

Bayesian Coursework

2024-04-08

```
# Load dataset.  
sgaj <- read.csv("sgaj.csv")  
head(sgaj, 7)
```

```
##   patient female age time measured  
## 1       1     1  19    0    199.5  
## 2       1     1  19    6    239.4  
## 3       1     1  19   12    163.5  
## 4       1     1  19   18    268.1  
## 5       1     1  19   24    228.9  
## 6       2     0  19    0    172.4  
## 7       2     0  19    6    159.3
```

Part 1

```
summary(sgaj$age)  
  
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.  
## 19.00  40.75  56.00  55.67  77.25  88.00
```

```
# Create age groups  
sgaj_with_age_groups <- sgaj %>%  
  mutate(age_group = case_when(  
    age >= 18 & age <= 30 ~ "18-30",  
    age >= 31 & age <= 50 ~ "31-50",  
    age >= 51 & age <= 70 ~ "51-70",  
    age > 70 ~ "71+"  
)
```

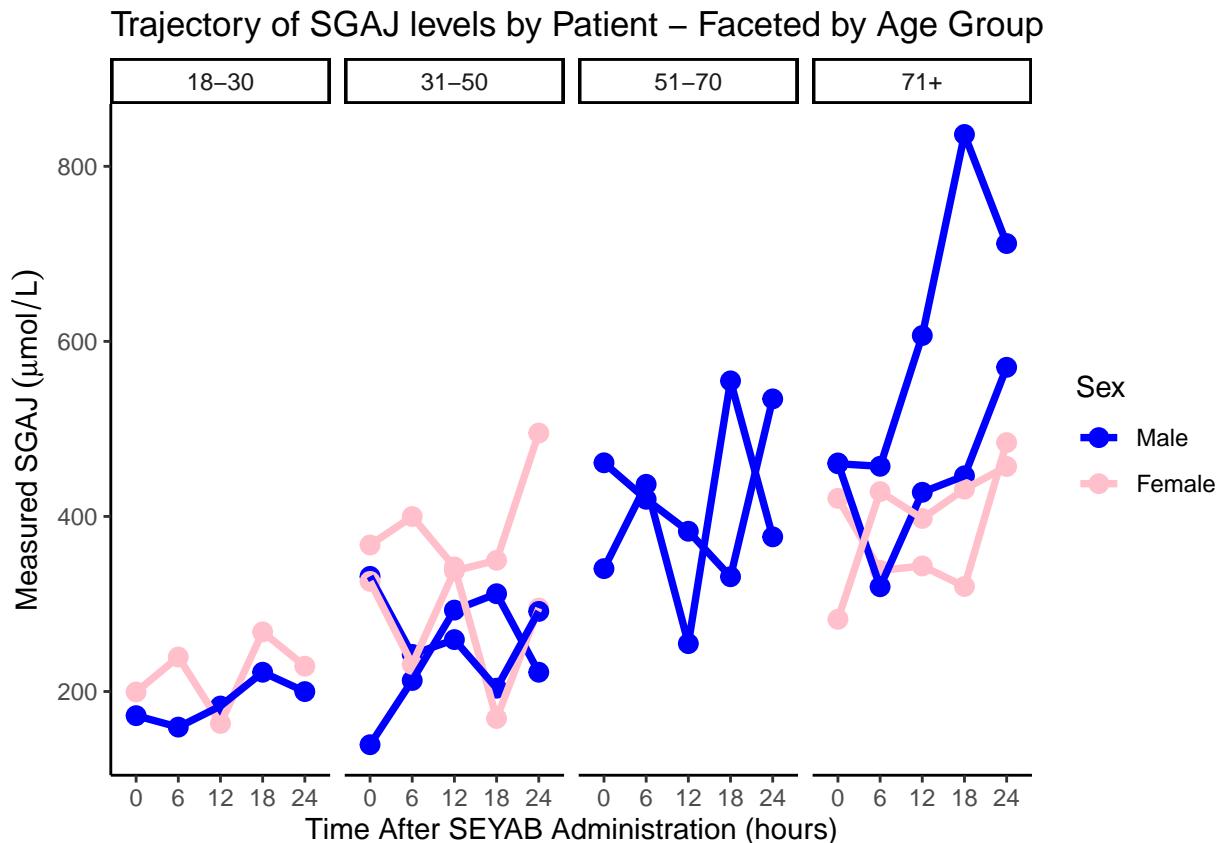
After reviewing the age distribution in the dataset, I decided to categorise the age variable into four groups: 18-30 (young adults), 31-50 (middle-aged adults), 51-70 (older adults), and 71+ (seniors). This is done solely for visualisation purposes (not used in the model) as it allows for the ordering (and separation in facets) of patients based on their age, making the exploration of how age and gender affect SGAJ easier to interpret.

```
ggplot(sgaj_with_age_groups, aes(x = as.factor(time), y = measured, group = patient, color = as.factor(  
  geom_point(size=3) +  
  geom_line(lineWidth=1.3) +  
  facet_wrap(~age_group, ncol=4) +  
  theme_classic() +  
  labs(
```

```

title = "Trajectory of SGAJ levels by Patient – Faceted by Age Group",
x = "Time After SEYAB Administration (hours)",
y = expression(Measured~SGAJ~(mu*mol/L)),
color = "Sex"
) +
scale_color_manual(
  values = c("blue", "pink"),
  labels = c("Male", "Female") # Set labels for Male and Female
)

```



The plot shows the trajectory of SGAJ level (across the timepoints) for each patient, faceted by the different age groups (to aid visualisation). The following insights can be drawn from the plot:

- There appears to be a positive association between age and SGAJ levels, suggesting that older patients tend to have higher biomarker concentrations.
- In younger age groups, female patients display higher SGAJ levels than male patients (in most time points). In contrast, in the senior age group (71+ years), this pattern is reversed, with male patients presenting higher SGAJ levels.
- Each patient has a different starting level of SGAJ (time=0) which suggests the need for a random intercept. Furthermore, the rate at which SGAJ levels change over time is not constant across patients; In general, older patients have a faster rate SGAJ growth than younger patients but there is still a lot of variability between patients, suggesting the need for a random slope on time.
- Given the limited samples (12 patients), not strong conclusions can be drawn. Furthermore, I acknowledge that unmeasured factors (such as medication, lifestyle) might influence SGAJ levels and could confound the observed associations with age and sex.

