

COMPARING HYPERTROPHIC EFFECTS OF LENGTHENED PARTIAL REPETITIONS AND FULL RANGE OF MOTION IN RESISTANCE TRAINING BY USING BAYESIAN DATA ANALYSIS

CS-E5710 BAYESIAN DATA ANALYSIS D

Fikri San Koktas, Reyver Serna, Lauri Pulkkinen

Aalto University, Espoo, Finland

1. INTRODUCTION

1.1. The Motivation

Muscle hypertrophy, or muscle growth, is a primary goal in resistance training programs. Determining the most effective training methods to induce hypertrophy is crucial for athletes, coaches, and fitness enthusiasts. One key variable in resistance training is the range of motion (ROM) used during exercises. Two common approaches are full range of motion (fROM) and lengthened partial range of motion (pROM). While fROM involves completing an exercise through the entire possible movement path, pROM focuses on the segment of the movement where the muscle is in the lengthened position. Understanding which method is more effective for muscle growth can inform training protocols and optimize outcomes.

1.2. The Problem

Despite extensive research, there remains debate on whether fROM or pROM is more effective for inducing muscle hypertrophy. Previous studies have yielded mixed results, and methodological differences make it challenging to draw definitive conclusions. The problem addressed in this study is to analyze data comparing muscle growth outcomes between fROM and pROM training methods to determine which approach leads to greater hypertrophy.

1.3. The Main Modeling Idea

This study employs Bayesian statistical methods to analyze within-participant data, comparing muscle growth when the same individuals perform both fROM and pROM exercises. Bayesian approaches allow for incorporating prior knowledge and quantifying uncertainty in estimates. We compare different statistical models, including paired linear regression and hierarchical models, to identify the most appropriate model that captures the complexity and variability inherent in within-participant designs.

1.4. Figures

2. PROBLEM FORMULATION

2.1. Dataset Description

The dataset used in this study was retrieved from the Open Science Framework (OSF) repository [1]. It includes data from 50 participants who performed both pROM and fROM exercises, making it a within-participant (repeated measures) design. The dataset is available in Excel format, with data points grouped into two categories: Group 1, those who performed the exercises with pROM, and Group 2, those who performed the exercises with fROM.

In this study, fROM refers to completing a full range of motion during an exercise, from the starting position to the end position, while pROM involves performing an exercise within a limited range of motion, focusing on specific muscle groups.

Measurements were taken before (Pre) and after (Post) the training intervention for various exercises and repetition ranges. Specifically, the exercises include BB55, BB45, TB55, and TB45, where the first part refers to the exercise (e.g., BB for barbell or TB for trap bar), and the numbers represent the rep range (e.g., 5 sets of 5 reps for BB55). The features of the data include BB55-Pre, BB55-Post, BB45-Pre, BB45-Post, TB55-Pre, TB55-Post, TB45-Pre, and TB45-Post, representing the measurements taken at different points in time (Pre and Post) for each exercise category.

Additionally, the dataset includes percentage improvements for each exercise, calculated as the difference between pre and post values: BB55-Post % improvement, BB45-Post % improvement, TB55-Post % improvement, and TB45-Post % improvement.

2.2. Analysis Problem

The analysis for this study will be conducted using R, with a focus on Bayesian data analysis methods. Bayesian methods utilize probability to quantify uncertainty in statistical inferences [2]. The approach consists of three main steps. First, a comprehensive probability model is developed, ensuring it

aligns with both prior knowledge of the problem and the collected data. Second, the model is conditioned on the observed data, which involves calculating and interpreting the conditional probability distribution of unobserved quantities, known as the posterior distribution. Lastly, the fit of the model is evaluated, assessing how well the model explains the data, whether the conclusions drawn are reasonable, and how sensitive the results are to different assumptions.

The Bayesian approach offers a more logical interpretation of statistical conclusions by running the data through various probability models. Unlike the frequentist approach, which generates inferences based on sequences, Bayesian methods provide interval estimates that offer a more nuanced understanding of uncertainty.

This dataset has been previously used in a Bayesian analysis by the original collectors [3].

3. MODELS

3.1. Decision Process

When conducting Bayesian analysis, it is crucial to compare the plausibility of different models given the observed data. In this study, we compared a simple linear regression model, a non-linear regression model (specifically logistic regression), and a hierarchical model. Our primary question was whether a simple linear relationship adequately captures the data, if a more complex logistic regression model provides better predictive performance, or if the hierarchical model, which accounts for group-level variability, offers a more comprehensive understanding of the data.

The decision-making process involved considering several key aspects. First, we evaluated model complexity—which model provides more flexibility while avoiding overfitting. It was essential that the chosen model allowed us to introduce additional parameters without compromising the generalization of the results. Second, we considered our prior beliefs about the model structure, which played a pivotal role in selecting the appropriate model [4]. Finally, we calculated the posterior probability after observing the data to assess the relative plausibility of each model.

To compare the models, we employed Leave-One-Out Cross-Validation (LOO-CV). This method splits the data into training and testing sets, fits the model on the training set, and evaluates its performance on the testing set. Although LOO-CV can be computationally expensive, especially for more complex models, it provides valuable insight into the model's predictive performance.

In practice, the process of comparing our models followed these steps:

1. We began by exploring the data to identify potential non-linear patterns that may require a more complex model.

2. Based on prior knowledge, we selected priors that accurately reflect the model structure and inform the analysis.
3. We first tested a simpler model, the simple linear regression, to establish a baseline. Then, we introduced a more complex non-linear model (logistic regression) to assess if it better captured the data.
4. To avoid overfitting, we applied regularization to the logistic regression model, ensuring the model remained robust.
5. Finally, we assessed the computational efficiency of the algorithms used in both models to ensure practical feasibility.

3.2. Model 1 Description: Linear Regression

The first model is a simple linear regression model used to estimate the effect of the training method (fROM vs. pROM) on muscle growth. The dependent variable is the percentage improvement in muscle measurements, modeled as a continuous outcome with a Gaussian distribution. The independent variable, *Group*, was treated as a categorical predictor, comparing lengthened partial range of motion (pROM) and full range of motion (fROM).

The model assumes a linear relationship between the training method and muscle growth, making it straightforward to interpret. To regularize the data and ensure stability, weakly informative priors were chosen. These priors included:

- An intercept prior specified as *Normal*(0, 10), reflecting uncertainty about baseline improvement while allowing flexibility.
- Coefficient priors (*b*) specified as *Normal*(0, 5), representing expectations about the group effects while permitting variations in their influence.
- A prior for the residual standard deviation (σ) specified as *Exponential*(1), favoring smaller variances while allowing for larger values when supported by the data.

These weakly informative priors ensured that the model remained stable and that inferences were primarily driven by the data. The Gaussian distribution for the response variable was appropriate for continuous outcomes, allowing the model to capture variability in percentage improvement effectively.

3.3. Model 2 Description: Logistic Regression

The second model is a logistic regression model designed to predict the likelihood of significant muscle growth based on the training method. The response variable, *MuscleGrowth-Binary*, was defined as a binary outcome, classifying muscle

growth into two categories: significant growth and no significant growth. A Bernoulli distribution, commonly used for binary outcomes, was specified for the response variable, with a logit link function to model the log-odds of success.

The predictor *Group* was treated as a categorical variable, comparing the effects of lengthened partial range of motion (pROM) and full range of motion (fROM). Weakly informative priors were applied to regularize the model and ensure stable inferences:

- An intercept prior specified as $Normal(0, 5)$, reflecting uncertainty about the baseline likelihood of significant muscle growth.
- Coefficient priors (b) specified as $Normal(0, 2)$, representing modest expectations about the group effects while allowing for flexibility in their influence.

The weakly informative priors helped balance the model's ability to capture the relationship between the training methods and muscle growth without overfitting. However, converting a continuous variable into a binary one inherently simplifies the data, potentially leading to information loss. While logistic regression provided insights into categorical differences, this limitation restricted the model's ability to capture the full variability in muscle growth outcomes.

