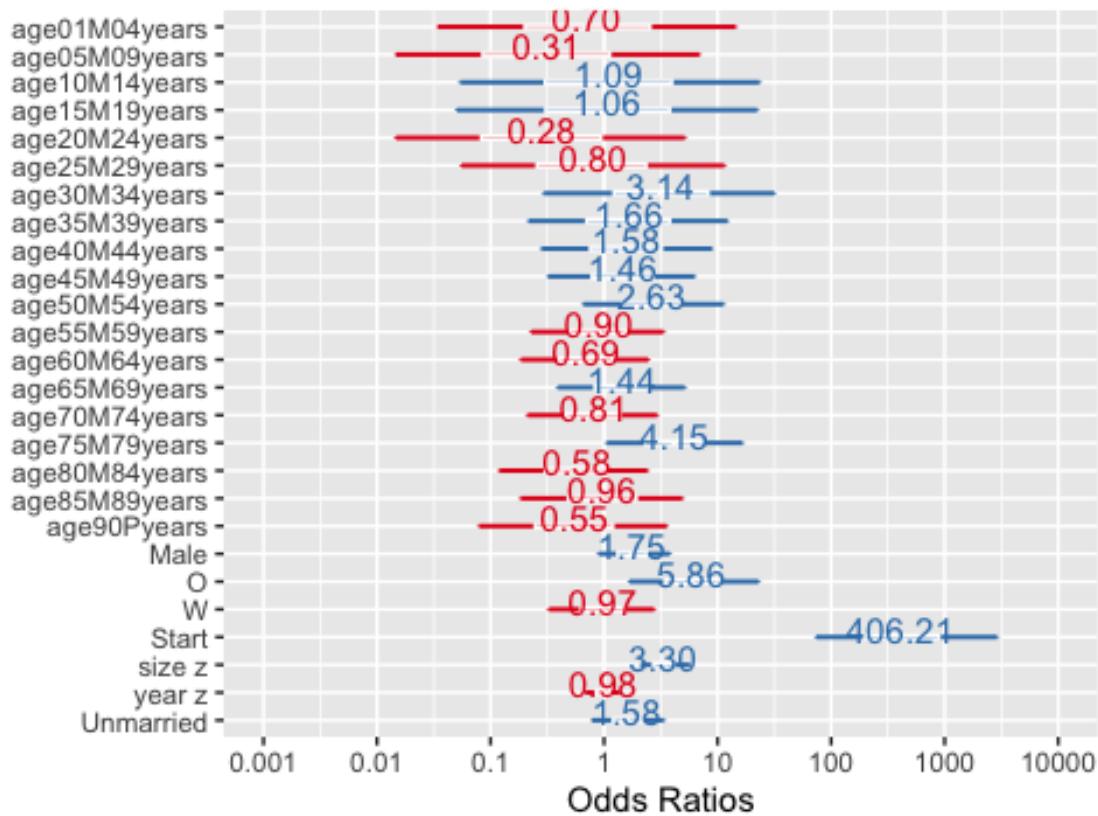
```
##                   Estimate Est.Error       Q2.5      Q97.5
## Intercept       -4.32250000  1.362928 -7.174176 -1.752077
## age01M04years   -0.33931969  1.886682 -4.046930  3.323349
## age05M09years   -1.16955713  1.915065 -4.824186  2.580967
## age10M14years    0.08948779  1.882779 -3.473392  3.905725
## age15M19years    0.07176377  1.904186 -3.653903  3.866177
## age20M24years   -1.27660843  1.826520 -4.950319  2.311217
```

```r
plot_model(brm_model2, type = "est", show.values = TRUE, value.offset = 0.3)
```


1.3 Random Effects (group level variation)

The following summarizes the estimated random intercepts (shrinkage effects) for patient and region. These account for individual and regional variability unexplained by covariates.

```r
re <- ranef(brm_model2)
head(re$patientid)
```

```
## , , Intercept
##
##       Estimate Est.Error      Q2.5      Q97.5
## 14    5.911493  6.108808 -2.503997 20.7671906
## 17    6.003108  6.421308 -3.264953 21.3294113
## 61    4.968166  6.615216 -4.880852 20.4883465
```

```
## 65   -8.835120   5.764587 -22.694805 -0.5660145
## 103   5.323214   6.085667   -3.444156 19.8594366
## 108   6.667558   6.200311   -2.043807 21.9015476

head(re$regionid)

## , , Intercept
##
##        Estimate Est.Error       Q2.5    Q97.5
## 1 -0.03564271 0.7260108 -1.559148 1.469797
## 2 -0.08212961 0.7922132 -1.831757 1.453806
## 3  0.15199696 0.7598713 -1.249647 1.824940
```