Part 2

2.1 Assumptions

- The SGAJ level for patient i at time t follows a Normal distribution.
- *Linearity:* The model is linear. The relationship between the dependent variable (SGAJ) and each of the independent variables is assumed to be linear.
- *Additive:* Simplifying assumption that there is no interaction between the covariates (model is additive).
- *Exchangeability:* Before considering the data, we treat all patients as essentially the same. The prior distributions for the group-specific parameters α_i and γ_i reflect this assumption. After seeing the data, the model allows these parameters to differ, but the starting point is a belief in their similarity (common prior distribution).

2.2 Model formulation

The SGAJ level for each patient i at time t is assumed to follow a normal distribution with mean μ_{it} and variance ψ^2 . The expected SGAJ measurement for patient i at time t (at reference values) is modeled as μ_{it} . Following the suggestion from the brief, I am fixing the measurement error's standard deviation (ψ) to 75.

$$\begin{aligned}y_{it} &\sim \mathcal{N}(\mu_{it}, \psi^2) \\ \mu_{it} &= \alpha_i + \beta_{\text{age}} (\text{age}_i - 50) / 10 + \beta_{\text{female}} (\text{female}_i) + \gamma_i (\text{time}_t - 12) / 6 \\ \psi &= 75\end{aligned}$$

2.2.1 Priors on alpha

- The variability between patients i is represented by a random effect α_i (patient-specific random intercept), which is assumed to be drawn from a Normal distribution with mean μ_a and variance σ_a^2 .

$$\begin{aligned}\alpha_i &\sim \mathcal{N}(\mu_a, \sigma_a^2) \\ \mu_a &\sim \mathcal{N}(225, 20.41^2) \\ \sigma_a &\sim \text{Uniform}(80, 100)\end{aligned}$$

- I am assuming that the random intercept a_i representing the deviation in baseline SGAJ level for each patient from the population average under reference conditions (age=50, female, time=12 hours), follows a Normal distribution.
- μ_a represents the average value (intercept) across the population for the reference group when other predictors (e.g., age, female, time) are held at their reference levels. Being influenced by the information from the previous study, I am assuming that μ_a has an average value of 225. The standard deviation of μ_a is calculated as $(265-185)/(1.96*2)=20.41$, with a 95% upper and lower credible limit of 265 and 185 respectively.
- For the random effects standard deviation σ_a , I am setting it as a uniform distribution between 80 and 100 to account for some uncertainty around the value 90 suggested from the previous study (because they said it is roughly 90).

2.2.2 Priors on beta_age and beta_female

- The terms involving β_{age} , and β_{female} correspond to the fixed effects of age (centered around the age of 50 and scaled such that a unit increase is in 10 years), and female (has the value 1 if female and zero otherwise), respectively.

$$\beta_{age} \sim \mathcal{N}(0, 10^2)$$

$$\beta_{female} \sim \mathcal{N}(-0.1, 10^2)$$

- For the β_{age} and β_{female} as the brief explains, the effect is not currently understood well enough and as such I am assuming vague priors, allowing the data to play a primary role in informing these relationships. For the prior of β_{female} I am subtly incorporating the expectation that SGAJ levels might be slightly lower in females than in males (as explained by the principal investigator).

2.2.3 Priors on gamma

- γ_i represents the patient-specific rate of change of SGAJ levels per day, accounting for individual variability in response to the drug over time. This is modeled as a random effect to capture the difference in how each patient's SGAJ levels evolve from the baseline measurement at 12 hours after SEYAB administration. Specifically, $\gamma_i (time_t - 12) / 6$ calculates the deviation in SGAJ levels for each patient i at any given time t , normalized to a 6-hour period (hours per measurement). This reflects the biological variability among patients - without it we would just have different starting points of SGAJ levels among patients (random intercept) but the rate of change over time would be constant across individuals. I am assuming that γ_i is drawn from a Normal distribution with mean μ_γ and variance σ_γ^2 .

$$\gamma_i \sim \mathcal{N}(\mu_\gamma, \sigma_\gamma^2)$$

$$\mu_\gamma \sim \mathcal{N}(0, 10^2)$$

$$\sigma_\gamma \sim \text{Uniform}(0, 100)$$

- For mu_{gamma} and $sigma_{gamma}$ I chose vague priors due to the lack of specific prior information, ensuring that the inferences are driven predominantly by the observed data rather than by subjective assumptions. I chose wide uniform and large variance normal distributions for these priors to offer the flexibility needed for the model to adapt and learn from the data.

2.3 RJAGS

TODO: add pic of drawn model

```
# Initialisation of data
data <- list(n_patients = length(unique(sgaj$patient)),
             patient = sgaj$patient,
             age = sgaj$age,
             time = sgaj$time,
             female = sgaj$female,
             n = nrow(sgaj),
             y = sgaj$measured
)
```

Constructing a hierarchical linear regression model based on the assumptions and model formulation described above. Checking convergence of the model from two different chains - starting from slightly different initial values.

```

# Model definition
model <- "model{
  # Linear regression model
  for (i in 1:n) {
    mu[i] <- alpha[patient[i]] + beta_age*(age[i] - 50)/10 + beta_female*female[i] + gamma[patient[i]]*
      y[i] ~ dnorm(mu[i], prec_psi)
    yrep[i] ~ dnorm(mu[i], prec_psi) # posterior-predictive distribution
  }

  # Random effects
  for (p in 1:n_patients) {
    alpha[p] ~ dnorm(mu_alpha, prec_alpha)
    gamma[p] ~ dnorm(mu_gamma, prec_gamma)
    rate_of_change[p] <- gamma[p] * 4
  }

  # Priors for fixed effects
  beta_age ~ dnorm(0, 1/(10^2))
  beta_female ~ dnorm(-0.1, 1/(10^2))

  # Prior for the population average of the random intercepts
  mu_alpha ~ dnorm(225, 1/(20.41^2))

  # Prior for the standard deviation of the random intercepts
  sd_alpha ~ dunif(80,100)
  prec_alpha <- 1/(sd_alpha^2)

  # Prior for the population average of the random slope
  mu_gamma ~ dnorm(0, 1/(10^2))

  # Prior for the standard deviation of the random slope
  sd_gamma ~ dunif(0, 100)
  prec_gamma <- 1/(sd_gamma^2)

  prec_psi <- 1/(75^2)
}""

# Two different initial values - to check that the model is converging to the same conclusion
initial_values <- list(list(alpha=rep(200, data$n_patients), sd_alpha=80, mu_gamma=0, sd_gamma=20, beta=0))

# Model setup
jags_model <- jags.model(textConnection(model),
                          data = data,
                          inits = initial_values,
                          n.chains = 2)

## Compiling model graph
## Resolving undeclared variables
## Allocating nodes
## Graph information:
##   Observed stochastic nodes: 60
##   Unobserved stochastic nodes: 90