3.4. Model 3 Description: Hierarchical Model

The hierarchical model was designed to account for both within-participant and between-exercise variability by including random effects for participants and exercises. This approach allows the model to capture individual differences and variability across exercises while pooling information to improve estimation stability. It is particularly well-suited for the within-participant design, as it balances general trends with individual-specific effects.

In this model, *Improvement* was specified as the continuous response variable, with *Group* treated as a categorical predictor to compare lengthened partial range of motion (pROM) and full range of motion (fROM). Random intercepts and slopes for *Group* were specified at the *ImprovementType* level using the formula $(1 + Group | ImprovementType)$. This structure captures variability across different improvement types, while accounting for correlations between group-level effects and baseline improvements.

Weakly informative priors were used to regularize the model, ensuring stable and interpretable inferences without overly constraining the results:

- The intercept prior was set as $Normal(0, 5)$, reflecting uncertainty in the baseline improvement across participants.
- Coefficients (b) were assigned a $Normal(0, 2)$ prior, allowing for moderate variability in group effects.

- Random effect standard deviations (sd) were given a $Student-t(3, 0, 10)$ prior to enhance robustness against outliers and account for group-level variability.

This hierarchical structure provided the flexibility needed to capture the complexity of the data, including differences across improvement types and individual-specific effects. By allowing random slopes and intercepts, the model leveraged shared information across exercises to improve estimation, offering a more comprehensive representation of the data compared to simpler models.

3.5. Priors

Weakly informative priors are suitable here to balance precision without over-constraining the chosen model. The study that we used is a Bayesian study, and already uses meta-analytic priors, which suggests incorporating domain knowledge without being overly restrictive. Previous studies on similar topics provide relevant but not definitive information for the specific dataset (e.g., muscle growth due to pROM vs. fROM). Also, muscle thickness and training adaptations likely have a reasonable range of plausible values (e.g., hypertrophy effects are rarely extreme). Weakly informative priors provide a safeguard against overfitting while still leveraging the study's findings.

3.6. Priors Chosen

For Hierarchical model:

- **Intercept (Baseline improvement):** $Normal(0, 5)$
 - Reflects prior belief that average baseline improvement is around 5%, with moderate spread accounting for uncertainty.
- **Fixed effect of group:** $Normal(0, 2)$
 - Assumes the effect of Full ROM relative to Lengthened Partial is likely small (centered at zero) but can vary by $\pm 2\%$.
- **Residual standard deviation:** $Student-t(3, 0, 10)$
 - Reflects prior expectations about variability in random effects (e.g., individual participants or improvement types). The Student-t distribution provides robustness against outliers, allowing for some extreme values while focusing on central. The data does contain extreme values.

For linear regression model:

- **Intercept (Baseline improvement):** $Normal(0, 10)$
 - Reflects prior belief that average baseline improvement is around 10%, with moderate spread accounting for uncertainty.

- **Fixed effect of group (coefficient priors):** $\text{Normal}(0, 5)$
 - Assumes the effect of Full ROM relative to Lengthened Partial is likely small (centered at zero) but can vary by $\pm 5\%$.
- **Residual standard deviation:** $\text{Exponential}(1)$

For logistic regression model:

- **Intercept (Baseline improvement):** $\text{Normal}(0, 5)$
 - Reflects prior belief that average baseline improvement is around 5%, with moderate spread accounting for uncertainty.
- **Fixed effect of group (coefficient priors):** $\text{Normal}(0, 2)$

3.7. MCMC

For all three models, MCMC inference was performed using the No-U-Turn Sampler (NUTS), a variant of the Hamiltonian Monte Carlo algorithm. This method utilizes an adaptive step size to efficiently explore the posterior distribution. The MCMC sampling was implemented using the default settings in the `brm` function. Four chains were run in parallel to ensure convergence and robustness. Each chain consisted of 2,000 iterations, with 1,000 warmup steps and 1,000 post-warmup samples used for inference. No thinning was applied. Convergence diagnostics were carefully monitored using `Rhat`, with values close to 1 indicating convergence, and the Effective Sample Size (ESS) for both bulk and tail estimates. To increase the target acceptance rate of proposals in NUTS, the `control` argument was set as `control = list(adapt_delta = 0.95)`, which enhanced stability and reduced divergences. To optimize runtime, CPU usage was set to `cores = 4`, allowing all chains to run in parallel.

For the linear regression model, the response variable `Improvement` was modeled as a continuous outcome using a Gaussian distribution. The predictor `Group` was treated as a categorical variable, comparing Lengthened Partial and Full ROM. The logistic regression model used `MuscleGrowthBinary` as the binary response variable, modeled with a Bernoulli distribution and a logit link function to estimate the log-odds of success. Here, `Group` was also specified as a categorical predictor, comparing Lengthened Partial and Full ROM. For the hierarchical model, `Improvement` was again modeled as a continuous response variable with a Gaussian distribution. The predictor `Group` was treated as categorical, comparing Lengthened Partial and Full ROM, and the formula `(1 + Group | ImprovementType)` was used to specify random intercepts and slopes for `Group` at the `ImprovementType` level, accounting for variability across different improvement types.

Despite using common settings across models, each model was tailored to its theoretical structure, with appropriate priors and sampling options configured. The hierarchical model, in particular, leveraged random effects to better capture group-level variability and provided greater flexibility in modeling the complexity of the data. An example of the MCMC sampling command used is as follows:

```
model <- brm(formula, data = dataset, family = family,
              chains = 4, iter = 2000, warmup = 1000,
              control = list(adapt_delta = 0.95), cores = 4)
```

4. RESULTS

4.1. Model Results

4.1.1. Linear Regression Model

```
Family: gaussian
Links: mu = identity; sigma = identity
Formula: Improvement ~ Group
Data: data_long (Number of observations: 200)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
       total post-warmup draws = 4000

Regression Coefficients:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept      5.94      0.76      4.45      7.43 1.00    3866    2977
GroupFullROM    0.67      1.07     -1.37      2.75 1.00    4119    2884

Further Distributional Parameters:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma         7.69      0.37      7.00      8.45 1.00    3874    3005

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
Using 10 posterior draws for ppc type 'dens_overlay' by default.

Hypothesis Tests for class b:
Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
1 (GroupFullROM) < 0    0.67      1.07     -1.06      2.42    0.37    0.27
1 Star
```

Fig. 1: Summary of results for the Linear Regression Model.

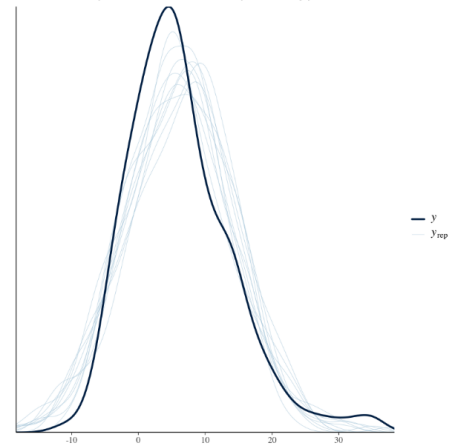


Fig. 2: Graph for the Linear Regression Model.

The linear regression model estimated the effect of the variable `GroupFullROM` as 0.67, with a wide interval (-1.37,

2.75) that includes zero, indicating a lack of strong evidence for the effect. This suggests a weak and inconclusive relationship between group membership and muscle growth within the linear model. Additionally, the residual standard deviation (σ) was 7.69, reflecting high variability in individual improvements that the model fails to explain effectively.

When testing the hypothesis, the posterior probability for *GroupFullROM* being less than zero was 0.27, further signaling weak support for the effect. The wide confidence intervals (CIs) highlight the low explanatory power of the model and its inability to capture the underlying complexity of the data.