1.4 Odds Ratios

The odds ratio (OR) is the exponentiated posterior mean (or median) of each coefficient.

```
odds_ratios <- exp(fe)
odds_df <- as.data.frame(odds_ratios)
odds_df <- odds_df[order(abs(odds_df$Estimate - 1), decreasing = TRUE), ]
head(odds_df, 10)

##                  Estimate Est.Error        Q2.5        Q97.5
## gradeStart   427.47742764  3.080237 5.260676e+01 4472.3118442
## raceth0        5.99385960  2.256193 1.337940e+00   30.9401709
## age75M79years  4.13938554  2.364600 7.846040e-01   22.6481428
## size_z         3.37119134  1.355896 1.988640e+00    6.4890874
## age30M34years  3.11891155  4.312735 1.632972e-01   59.1595807
## age50M54years  2.65375774  2.435053 4.823600e-01   15.8427305
## Intercept      0.01326668  3.907619 7.661164e-04    0.1734133
## sexMale        1.78551454  1.575105 7.537673e-01    4.5340126
## age20M24years  0.27898188  6.212231 7.081147e-03   10.0866975
## age05M09years  0.31050442  6.787379 8.033092e-03   13.2099102
```

1.5 Marginal Effects

```
library(dplyr)
library(purrr)

# Get conditional effects
me_list <- conditional_effects(brm_model2)

# Safe cleaning function
clean_ce <- function(df, varname) {
  if ("effect1__" %in% names(df) && !is.character(df$effect1__)) {
    df$effect1__ <- as.character(df$effect1__)
  }
  df$variable <- varname
  df
}
```

```
# Apply cleaning to all conditional effects
me_combined <- bind_rows(
  lapply(names(me_list), function(v) clean_ce(me_list[[v]], v))
)

# Compute range and summarize by variable
me_summary <- me_combined %>%
  mutate(range = upper__ - lower__) %>%
  group_by(variable) %>%
  summarise(
    avg_range = mean(range, na.rm = TRUE),
    avg_estimate = mean(estimate__, na.rm = TRUE),
    n_levels = n()
  ) %>%
  arrange(desc(avg_range))

# Show top 10 most variable effects
head(me_summary, 10)

## # A tibble: 7 × 4
##    variable avg_range avg_estimate n_levels
##    <chr>        <dbl>        <dbl>    <int>
## 1 size_z       0.470        0.227      100
## 2 grade        0.384        0.429        2
## 3 age          0.286        0.0174      20
## 4 raceth       0.241        0.0337       3
## 5 sex          0.183        0.0188       2
## 6 marry        0.171        0.0177       2
## 7 year_z       0.146        0.0136      100
```