```

```

##      Total graph size: 649
##
## Initializing model

# Burn-in
update(jags_model, 1000)

pars <- c("alpha", "mu_alpha", "sd_alpha", "beta_age", "beta_female", "gamma", "mu_gamma", "sd_gamma", "sigma_e")

# Sampling
set.seed(2292)
samples <- coda.samples(jags_model,
                        variable.names = pars,
                        n.iter = 200000,
                        thin=10)

# Get draws
sgaj_draws <- as_draws(samples)

```

2.4 Summarising posterior distributions of key quantities of interest

```

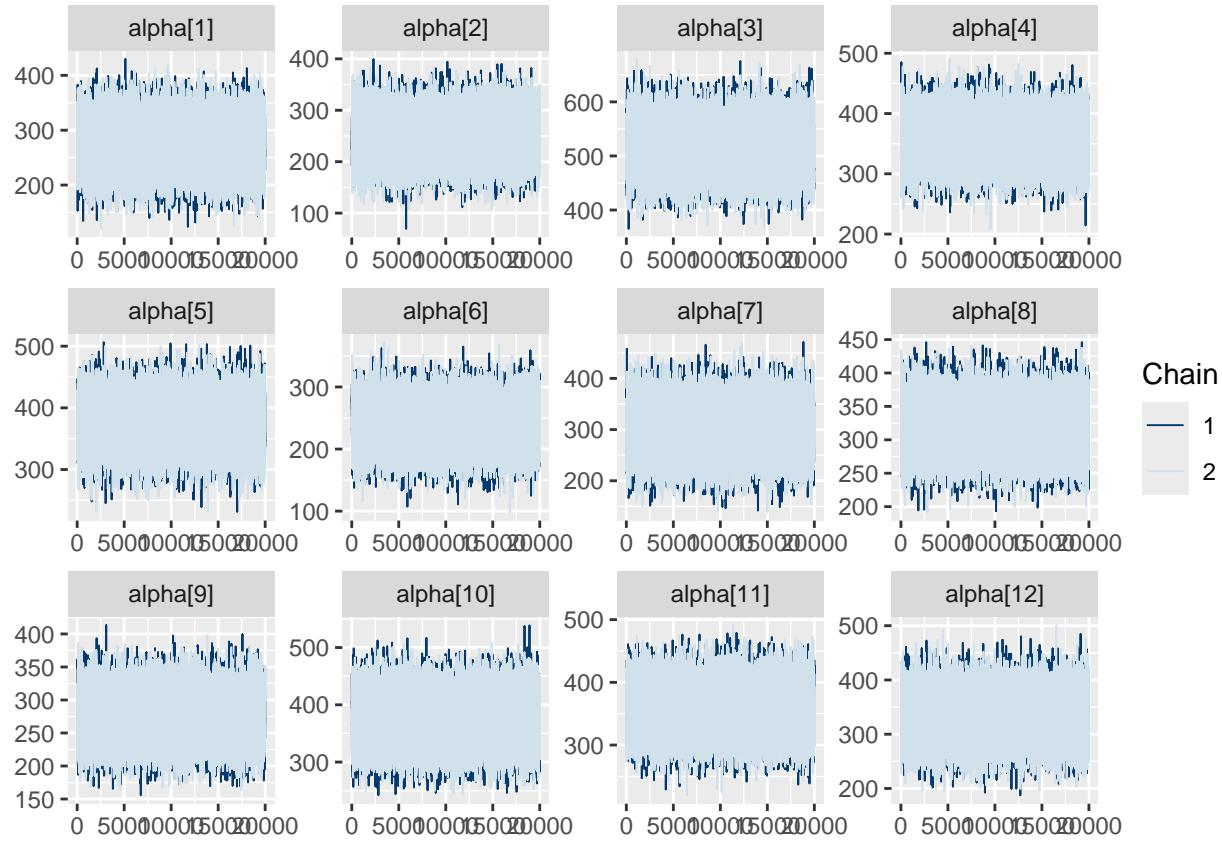
# Function that returns posterior summary for only specified parameters.
# The summary includes the 2.5%, 50%, 97.5% quantiles, convergence measures (rhat, ESS), and monte carlo standard error.
get_posterior_summary <- function(draws, pars) {
  subsetted_draws <- subset_draws(draws, variable = pars)
  summary_subsetted_draws <- summary(subsetted_draws, quantile(.x, probs = c(0.025, 0.5, 0.975)), defatult = TRUE)

  return(summary_subsetted_draws)
}

```

2.4.1 Random intercept

```
mcmc_trace(samples, pars=c(paste0("alpha[",1:12,"]")))
```



```
print(get_posterior_summary(sgaj_draws, c(paste0("alpha[", 1:12, "]"))))
```

```
## # A tibble: 12 x 9
##   variable `2.5%` `50%` `97.5%` rhat ess_bulk ess_tail mcse_mean mcse_median
##   <chr>     <dbl>   <dbl>    <dbl>  <dbl>     <dbl>     <dbl>      <dbl>
## 1 alpha[1]    200.    276.    351.  1.00  38254.   37962.    0.199   0.254
## 2 alpha[2]    175.    250.    324.  1.00  37372.   37761.    0.197   0.251
## 3 alpha[3]    440.    516.    594.  1.00  37792.   38858.    0.202   0.209
## 4 alpha[4]    293.    357.    421.  1.00  39815.   39548.    0.164   0.214
## 5 alpha[5]    307.    374.    442.  1.00  38952.   38994.    0.174   0.233
## 6 alpha[6]    177.    239.    301.  1.00  40746.   38694.    0.157   0.222
## 7 alpha[7]    218.    300.    382.  1.00  36867.   38142.    0.218   0.250
## 8 alpha[8]    255.    320.    385.  1.00  39794.   39515.    0.167   0.201
## 9 alpha[9]    214.    276.    339.  1.00  39815.   39587.    0.160   0.187
## 10 alpha[10]   302.    375.    448.  1.00  36794.   39043.    0.194   0.234
## 11 alpha[11]   291.    358.    424.  1.00  40342.   39924.    0.167   0.224
## 12 alpha[12]   260.    334.    408.  1.00  38165.   39076.    0.193   0.238
```