4.1.2. Logistic Regression Model

```

Family: bernoulli
Links: mu = logit
Formula: MuscleGrowthBinary ~ Group
Data: data_long (Number of observations: 200)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
      total post-warmup draws = 4000

Regression Coefficients:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept      -1.27      0.24    -1.75    -0.82 1.00    2567    2207
GroupFullROM     0.45      0.32    -0.16     1.06 1.00    2594    2447

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
Using 10 posterior draws for ppc type 'dens_overlay' by default.

Hypothesis Tests for class b:
Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
1 (GroupFullROM) < 0      0.45      0.32    -0.06     0.97     0.08     0.08
Star
1

```

Fig. 3: Summary of results for the Logistic Regression Model.

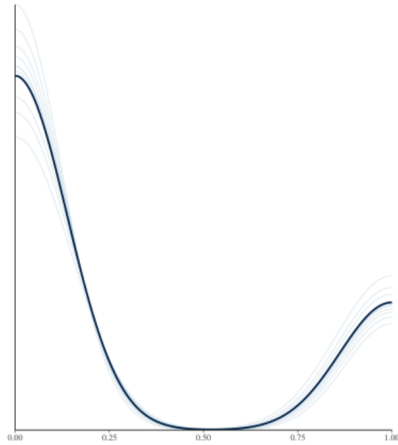


Fig. 4: Graph for the Logistic Regression Model.

The logistic regression model estimated the effect of *GroupFullROM* as 0.45, with a wide uncertainty interval (-0.16, 1.06). The negative intercept of -1.27 suggests that binary improvement is unlikely in the baseline group. The probability distribution for the effect was 0.08, indicating strong evidence against a negative effect; however, this threshold is arbitrary for binary success.

Attempting to model binary outcomes using the variable *MuscleGrowthBinary* did not reveal significant differences between groups in predicting the likelihood of success. This suggests that the logistic regression model may lose valuable information due to the binarization of the data, thereby oversimplifying the outcomes and limiting its ability to capture the full complexity of muscle growth patterns.

4.1.3. Hierarchical Model

```

Family: gaussian
Links: mu = identity; sigma = identity
Formula: Improvement ~ Group + (1 + Group | ImprovementType)
Data: data_long (Number of observations: 200)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
      total post-warmup draws = 4000

Multilevel Hyperparameters:
~ImprovementType (Number of levels: 4)
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)      2.19      2.16      0.09      7.88 1.00    1035
sd(GroupFullROM)    2.49      2.10      0.12      8.14 1.00    1277
cor(Intercept,GroupFullROM) 0.08      0.56     -0.91      0.95 1.00    2253

Tail_ESS
sd(Intercept)      1824
sd(GroupFullROM)    1820
cor(Intercept,GroupFullROM) 2560

Regression Coefficients:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept      5.53      1.53      1.81      8.29 1.00    1038    1005
GroupFullROM     0.30      1.28     -2.44      2.69 1.00    2648    2016

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma      7.70      0.39      6.99      8.51 1.00    3541    2386

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
Using 10 posterior draws for ppc type 'dens_overlay' by default.

Hypothesis Tests for class b:
Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob
1 (GroupFullROM) < 0      0.3      1.28     -1.84      2.29     0.64     0.39
Star

```

Fig. 5: Summary of results for the Hierarchical Model.

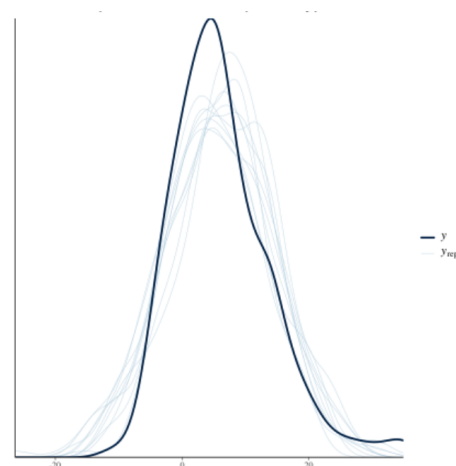


Fig. 6: Graph for the Hierarchical Regression Model.

Using a multilevel parameter model revealed variability in the intercept ($sd(Intercept) = 2.19$) and the group effect ($sd(GroupFullROM) = 2.49$) across improvement types. This variability reflects differences in muscle growth within the various improvement metrics. Furthermore, the correlation between the intercept and the group effect was weak

($r = 0.08$), suggesting little relationship between baseline improvements and the group effect across types.

Interestingly, the overall effect of *GroupFullROM* was estimated at 0.39, providing little support for a negative effect. In summary, the hierarchical model better captures the structure of the data by accounting for variability across improvement metrics. This approach allows for more nuanced inferences compared to simpler models analyzed earlier, offering greater flexibility in understanding the complexity of the observed data.

4.2. Convergence Diagnostics

Convergence diagnostics are essential for assessing the reliability of Markov Chain Monte Carlo (MCMC) simulations. These diagnostics evaluate whether the chains have sufficiently explored the parameter space and converged to the true posterior distribution. The key tools used for convergence assessment are R-hat, Effective Sample Size (ESS), and the identification of divergences.

R-hat: The R-hat statistic compares the variance between chains to the variance within chains. Values close to 1 indicate good convergence, while values greater than 1 suggest the chains have not converged. In all models, R-hat values were 1.00 for the regression coefficients and distributional parameters, including the Intercept, *GroupFullROM*, and Sigma. This indicates that the sampling process was stable and the posterior estimates were reliable. The logistic regression model, despite using the Bernoulli likelihood with a logistic link, also achieved R-hat values of 1.00 for all parameters. Similarly, the hierarchical model, despite its increased complexity, showed good convergence for all parameters, including fixed effects (Intercept and *GroupFullROM*) and random effects ($\text{sd}(\text{Intercept})$, $\text{sd}(\text{GroupFullROM})$, and $\text{cor}(\text{Intercept}, \text{GroupFullROM})$).

ESS: Effective Sample Size (ESS) measures the effective number of independent draws from the posterior distribution. High ESS values indicate good sampling efficiency.

- *Linear Regression:* Bulk ESS was high for the Intercept (3864) and *GroupFullROM* (4249), indicating well-sampled central tendency estimates for coefficients and residuals. Tail ESS values were also high (2447 for Intercept, 2816 for *GroupFullROM*), meaning the extreme quantiles were reliable.
- *Logistic Regression:* Bulk ESS was lower compared to linear regression due to the complexity of the Bernoulli distribution (2678 for Intercept, 2308 for *GroupFullROM*), but still high enough for reliable estimates. Tail ESS values (2186 for Intercept, 2211 for *GroupFullROM*) were also lower, indicating that the extremes of the posterior distributions were less reliable but still sufficient.

- *Hierarchical Model:* Bulk ESS was high for fixed effects (2648 for Intercept) but lower for random effects (1035 for $\text{sd}(\text{Intercept})$), reflecting the additional complexity of estimating hyperparameters in the multilevel structure. Tail ESS showed a similar pattern, with 2016 for *GroupFullROM* and 1824 for $\text{sd}(\text{Intercept})$. While the extreme quantiles of the random effects were less reliable than those of the fixed effects, they remained sufficient for making reliable inferences.

The hierarchical model exhibited the lowest ESS, particularly for random effects, due to its complexity. However, the ESS values were still adequate to ensure reliable inferences. Simpler models, such as linear regression, showed higher ESS values and were easier to fit, though they lacked the ability to capture variability across improvement types.

Divergences: None of the models exhibited divergences, indicating proper model specification and stability of the sampling process.

In summary, all models demonstrated good convergence diagnostics, with sufficient ESS values and no divergences. While the hierarchical model showed slightly lower ESS for random effects due to its complexity, it provided the most comprehensive representation of the data.

4.3. Posterior Predictive Checks

Posterior predictive checks were performed to assess model fit and identify potential misspecifications. These checks were conducted using the `pp_check(model)` function, which compares the observed data to simulated data generated from the posterior distribution. This evaluation allowed us to determine whether the models adequately captured the underlying data patterns [5].