1.6   Shrinkage EFfects

```
patient_re <- as.data.frame(re$patientid)
region_re <- as.data.frame(re$regionid)
patient_re$group <- rownames(re$patientid)
region_re$group <- rownames(re$regionid)

# patient-level shrinkage
top_patient <- patient_re %>%
  arrange(desc(abs(Estimate.Intercept))) %>%
  slice(1:30)   # top 30 most extreme patients

ggplot(top_patient, aes(x = reorder(group, Estimate.Intercept),
                        y = Estimate.Intercept)) +
  geom_point() +
  geom_errorbar(aes(ymin = Q2.5.Intercept, ymax = Q97.5.Intercept), width = 0.
2) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "gray") +
  coord_flip() +
  labs(
```
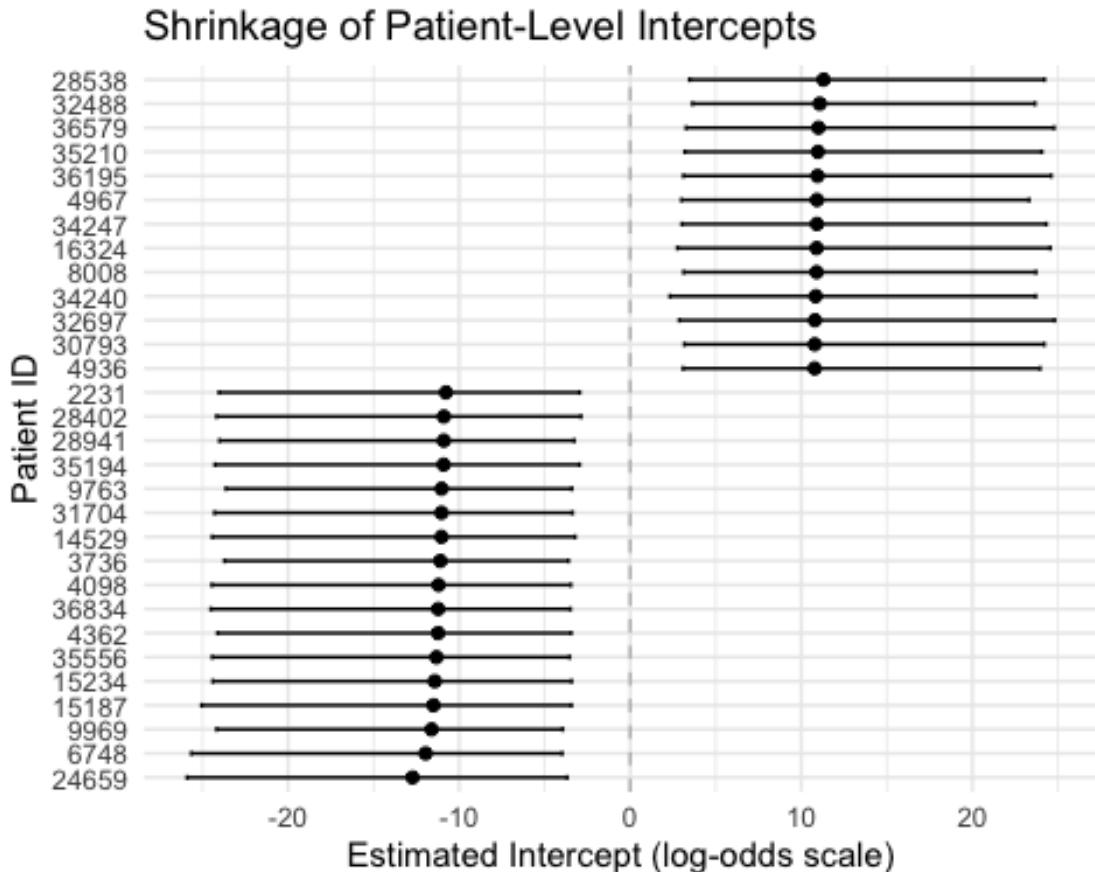
```
    title = "Shrinkage of Patient-Level Intercepts",
    x = "Patient ID",
    y = "Estimated Intercept (log-odds scale)"
  ) +
  theme_minimal()
```


Shrinkage of Patient-Level Intercepts
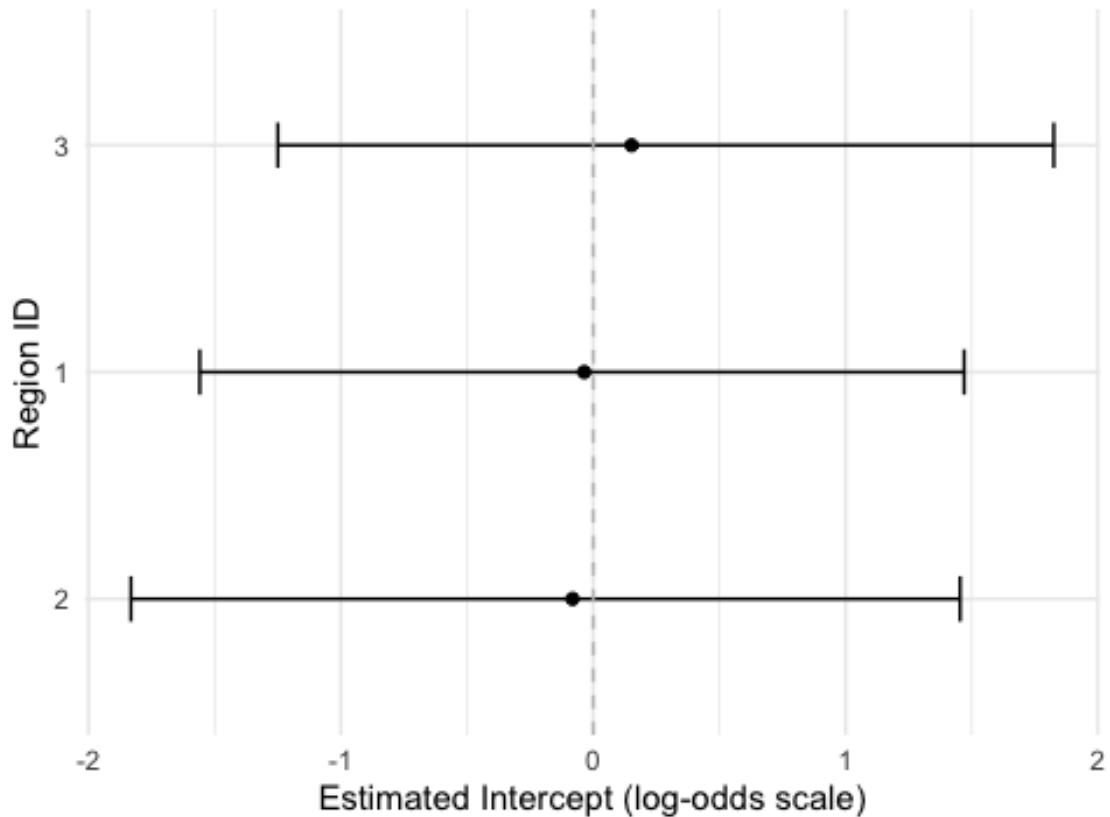
```
# region-level shrinkage
top_region <- region_re %>%
  arrange(desc(abs(Estimate.Intercept)))

ggplot(top_region, aes(x = reorder(group, Estimate.Intercept),
                       y = Estimate.Intercept)) +
  geom_point() +
  geom_errorbar(aes(ymin = Q2.5.Intercept, ymax = Q97.5.Intercept), width = 0.
2) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "gray") +
  coord_flip() +
  labs(
    title = "Shrinkage of Region-Level Intercepts",
    x = "Region ID",
    y = "Estimated Intercept (log-odds scale)"
  ) +
  theme_minimal()
```