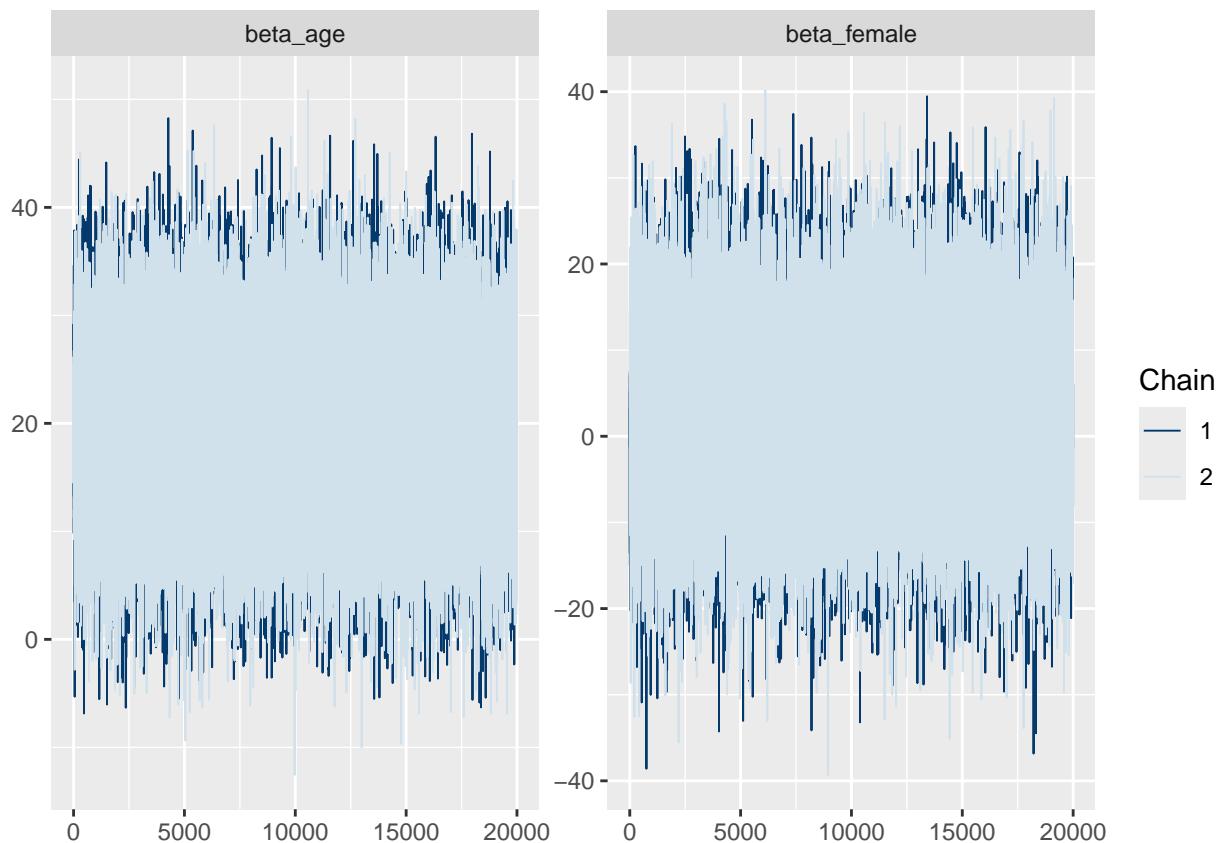
Convergence: The traceplots suggest a well-mixed distribution (“fat hairy caterpillar” appearance), indicating stable estimates with random dispersion around the mean. The overlap of both chains and an \hat{R} value of ~ 1.00 for all parameters show no signs of convergence issues. This implies that within-chain and between-chain variances are comparable. The ESS values range from approximately 36,000 to 40,000 for both bulk and tail. High ESS values suggest that the sample provides a reliable approximation of the posterior distribution, reducing the error in our estimates and increasing the confidence in statistical inferences. The monte carlo standard error (MCSE) values for the mean range from 0.16 to 0.22 and for the median from

0.19 to 0.25. These values indicate the variability of the estimate if the entire Monte Carlo simulation were repeated under the same conditions. Given the magnitude of the estimated parameters, the observed MCSE values are notably low, which substantiates the precision and reliability of our estimates.

Interpretation: The median values (50% quantile), representing the central tendency of the random intercept estimates, demonstrate considerable variation across patients. Notably, patient 6 exhibits both the lowest median at 239 and the narrowest 2.5% to 97.5% quantile interval, ranging from 177 to 301. Conversely, patient 3 shows the highest median at 516, accompanied by the broadest interval from 440 to 594.

2.4.2 Fixed effects

```
mcmc_trace(samples, pars=c("beta_age", "beta_female"))
```



```
print(get_posterior_summary(sgaj_draws, c("beta_age", "beta_female")))
```