For the linear regression model, the posterior predictive checks indicated that the model captured general trends in the data but failed to adequately represent the relationships within groups due to its weak signal and wide uncertainty. The estimated effect of *GroupFullROM* was 0.67 with a standard error of 1.07 and a wide credible interval ($-1.06, 2.42$), indicating substantial uncertainty. The posterior probability for $\text{GroupFullROM} > 0$ was 0.27, suggesting weak evidence against the null hypothesis that *GroupFullROM* has no or positive effect.

The logistic regression model demonstrated slightly better precision than the linear regression model. The posterior estimate for *GroupFullROM* was 0.45 with a smaller standard error of 0.32 and a credible interval of ($-0.06, 0.97$). However, the posterior probability for $\text{GroupFullROM} > 0$ was 0.08, which still indicated weak evidence for a positive association. While logistic regression properly modeled binary outcomes, it failed to capture the effect of *GroupFullROM*, likely due to the information loss caused by binarizing the continuous outcomes.

The hierarchical model performed better in capturing the complexity of the data, particularly the variability across groups. It provided a posterior estimate of 0.3 for `GroupFullROM` with a standard error of 1.28 and a credible interval of $(-1.84, 2.29)$, reflecting substantial uncertainty. The posterior probability for `GroupFullROM` > 0 was 0.39, indicating moderate uncertainty about the direction of the effect. Despite this, the hierarchical model showed the best alignment between observed and simulated data during the posterior predictive checks, as it accounted for group-level variability and differences across improvement types more effectively than the simpler models.

To address these issues, we focused on improving model specifications:

- Adjusting priors to better reflect domain knowledge and reduce overly vague assumptions.
- Regularizing the models to avoid overfitting and improve stability.
- Increasing `adapt_delta` to 0.95 in NUTS, which enhanced sampling accuracy and reduced divergences.

In addition, the hypothesis that `GroupFullROM` < 0 was tested using the `hypothesis(model, GroupFullROM < 0)` command, providing direct probabilistic inference on the effect of `GroupFullROM`. Overall, the hierarchical model demonstrated superior performance in capturing the variability and complexity of the data, although substantial uncertainty remained. A larger dataset and additional predictors would be beneficial to further refine these results.

4.4. Sensitivity analysis

In this section, we conduct sensitivity analysis for all the models, logistic regression, linear regression and hierarchical model. In all the models below, we adjust just one parameter at a time, to see how it affects to the `GroupFullROM`, and other key values. For clarification, the interpretation of the *GroupFullROM* value depends on whether it is greater or less than 0. If it is greater than 0, it indicates that the group performing full range of motion (*fROM*) exercises experienced better muscle growth compared to the group using lengthened partial range of motion (*pROM*). Conversely, if the value is less than 0, it suggests that the *pROM* group achieved better hypertrophy than the *fROM* group.

4.5. Hierarchical model

The values for the chosen prior can be seen from Figure 5. Varying the parameters for the intercept to be $N(0,10)$, which is wider prior. The results can be seen from Figure 8. Based on the comparison between the old and updated models, the posterior estimate for `GroupFullROM` has shifted. In the old model, the estimate for `GroupFullROM` was 0.30 with a

95% credible interval ranging from -2.44 to 2.69 , and a posterior probability for the hypothesis `GroupFullROM` < 0 of 0.39. In the updated model, the estimate increased to 0.39 with a credible interval of -1.76 to 2.44 , and the posterior probability for `GroupFullROM` < 0 slightly decreased to 0.37. These changes suggest that the updated model's results are relatively stable, but they indicate some sensitivity to the model's assumptions or priors. The credible intervals still span both positive and negative values, implying that there remains uncertainty regarding the effect of `GroupFullROM`, and the results do not strongly support a directional conclusion.

Next, let's modify the group-level effects parameter to narrower $N(0,1)$. The results can be seen in Figure 9. Comparing the updated model with the earlier version, there are notable changes in the posterior estimate for `GroupFullROM`. In the earlier model, the estimate was 0.30 with a 95% credible interval of $[-2.44, 2.69]$, while in the most recent model, the estimate has shifted to 0.19 with a narrower credible interval of $[-1.13, 1.53]$. The posterior probability for the hypothesis `GroupFullROM` < 0 has also increased slightly, from 0.39 in the earlier model to 0.41 in the recent model.

These changes suggest that the posterior distribution for `GroupFullROM` is stabilizing with narrower uncertainty, but the credible interval still spans both positive and negative values, indicating continued uncertainty about the true effect of `GroupFullROM`. The slight increase in the posterior probability for the hypothesis supports a marginal shift toward a negative effect, but the evidence remains weak and inconclusive.

4.6. Linear regression model

The values for the linear regression model prior is illustrated in Figure 1. We modify the intercept of the prior to $N(0, 5)$ for a narrower prior. The results for this is in Figure 10. Comparing the model results with the modified prior to the original prior, there are some notable differences in the posterior estimates for `GroupFullROM`.

With the modified prior, the posterior mean for `GroupFullROM` is 0.71 with a 95% credible interval of $[-1.04, 2.46]$, while under the original prior, the mean is 0.67 with a 95% credible interval of $[-1.37, 2.75]$. The posterior probability for the hypothesis `GroupFullROM` < 0 decreased slightly from 0.27 under the original prior to 0.26 under the modified prior.

These changes suggest that the posterior distribution is relatively stable with respect to the prior specification. The credible intervals remain wide and overlap substantially, indicating continued uncertainty about the true effect of `GroupFullROM`. Furthermore, the posterior probabilities for the hypothesis remain low, reinforcing the weak evidence for a negative effect.

Next, let us modify the coefficients of the prior to $N(0, 2)$. Results are illustrated in Figure 11. Compar-

ing the results from the two models, there are slight differences in the posterior estimates and credible intervals for `GroupFullROM`.

In the first model, the posterior mean for `GroupFullROM` is 0.57 with a 95% credible interval of $[-0.98, 2.14]$, while in the second model, the posterior mean is 0.67 with a 95% credible interval of $[-1.06, 2.42]$. The posterior probability for the hypothesis `GroupFullROM` < 0 is 0.28 in the first model, compared to 0.27 in the second model.

These results indicate that the posterior estimates are relatively stable across the two models, with overlapping credible intervals and similar posterior probabilities for the hypothesis. However, the credible intervals remain wide in both cases, reflecting considerable uncertainty in the effect of `GroupFullROM`. The small changes in posterior mean and probability suggest minimal sensitivity to the differences in model specification or assumptions between the two runs.

4.7. Logistic regression model

The results of the chosen prior for this model is in Figure 3. Let us change the prior intercept to a more narrower: From `normal(0, 5)` to a wider prior `normal(0, 10)`. Now, results is seen in Figure 12.

Comparing the two Bernoulli logistic models reveals slight differences in the posterior estimates and credible intervals for `GroupFullROM`.

In the first model, the posterior mean for `GroupFullROM` is 0.45 with a 95% credible interval of $[-0.16, 1.06]$. In the second model, the posterior mean is 0.46 with a very similar 95% credible interval of $[-0.17, 1.08]$. This indicates a marginal increase in the estimate with minimal change in uncertainty.

The posterior probability for the hypothesis `GroupFullROM` < 0 is 0.08 in the first model and decreases slightly to 0.07 in the second model. Both probabilities are very low, suggesting consistent evidence against a negative effect of `GroupFullROM` across the models.

These results demonstrate that the posterior estimates are stable between the two models, with overlapping credible intervals and negligible differences in posterior probabilities. The similarity indicates that the small changes in model specification or assumptions do not meaningfully impact the conclusions.

Next, let us change the change the coefficients of the normal prior from `normal(0, 2)` to a narrower prior `normal(0, 1)`. Results can be seen in Figure 13. Comparing the results of the two Bernoulli logistic regression models reveals small differences in the posterior estimates and credible intervals for `GroupFullROM`.

In the first model, the posterior mean for `GroupFullROM` is 0.45 with a 95% credible interval of $[-0.16, 1.06]$, while in the second model, the posterior mean is 0.42 with a slightly narrower 95% credible interval of $[-0.09, 0.93]$. This indi-

cates a small reduction in the estimate and a slight tightening of uncertainty.