Shrinkage of Region-Level Intercepts

1.7 Raw vs Shrunk Estimates

Red points represent posterior-predicted probabilities, which incorporate both observed data and model priors. Blue points are empirical averages (raw rate). Posterior estimates are shrunk toward the global mean, especially for small regions.

```r
raw_region <- df %>%
  group_by(regionid) %>%
  summarise(
    raw_rate = mean(latestage),
    n = n()
  )

region_re <- as.data.frame(ranef(brm_model2)$regionid)
region_re$regionid <- as.integer(rownames(ranef(brm_model2)$regionid))
region_post <- region_re %>%
  select(regionid, estimate = Estimate.Intercept,
         lower = Q2.5.Intercept, upper = Q97.5.Intercept)

region_compare <- left_join(raw_region, region_post, by = "regionid") %>%
  mutate(
    posterior_prob = plogis(estimate),  # convert log-odds to probability
```
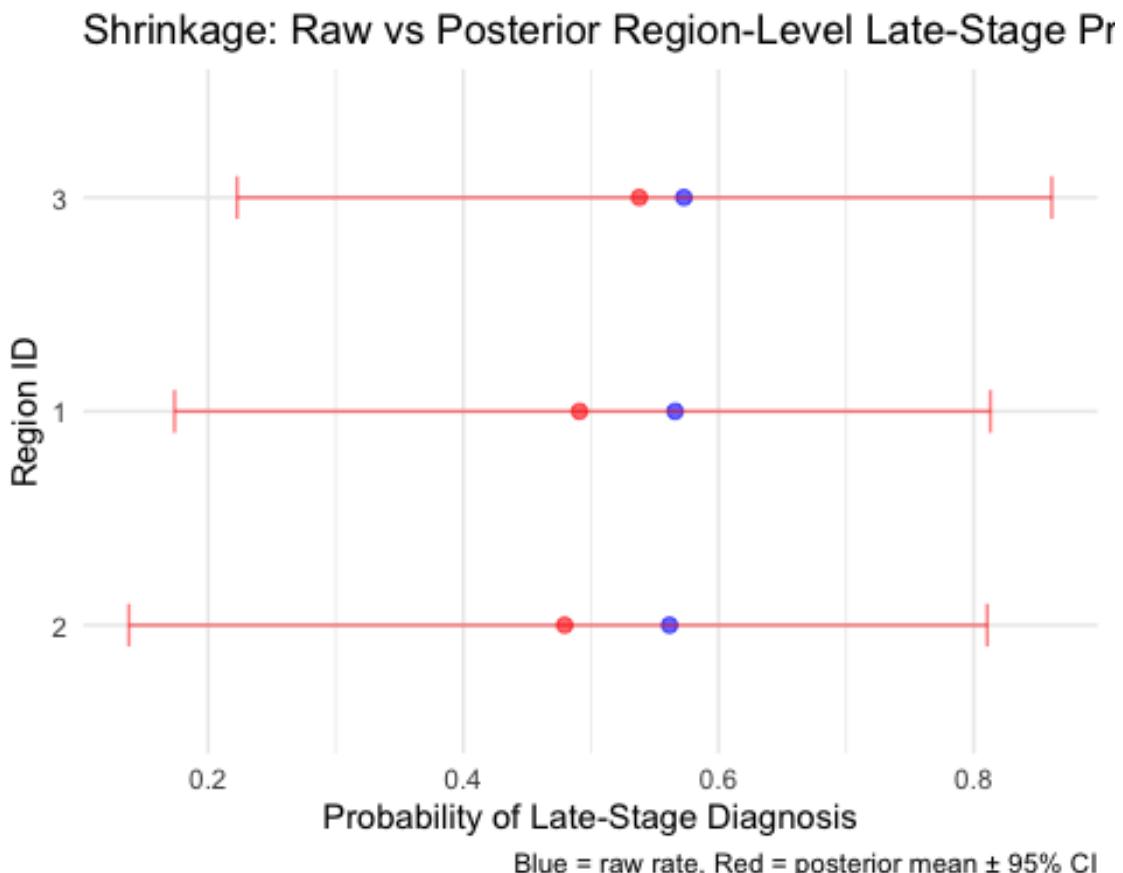
```
    lower_prob = plogis(lower),
    upper_prob = plogis(upper)
  )

ggplot(region_compare, aes(x = reorder(regionid, raw_rate))) +
  geom_point(aes(y = raw_rate), color = "blue", size = 2, alpha = 0.6) +
  geom_point(aes(y = posterior_prob), color = "red", size = 2, alpha = 0.6) +
  geom_errorbar(aes(ymin = lower_prob, ymax = upper_prob), color = "red", wid
th = 0.2, alpha = 0.5) +
  coord_flip() +
  labs(
    title = "Shrinkage: Raw vs Posterior Region-Level Late-Stage Probabilitie
s",
    x = "Region ID",
    y = "Probability of Late-Stage Diagnosis",
    caption = "Blue = raw rate, Red = posterior mean ± 95% CI"
  ) +
  theme_minimal()
```