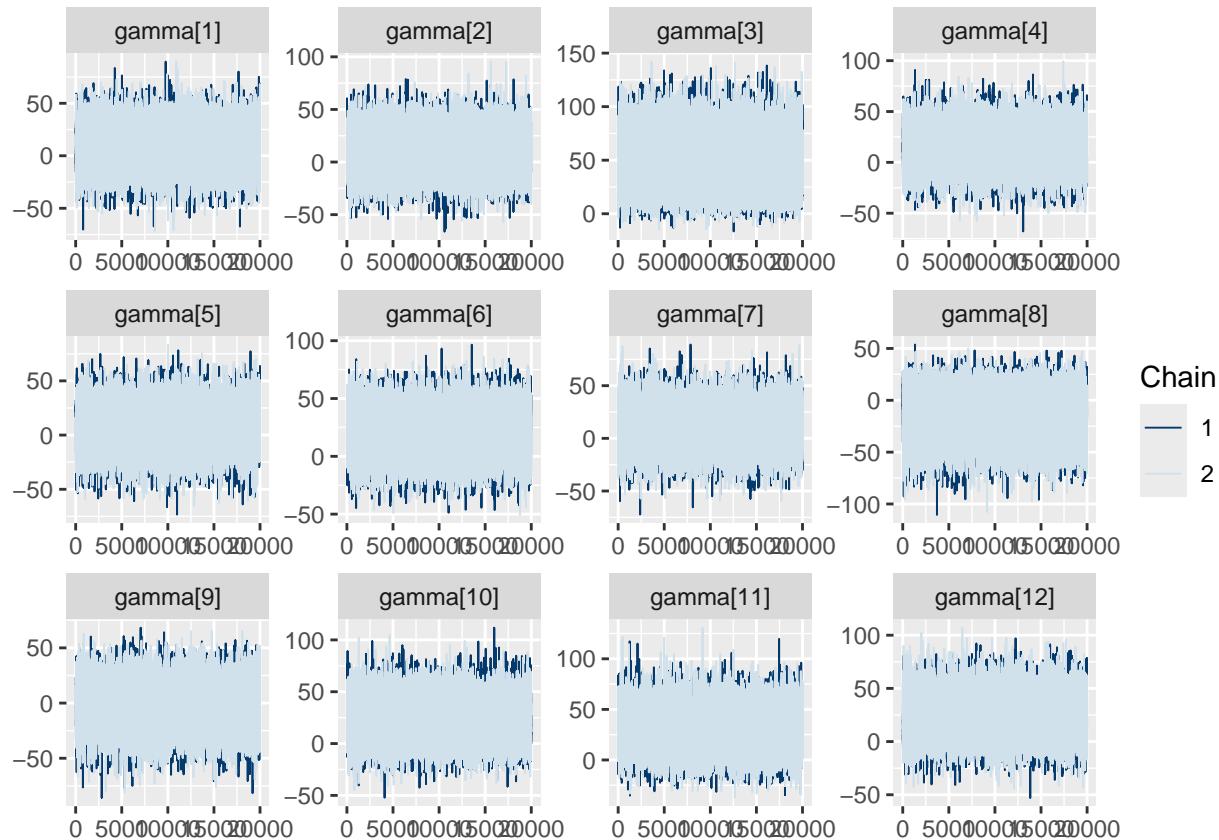
```
## # A tibble: 2 x 9
##   variable    '2.5%'  '50%'  '97.5%' rhat ess_bulk ess_tail mcse_mean mcse_median
##   <chr>      <dbl>   <dbl>   <dbl>     <dbl>     <dbl>     <dbl>       <dbl>
## 1 beta_age    4.43   19.5    34.7    1.00    34913.   38777.    0.0414    0.0465
## 2 beta_female -16.1   3.10    22.2    1.00    39373.   40434.    0.0492    0.0685
```

Convergence: The traceplots show stable mean with random dispersion around the mean and both chains overlap. The \hat{R} statistic is estimated to be ~ 1.00 , giving no suggestion of any lack of convergence. The mean MCSE values for both are smaller than 0.05.

Interpretation: The median value of the β_{age} coefficient at 19.5 indicates that SGAJ levels typically rise by about 19.5 (4.43-34.7) units for each decade past 50 years, and decrease by the same amount per decade before 50. The β_{female} coefficient's median at 3.1 suggests females have SGAJ levels that are typically 3.1 units higher than males (all other things constant), but with a broad 95% credible interval from -16.11 to 22.24, there is considerable uncertainty around this effect.

2.4.3 Random slope

```
mcmc_trace(samples, pars=c(paste0("gamma[", 1:12, "]")))
```



```
print(get_posterior_summary(sgaj_draws, c(paste0("gamma[", 1:12, "]"))))
```

```
## # A tibble: 12 x 9
##   variable `2.5%` `50%` `97.5%` rhat ess_bulk ess_tail mcse_mean mcse_median
##   <chr>     <dbl>   <dbl>   <dbl>    <dbl>     <dbl>     <dbl>      <dbl>
## 1 gamma[1]  -24.0    10.8    42.2    1.00   39407.   38758.    0.0821    0.101
## 2 gamma[2]  -22.3    11.8    43.2    1.00   37240.   37669.    0.0836    0.0927
## 3 gamma[3]     7.67   43.4    94.0    1.00   15031.   12845.    0.181     0.230
## 4 gamma[4]   -18.1    14.4    47.8    1.00   39548.   37645.    0.0813    0.0901
## 5 gamma[5]   -25.8    9.73   40.6    1.00   36434.   37357.    0.0860    0.101
## 6 gamma[6]   -13.7   17.0    52.4    1.00   37918.   38577.    0.0836    0.112
## 7 gamma[7]   -22.7   11.4    43.3    1.00   36989.   38443.    0.0839    0.0918
## 8 gamma[8]   -53.2   -8.38   22.7    1.00   18092.   23322.    0.147     0.191
```

```

##  9 gamma[9]   -36.1    2.77    31.7   1.00   28269.   37753.   0.104   0.146
## 10 gamma[10]   -9.64   20.1     56.9   1.00   32611.   38082.   0.0923  0.119
## 11 gamma[11]   -4.33   24.7     65.4   1.00   27651.   37364.   0.107   0.148
## 12 gamma[12]   -8.72   20.4     58.0   1.00   32777.   37746.   0.0929  0.112

```

Convergence: The traceplots show stable mean with random dispersion around the mean and both chains overlap. The \hat{R} statistic is estimated to be ~ 1.00 across all the gammas (no suggestion of any lack of convergence). The mean MCSE values are between 0.08 and 0.18.

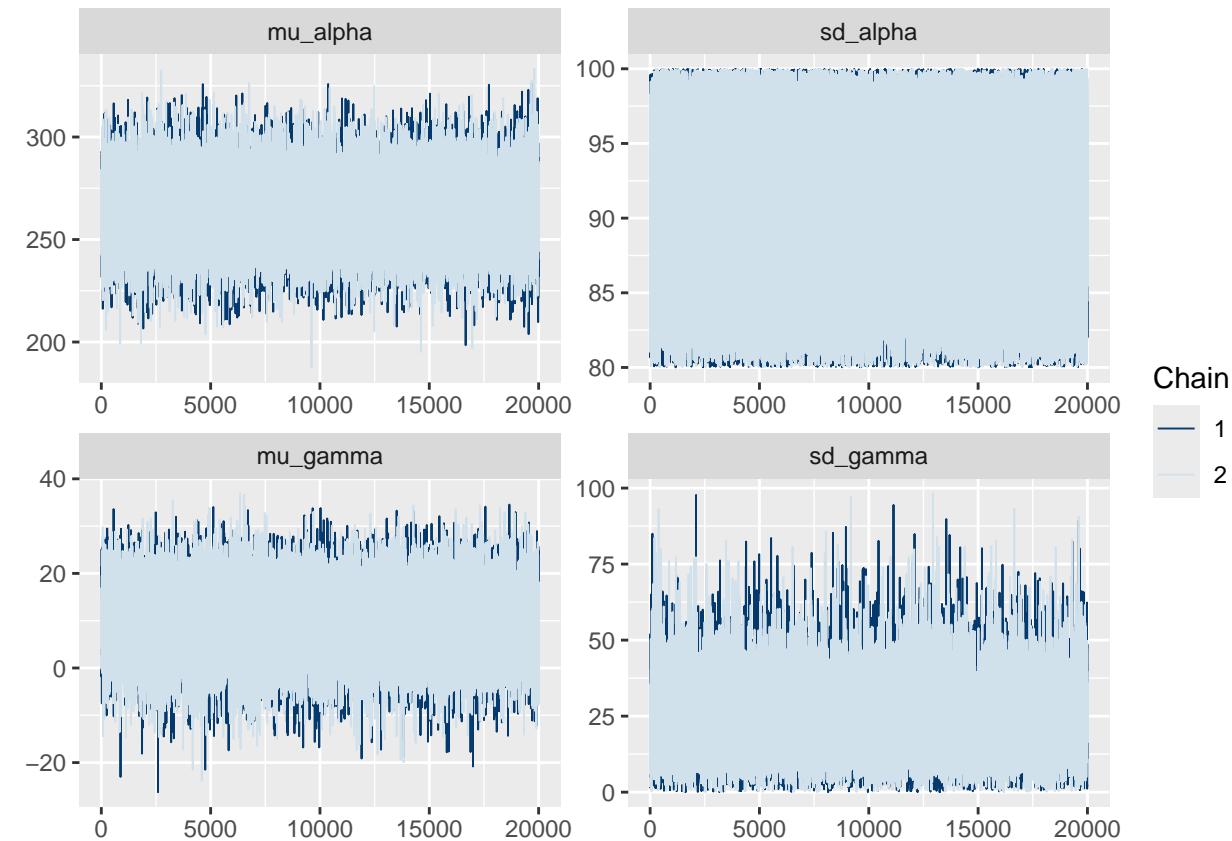
#TODO: *Interpretation:* The median values (50% quantile), representing the central tendency of the random slope estimates, demonstrate significant variation across patients. Patient 3 has the highest random slope