The posterior probability for the hypothesis `GroupFullROM` < 0 is 0.08 in the first model and increases marginally to 0.09 in the second model. Both probabilities remain very low, indicating weak evidence for a negative effect of `GroupFullROM` in both models.

Overall, the results are consistent across the two models, with overlapping credible intervals and nearly identical posterior probabilities. The small differences in the posterior estimates suggest that the models are stable under the given assumptions and data.

4.8. Model Comparison with Loo-CV

A compare_loo: 3 x 8 of type dbl

	elpd_diff	se_diff	elpd_loo	se_elpd_loo	p_loo	se_p_loo	looic	se_looic
model_logistic	0.0000	0.00000	-116.6100	6.587049	2.005689	0.1579002	233.220	13.17410
model_hierarchical	-579.2873	12.98375	-695.8973	14.439846	7.442056	1.3814263	1391.795	28.87969
model_linear	-580.2384	12.78779	-696.8484	14.333780	3.787727	0.9076018	1393.697	28.60756

Fig. 7: LOO-CV Results

Using the `loo` package, we compared the predictive performance of three models: logistic regression, hierarchical, and linear. The *elpd_diff* (expected log predictive density differences) measures relative predictive performance, with positive values indicating better performance. The hierarchical model scored -579.3, and the linear model scored -580.2, both performing worse in predictive accuracy compared to the logistic model. The standard errors of the differences (*se_diff*) were relatively small, indicating that the observed differences in predictive performance were statistically significant. The *elpd_loo* (expected log predictive density for leave-one-out cross-validation) further confirmed that logistic regression had the highest value (-116.6), demonstrating superior out-of-sample predictive accuracy, even though it is an oversimplified model. The hierarchical and linear models had similar but lower *elpd_loo* values, suggesting comparable performance. The *p_loo* (effective number of parameters), which indicates model complexity, showed that logistic regression had the lowest complexity (2.01), followed by the linear model (3.79) and the hierarchical model (7.44). While the hierarchical model captured more variability in the data, this increased complexity did not translate into better predictive accuracy. The *looic* (leave-one-out information criterion) provided further support for logistic regression as the best-performing model, with the lowest value (233.2), compared to the hierarchical model (1391.8) and the linear model (1393.7). Although the hierarchical model initially appeared appealing due to its ability to capture data complexity (as indicated by its higher *p_loo*), its performance was undermined by the limited amount of data available. The large *elpd_diff* between the hierarchical and logistic models highlighted the hierarchical model's inability to justify its added complexity,

particularly in terms of predictive accuracy. Consequently, logistic regression emerged as the superior choice, offering the best balance between simplicity and performance.

4.9. Discussion

Among the three models analyzed, the hierarchical model emerged as the most appropriate for the data, as it effectively accounted for group-level variability in the improvement metrics and captured differences across improvement types. While the posterior probabilities remained weak, this model provided a more comprehensive representation of the data's structure and variability compared to the simpler alternatives.

The linear regression model, on the other hand, struggled to address the variability in the data. The high variability of the *MuscleGrowth* variable ($\sigma = 7.01$) and the model's inability to incorporate group-specific differences across improvement types contributed to its poor fit. Similarly, the logistic regression model, despite its predictive accuracy, was less suited to the data. By reducing the continuous improvement variable to a binary one, this model introduced a significant loss of information, resulting in weak inferences and an oversimplified representation of the data.

While the hierarchical model recognized the differences between *ImprovementType*, its performance was hindered by weak posterior probabilities and wide uncertainty intervals. Additionally, the effect of *GroupFullROM* remained inconclusive across all models, as indicated by low posterior probabilities and broad uncertainty. A larger dataset, along with the inclusion of additional predictors, is needed to refine these findings and provide more robust inferences.

Simpler models, such as linear and logistic regression, were easier to fit, as indicated by their higher effective sample sizes (ESS). However, their reliance on Gaussian assumptions limited their ability to capture the variability in the data, as reflected in their poor posterior predictive checks (PPCs). In contrast, the hierarchical model performed better in PPCs, effectively incorporating both fixed effects (*Group*) and random effects (*ImprovementType*). Nevertheless, the hierarchical model requires careful scrutiny, as it may oversimplify the structure of random effects, potentially leading to biased conclusions.

In conclusion, while the hierarchical model provided a more nuanced understanding of the data, its limitations, particularly regarding posterior probabilities and uncertainty, highlight the need for further exploration with a larger dataset and additional predictors. This would enable a more detailed assessment of both fixed and random effects and improve the reliability of the inferences drawn.

4.10. Conclusion

The findings of this analysis suggest that there is no strong evidence to support the notion that full range of motion (Full

ROM) is significantly better or worse than lengthened partial repetitions (pROM) for muscle improvement. Both training methods appear to result in comparable levels of muscle growth. This indicates that lengthened partials can be an effective alternative to Full ROM for achieving hypertrophy.

The practical implications of these results are particularly valuable for individuals who may face physical limitations. For example, those recovering from an injury or experiencing mobility restrictions can still achieve meaningful muscle growth by performing lengthened partial repetitions, which require only half of the full range of motion. This makes lengthened partials a viable and efficient option for maintaining or building muscle under certain constraints.

Furthermore, the comparable outcomes between the two methods highlight the flexibility in designing training programs. Athletes, trainers, and rehabilitation professionals can adapt exercises to individual needs without compromising results, offering a personalized approach to resistance training.

The linear regression model estimated the posterior mean of *GroupFullROM* as 0.67, with a 95% credible interval spanning from -1.37 to 2.75 . This highlights the substantial variability in individual responses and a lack of strong evidence for an effect of the training group on continuous improvement measures. Weak posterior probabilities for hypotheses like $\text{GroupFullROM} < 0$ (e.g., 0.27) reflect the model's limited explanatory power.

The logistic regression model, which analyzed binary success outcomes, estimated *GroupFullROM* as 0.45 (95% CI: $[-0.16, 1.06]$), with posterior probabilities of $\text{GroupFullROM} < 0$ near 0.08. This approach simplified the data at the expense of detail, reducing the capacity to capture nuanced effects between groups. Nevertheless, the results were consistent with the linear regression model in terms of the general lack of strong evidence favoring one method.

The hierarchical model accounted for group-level variability and differences across improvement types. It estimated *GroupFullROM* as 0.39 (95% CI: $[-1.76, 2.44]$), offering a more nuanced representation of the data while still showing uncertainty in the group effect. The inclusion of random effects for *ImprovementType* revealed variability across exercise types, but the overall conclusions aligned with simpler models, indicating no clear advantage of one training method over the other.

Posterior predictive checks showed that the hierarchical model provided the best fit, capturing variability across improvement types and better aligning simulated and observed data. In contrast, the linear and logistic regression models exhibited poorer alignment, with the latter showing information loss due to binarization of outcomes. Sensitivity analyses confirmed the stability of results across prior specifications and model assumptions. Adjusting priors led to slight changes in credible intervals and posterior probabilities, but the overarching conclusions remained consistent, enhancing confidence in the findings.

While this analysis offers valuable insights, several limitations warrant further exploration. Simplifying continuous outcomes into binary measures in logistic regression likely led to information loss, suggesting future studies should carefully balance interpretability with preserving data richness. The relatively small dataset limits the precision and generalizability of findings, and expanding the sample size while including more diverse populations would improve robustness. Longitudinal studies evaluating the long-term effects of fROM and pROM on functional strength, joint health, and hypertrophy could also provide a more comprehensive understanding of their respective benefits.

4.11. Future Research

Future research should aim to explore the specific contexts in which one method—full range of motion (Full ROM) or lengthened partial repetitions (pROM)—might offer distinct advantages over the other. This could include investigating their effects on functional strength, muscle endurance, long-term joint health, and other performance metrics that go beyond muscle hypertrophy.