Shrinkage: Raw vs Posterior Region-Level Late-Stage Pr

Blue = raw rate, Red = posterior mean ± 95% CI

1.8   Intra-Class Correlation

ICC (intra-class correlation) quantifies the proportion of total variance attributable to grouping structure. Here we compute ICC for both patients and regions.

```r
posterior <- as_draws_df(brm_model2)
posterior <- posterior %>%
  mutate(
    var_patient = sd_patientid__Intercept^2,
    var_region  = sd_regionid__Intercept^2,
    var_resid   = (pi^2) / 3,

    ICC_patientid = var_patient / (var_patient + var_region + var_resid),
    ICC_regionid  = var_region  / (var_patient + var_region + var_resid)
  )
# Compute posterior summaries
icc_summary <- posterior %>%
  summarise(
    SD_patient = mean(sd_patientid__Intercept),
    SD_region = mean(sd_regionid__Intercept),

    ICC_patient_mean = mean(ICC_patientid),
    ICC_patient_low  = quantile(ICC_patientid, 0.025),
    ICC_patient_high = quantile(ICC_patientid, 0.975),

    ICC_region_mean = mean(ICC_regionid),
    ICC_region_low  = quantile(ICC_regionid, 0.025),
    ICC_region_high = quantile(ICC_regionid, 0.975)
  )

# Create simple summary table
icc_table <- tibble(
  Group = c("Patient", "Region"),
  SD = c(icc_summary$SD_patient, icc_summary$SD_region),
  ICC = c(icc_summary$ICC_patient_mean, icc_summary$ICC_region_mean),
  CI = c(
    sprintf("[%.3f, %.3f]", icc_summary$ICC_patient_low, icc_summary$ICC_pati
ent_high),
    sprintf("[%.3f, %.3f]", icc_summary$ICC_region_low, icc_summary$ICC_regio
n_high)
  )
)

knitr::kable(icc_table, caption = "Posterior Intra-Class Correlation Estimate
s")
```

Posterior Intra-Class Correlation Estimates

Group	SD	ICC	CI
Patient	9.1081563	0.9427376	[0.841, 0.978]
Region	0.7869934	0.0150378	[0.000, 0.112]

```r
posterior %>%
  select(ICC_patientid, ICC_regionid) %>%
```

```
  pivot_longer(everything(), names_to = "Level", values_to = "ICC") %>%
  ggplot(aes(x = ICC, fill = Level)) +
  geom_density(alpha = 0.5) +
  labs(title = "Posterior ICC Distributions",
      x = "ICC", y = "Density") +
  theme_minimal()
```

```
## Warning: Dropping 'draws_df' class as required metadata was removed.
```