2.4.4 Hyper-parameters

```

# Check convergence for random intercept (alpha).
mcmc_trace(samples, pars=c("mu_alpha", "sd_alpha", "mu_gamma", "sd_gamma"))

```



```
get_posterior_summary(sgaj_draws, c("mu_alpha", "sd_alpha", "mu_gamma", "sd_gamma"))
```

```

## # A tibble: 4 x 9
##   variable `2.5%` `50%` `97.5%` rhat ess_bulk ess_tail mcse_mean mcse_median
##   <chr>     <dbl>   <dbl>   <dbl>   <dbl>     <dbl>     <dbl>      <dbl>
## 1 mu_alpha 232.    265.    298.    1.00    40403.   39915.    0.0842   0.113

```

```

## 2 sd_alpha  80.8   92.1    99.7   1.00   38634.   37896.   0.0284   0.0409
## 3 mu_gamma -4.60   10.5    24.0   1.00   29915.   29618.   0.0418   0.0538
## 4 sd_gamma  2.17   22.1    50.1   1.00   9891.    6365.    0.109    0.0985

```

Convergence: The traceplots show stable mean with random dispersion around the mean and both chains overlap. All hyper-parameters indicate convergence with \hat{R} values at ~ 1.00 . The MCSE for is under 0.1 for all the hyperparameters.

Interpretation: The median posterior estimate for mu_alpha at 265 (232, 298) suggests that the average baseline SGAJ level across the population (reference conditions - age=50, female, time=12 hours) is approximately 265 units, a considerably higher value than the suggested average SGAJ (225) from the previous cohort. The posterior median for sigma_alpha (standard deviation across all individuals) at reference conditions is 92.1 (80.8, 99.7) which is also slightly higher than the value suggested from the previous study.

The median value for mu_gamma at 10.5 implies an average rate of increase across the population (reference conditions) of about 10.5 units in SGAJ levels every 6 hours. However, the broad 95% credible interval from -4.6 to 24 indicates substantial uncertainty, suggesting that in some scenarios, SGAJ levels might even decrease. The median of sd_gamma at 22.1, with a range from 2.17 to 50.1, reflects high variability in the rate of change of SGAJ levels among patients. This suggests that individual responses to treatment can vary notably.

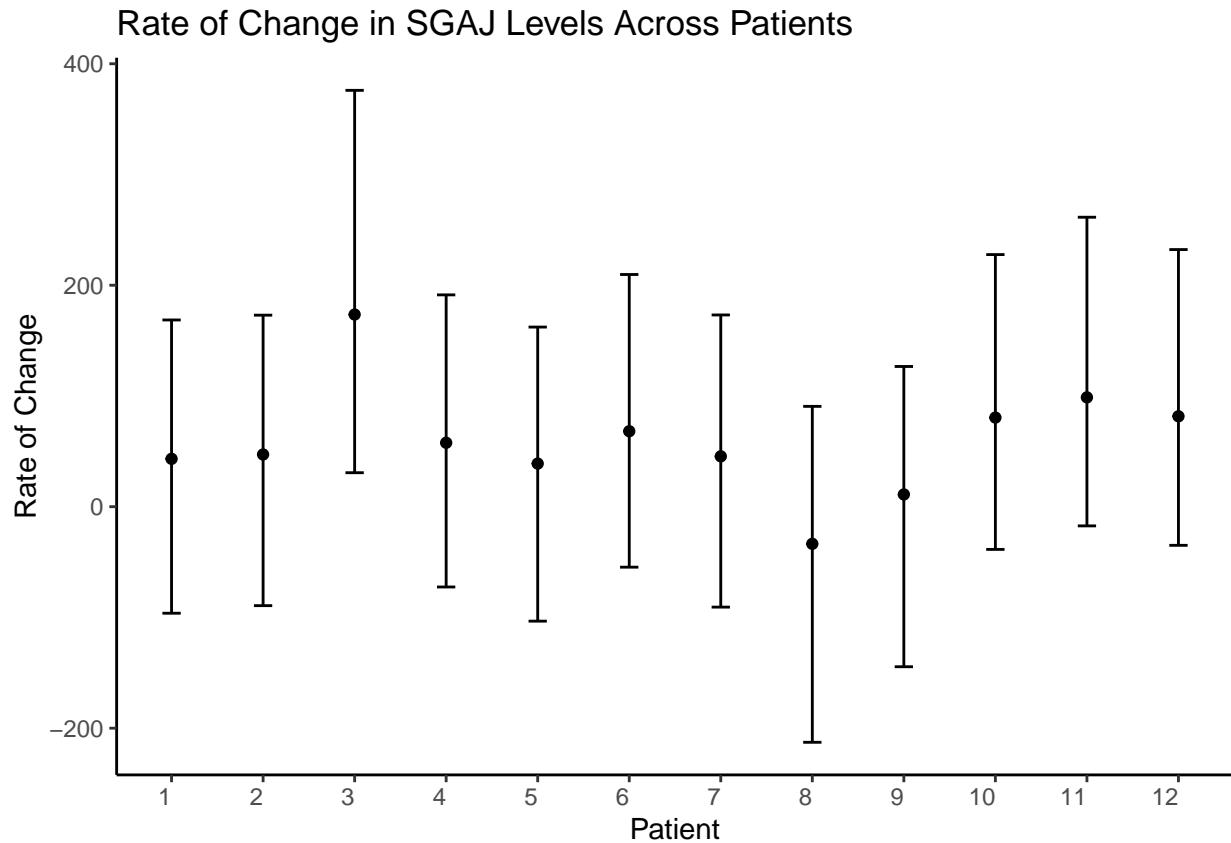
2.5 Visualising rate of change per day per patient

```

# Get dataframe for rate of change variable - showing
rate_of_change_df <- get_posterior_summary(sgaj_draws, c("rate_of_change"))[,c("2.5%", "50%", "97.5%")]
colnames(rate_of_change_df) <- c("lower", "median", "upper")
rate_of_change_df$patient <- as.factor(1:12)

ggplot(rate_of_change_df, aes(x=patient, y=median)) +
  geom_point() + # Add points for the median values
  geom_errorbar(aes(ymin=lower, ymax=upper), width=.2) + # Add error bars for the 95% CI
  theme_classic() +
  labs(y = "Rate of Change", x = "Patient", title = "Rate of Change in SGAJ Levels Across Patients") +
  theme(axis.text.x = element_text(hjust = 1)) # Improve x-axis label readability