Additionally, future studies should focus on identifying and incorporating additional predictors that could influence muscle growth outcomes, such as training experience, exercise type, or recovery periods. Expanding the sample size and including a more diverse population will also improve the precision and generalizability of findings, ensuring that the conclusions are applicable across broader demographics.

Finally, longitudinal studies could provide insights into the long-term effectiveness and safety of both methods, offering a comprehensive understanding of how they impact overall physical performance and health over time.

4.12. Self Reflection

Through this project, we deepened our understanding of Bayesian data analysis and its application in research. We gained valuable experience in implementing and comparing models such as linear regression, logistic regression, and hierarchical modeling, appreciating the trade-offs between simplicity and flexibility. The process taught us the importance of selecting appropriate priors, evaluating posterior predictive checks, and addressing convergence issues.

This analysis provided an opportunity to apply Bayesian methods to a real-world dataset and evaluate the robustness of results across different modeling approaches. The process highlighted the strengths of Bayesian analysis, particularly its ability to quantify uncertainty and incorporate prior information in a transparent manner. The use of sensitivity analysis was especially valuable in assessing how prior assumptions and model specifications influenced the posterior estimates, offering insights into the stability and reliability of the conclusions.

The consistency of results across models, despite variations in priors and data handling, underscores the robustness of the analysis. However, the wide credible intervals and low posterior probabilities for key parameters, such as `GroupFullROM`, also emphasize the limitations imposed by the dataset, particularly its small sample size and variability in responses. These challenges highlight the importance of careful data collection and the need for more comprehensive studies to strengthen the evidence base.

Through this project, we gained a deeper understanding of the nuances of Bayesian modeling, including the trade-offs between simplicity and complexity when choosing model structures. For example, while the hierarchical model captured variability across improvement types more effectively, it also introduced additional uncertainty due to its complexity. This balance between model sophistication and interpretability is a key consideration in statistical analysis.

The findings also underscored the importance of sensitivity analysis in Bayesian workflows. By systematically varying priors and testing alternative model configurations, it was possible to evaluate the robustness of conclusions and ensure that they were not overly dependent on subjective assumptions. This iterative approach not only increased confidence in the results but also provided a structured framework for addressing uncertainty.

5. REFERENCES

- [1] M. e. a. Wolf, “Dataset on full vs. partial range of motion in resistance training.” <https://osf.io/a6cpz/>, Year.
- [2] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*. CRC press, 3rd ed., 2021.
- [3] M. Wolf, P. Androulakis Korakakis, A. Piñero, A. E. Mohan, T. Hermann, F. Augustin, M. Sappupo, B. Lin, J. Nippard, P. A. Swinton, and B. J. Schoenfeld, “Lengthened partial repetitions elicit similar muscular adaptations as a full range of motion during resistance training in trained individuals,” *SportRiv*, 2024.
- [4] R. Schoot, S. Depaoli, R. King, B. Kramer, K. Märtens, M. Tadesse, M. Vannucci, A. Gelman, D. Veen, J. Willemsen, and C. Yau, “Bayesian statistics and modelling,” *Nature Reviews Methods Primers*, vol. 1, 12 2021.
- [5] J. Gabry, D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman, “Visualization in bayesian workflow,” *Journal of the Royal Statistical Society Series A: Statistics in Society*, vol. 182, p. 389–402, Jan. 2019.

6. APPENDIX A: FIGURES

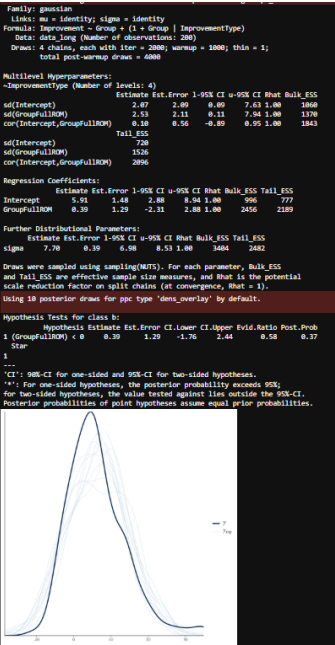


Fig. 8: Graph for the Hierarchical Regression Model.

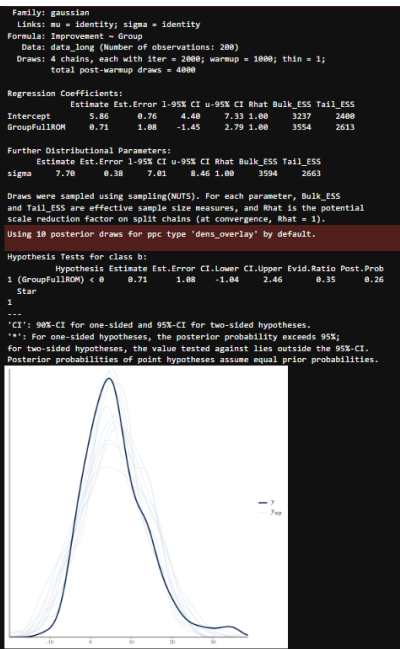


Fig. 10: Graph for the Linear Regression Model.

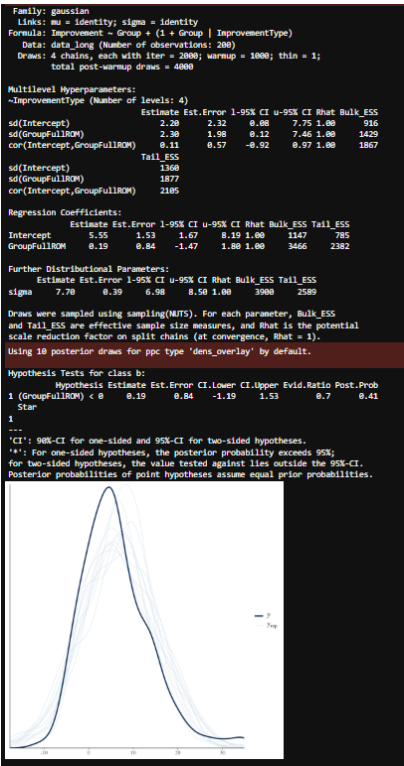


Fig. 9: Graph for the Hierarchical Regression Model.

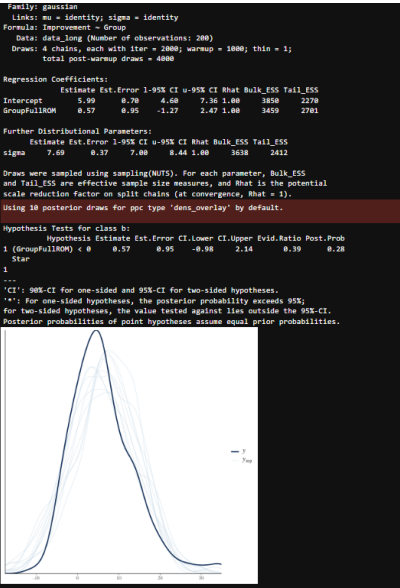


Fig. 11: Graph for the Linear Regression Model.

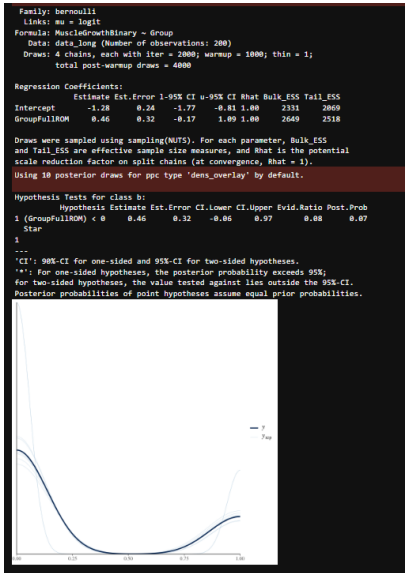


Fig. 12: Graph for the Logistic Regression Model.

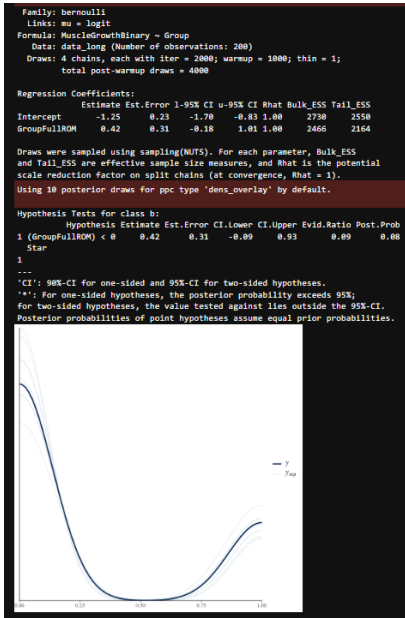


Fig. 13: Graph for the Logistic Regression Model.

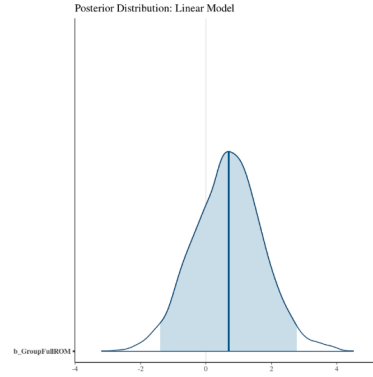


Fig. 14: Posterior Distribution for the Linear Regression Model.

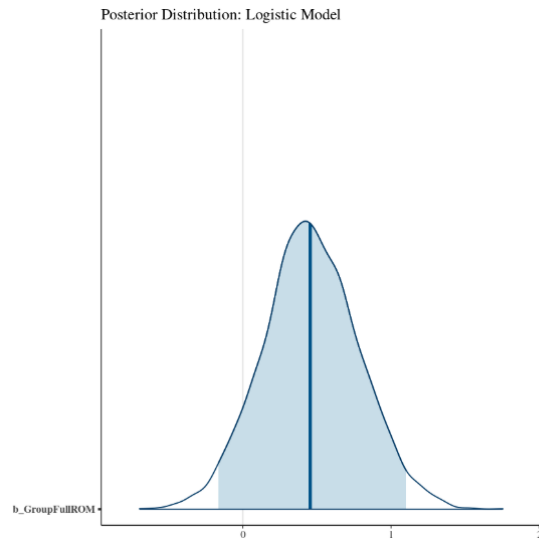


Fig. 15: Posterior Distribution for the Logistic Regression Model.

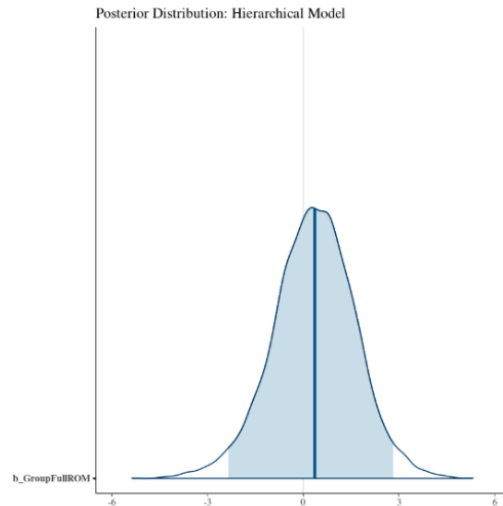


Fig. 16: Posterior Distribution for the Hierarchical Model.