```



TODO: show similar kind of plot facetting by age group and gender?

2.6 Model limitations

TODO: enumerate all limitations of the model.

- This is the starting model for a full analysis that would have been conducted in a real publication scenario.

Part 3

```
# Add new row for patient 12 with timepoint = 30 and measured = NA.
sgaj_miss<- rbind(sgaj, c(12, 1, 77, 30, NA))

data_miss <- list(n_patients = length(unique(sgaj_miss$patient)),
                  patient = sgaj_miss$patient,
                  age = sgaj_miss$age,
                  time = sgaj_miss$time,
                  female = sgaj_miss$female,
                  n = nrow(sgaj_miss),
```

```

        y = sgaj_miss$measured
    )

jags_model_miss <- jags.model(textConnection(model),
                                data = data_miss,
                                inits = initial_values,
                                n.chains = 2)

## Compiling model graph
## Resolving undeclared variables
## Allocating nodes
## Graph information:
##   Observed stochastic nodes: 60
##   Unobserved stochastic nodes: 92
##   Total graph size: 659
##
## Initializing model

# Burn-in
update(jags_model_miss, 1000)

pars_miss <- c("mu", "yrep")

# Sampling
samples_miss <- coda.samples(jags_model_miss,
                             variable.names = pars_miss,
                             n.iter = 200000,
                             thin=10)
samples_miss_draws <- as_draws(samples_miss)

# Get last yrep - which is the one for patient 12 at 30 hours.
y.pred <- extract_variable(samples_miss_draws, paste0("yrep[", data_miss$n, "]"))

# Find probability (based on posterior predictive distribution) that it is above 500.
mean(y.pred >= 500)

## [1] 0.315675

# get ids of patient 12 - to use to get yrep.
ids_patient_12 <- rownames(sgaj_miss[sgaj_miss$patient == 12, ])

# Get dataframe with observations only for patient 12
sgaj_miss_patient_12 <- sgaj_miss[sgaj_miss$patient == 12, ] %>%
  mutate(index = 1:6)

# Get summary for mu (fitted mean) and yrep (posterior predictive)
samples_miss_mu_pp <- summary(subset_draws(samples_miss_draws, c(
  paste0("mu[", ids_patient_12, "]"),
  paste0("yrep[", ids_patient_12, "]")))
  ), ~ quantile(.x, probs = c(0.025, 0.5, 0.975))) %>% mutate(
  index = rep(1:nrow(sgaj_miss_patient_12), 2),
  Uncertainty = if_else(

```

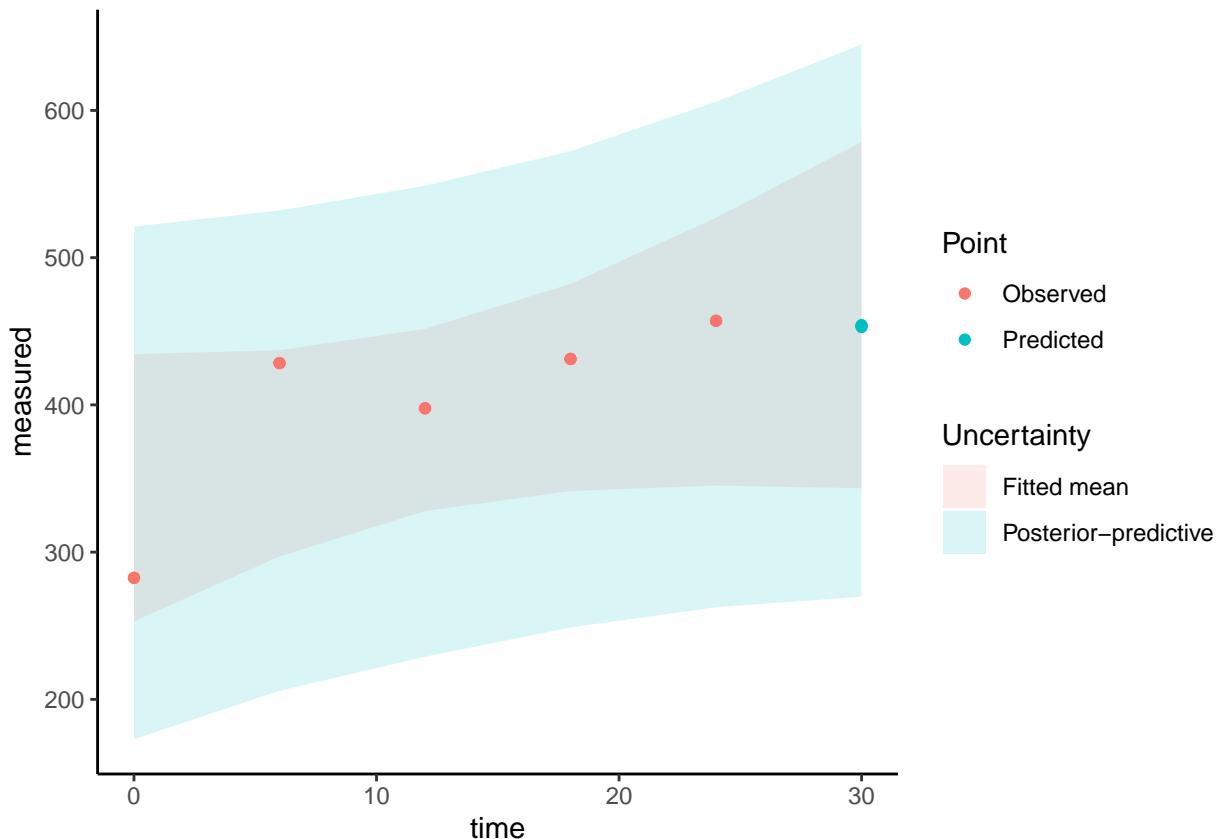
```

    str_detect(variable, "mu"),
    "Fitted mean",
    "Posterior-predictive"
)
)

# Combine datasets
combined_df_patient_12 <-
  samples_miss_mu_pp %>% left_join(sgaj_miss_patient_12, by = "index") %>%
  mutate(
    Point = if_else(is.na(measured), "Predicted", "Observed"),
    measured = if_else(is.na(measured), `50%`, measured)
  )

ggplot(combined_df_patient_12, aes(x=time, y=measured)) + geom_ribbon(aes(ymin=`2.5%`, ymax=`97.5%`, fill=

```



Part 4

Conduct a posterior predictive check for the measurement at 18 hours after SEYAB administration for patient id = 3. #TODO::: getting from the original model (not the one with missing) - think if this is fine

```
# get posterior predictive distribution for patient with id=3. From the original model (not the one with
index_patient_3_time_18 <- rownames(sgaj[sgaj$patient == 3 & sgaj$time==18,])
```

```

yrep_patient_3_time_18 <- extract_variable(sgaj_draws, paste0("yrep[", index_patient_3_time_18, "]"))
summary(yrep_patient_3_time_18)

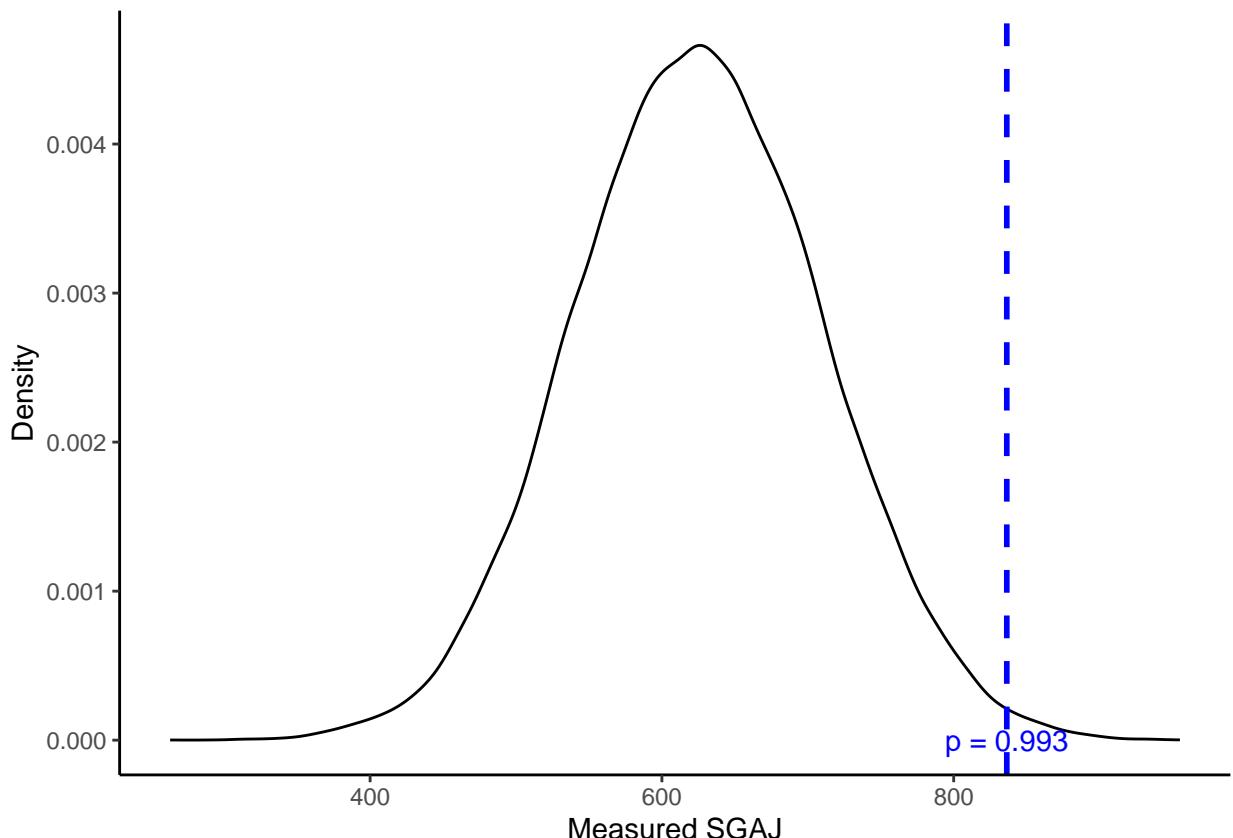
##      value
##  Min.   :262.9
##  1st Qu.:567.4
##  Median :624.8
##  Mean   :625.2
##  3rd Qu.:682.7
##  Max.   :955.1

obs_patient_3_time_18 <- as.numeric(sgaj[index_patient_3_time_18]$measured)

# Calculate the p-value
p_value <- mean(yrep_patient_3_time_18$value < obs_patient_3_time_18)

# Create the plot
yrep_patient_3_time_18 %>%
  ggplot(aes(x = value)) +
  geom_density() + # Plotting the density
  geom_vline(aes(xintercept = obs_patient_3_time_18), color="blue", linetype="dashed", linewidth=1) +
  annotate("text", x = obs_patient_3_time_18, y = 0, label = paste0("p = ", round(p_value, 3)), color = "blue") +
  labs(x = "Measured SGAJ", y = "Density") + # Adding labels
  theme_classic() # Using classic theme

```



Because of the conservatism of p-values the very large p-value that we got, likely indicates that the posterior distribution for patient 3 lacks fit to the data.

TODO: add comment however very high compared to others! perhaps outlier!