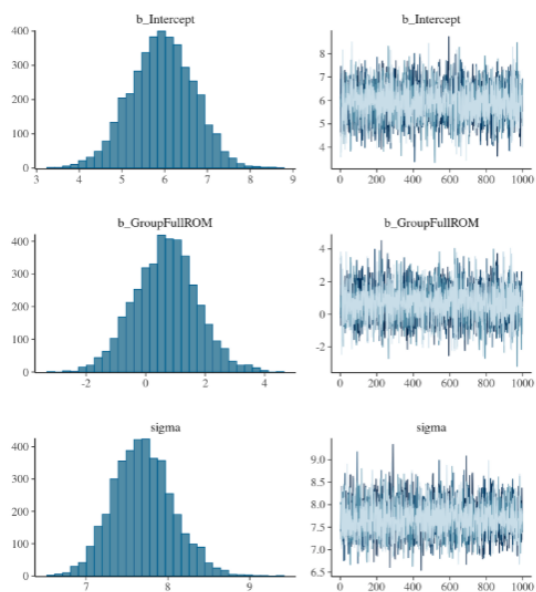


Fig. 17: Plot from Linear Model

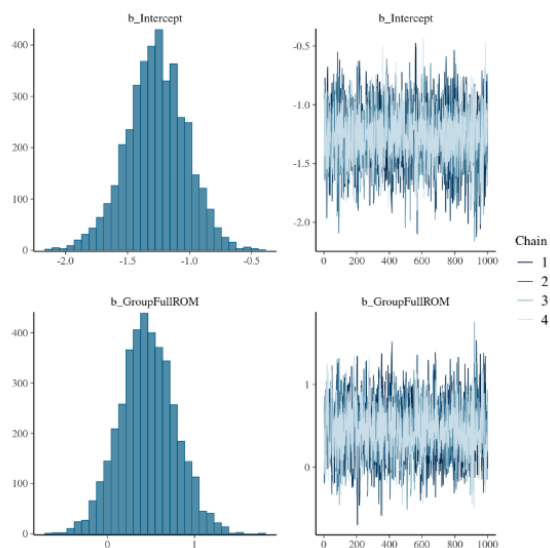


Fig. 19: Plot from Logistic Model

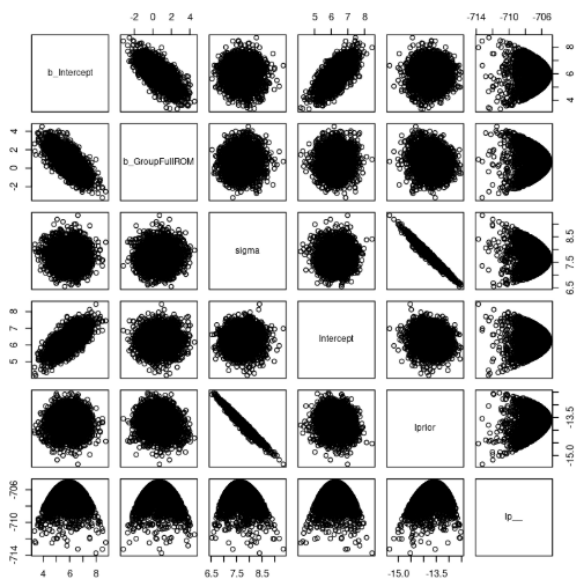


Fig. 18: Plot from Linear Model

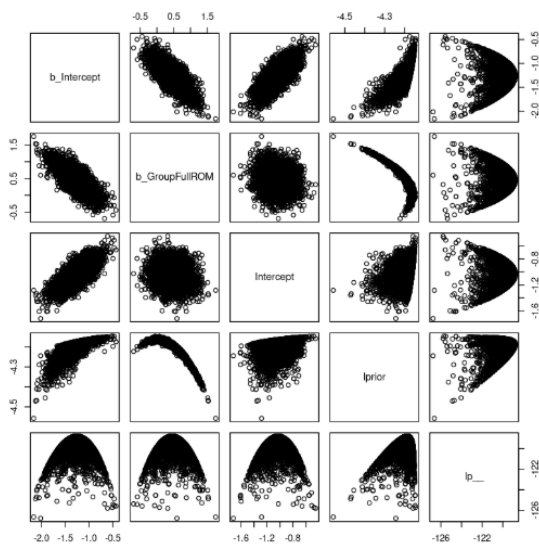


Fig. 20: Plot from Logistic Model

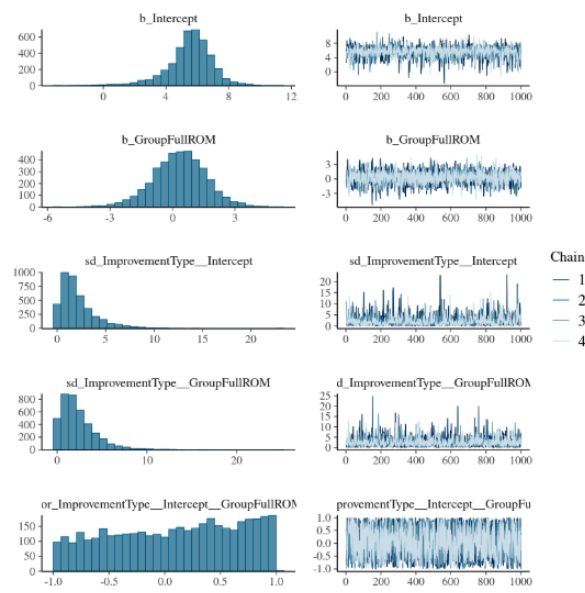


Fig. 21: Plot from Hierarchical Model

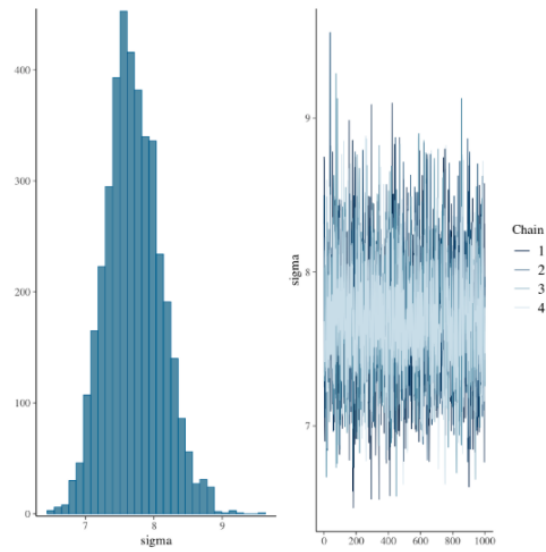


Fig. 22: Plot from Hierarchical Model

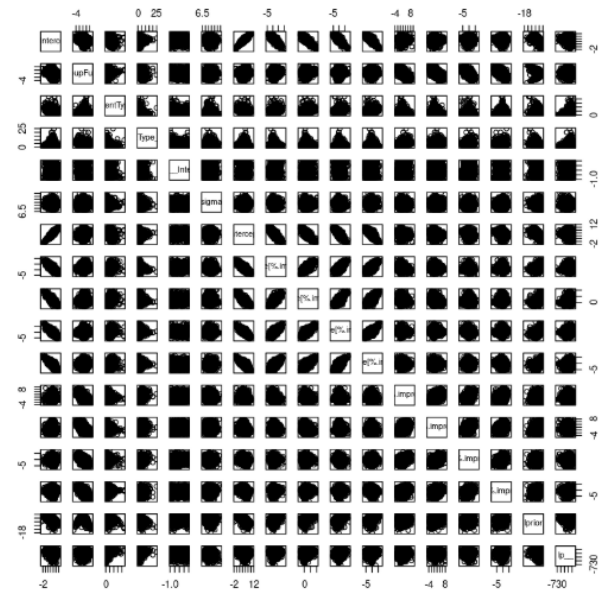


Fig. 23: Plot from Hierarchical Model

7. APPENDIX B: CODE

```
# Load necessary libraries
library(brms)
library(readxl)
library(tidyr)
library(dplyr)

# Load the data
data <- read_excel("/notebooks/Data Sheet.
  xlsx", sheet = "Muscle Thickness")

# Clean the GROUP column
data$GROUP <- ifelse(data$GROUP %in% c("1",
  "1 = pROM"), 1,
  ifelse(data$GROUP %in% c
    ("2", "2 = fROM"),
    2, NA))

# Convert GROUP to a factor with appropriate
  labels
data$Group <- factor(data$GROUP, levels = c
  (1, 2), labels = c("Lengthened Partial",
  "Full ROM"))

# Reshape the data to long format to include
  all '% improv' columns
data_long <- data %>%
  pivot_longer(
    cols = starts_with("% improv"),
    names_to = "ImprovementType",
    values_to = "Improvement"
  ) %>%
  drop_na(Improvement) # Remove rows with
    missing values in Improvement

# Verify the reshaped data
str(data_long)
table(data_long$Group,
  data_long$ImprovementType, useNA = "ifany
  ")

# Define proper priors
priors <- c(
  set_prior("normal(0, 10)", class = "
    Intercept"), # Prior for the intercept
  set_prior("normal(0, 5)", class = "b"),
    # Prior for coefficients
  set_prior("exponential(1)", class = "sigma
    ") # Prior for residual standard
    deviation
)

# Bayesian linear model with all improvement
  types
model <- brm(
  Improvement ~ Group,
  data = data_long,
```

```
family = gaussian(),
prior = priors,
control = list(adapt_delta = 0.95),
cores = 4
)

# Summary of the model
summary(model)

# Posterior predictive checks
pp_check(model)

# Hypothesis testing
hypothesis(model, "GroupFullROM < 0")
```

Listing 1: Code for the Linear Regression Model

```
# Load necessary libraries
library(brms)
library(readxl)
library(tidyr)
library(dplyr)

# Load the data
data <- read_excel("/notebooks/Data Sheet.
  xlsx", sheet = "Muscle Thickness")

# Clean the GROUP column
data$GROUP <- ifelse(data$GROUP %in% c("1",
  "1 = pROM"), 1,
  ifelse(data$GROUP %in% c
    ("2", "2 = fROM"),
    2, NA))

# Convert GROUP to a factor with appropriate
  labels
data$Group <- factor(data$GROUP, levels = c
  (1, 2), labels = c("Lengthened Partial",
  "Full ROM"))

# Reshape the data to long format to include
  all '% improv' columns
data_long <- data %>%
  pivot_longer(
    cols = starts_with("% improv"),
    names_to = "ImprovementType",
    values_to = "Improvement"
  ) %>%
  drop_na(Improvement) # Remove rows with
    missing values in Improvement

# Define a binary outcome for success/failure
  based on a threshold
threshold <- 10 # Set an appropriate
  threshold
data_long <- data_long %>%
  mutate(MuscleGrowthBinary = ifelse(
    Improvement >= threshold, 1, 0))
```

```

# Verify the reshaped data
str(data_long)
table(data_long$Group,
      data_long$ImprovementType,
      data_long$MuscleGrowthBinary, useNA = "
      ifany")

# Define priors
priors <- c(
  set_prior("normal(0, 5)", class = "
    Intercept"), # Prior for the intercept
  set_prior("normal(0, 2)", class = "b")
    # Prior for coefficients
)

# Bayesian logistic regression model with all
  improvement types
model <- brm(
  MuscleGrowthBinary ~ Group,
  data = data_long,
  family = bernoulli(link = "logit"), #
    Logistic regression
  prior = priors,
  control = list(adapt_delta = 0.95),
  cores = 4
)

# Summary of the model
summary(model)

# Posterior predictive checks
pp_check(model)

# Hypothesis testing
hypothesis(model, "GroupFullROM < 0")

```

Listing 2: Code for the Logistic Regression Model

```

# Load necessary libraries
library(brms)
library(readxl)

# Load the data
data <- read_excel("/notebooks/Data Sheet.
  xlsx", sheet = "Muscle Thickness")

# Clean the GROUP column
data$GROUP <- ifelse(data$GROUP %in% c("1",
  "1 = pROM"), 1,
  ifelse(data$GROUP %in% c
    ("2", "2 = fROM"),
    2, NA))

# Convert GROUP to a factor with appropriate
  labels
data$Group <- factor(data$GROUP, levels = c
  (1, 2), labels = c("Lengthened Partial",
    "Full ROM"))

```

```

# Gather improvement columns into long format
  for hierarchical modeling
library(tidyr)
data_long <- data %>%
  pivot_longer(cols = starts_with("% improv")
    , # Adjust this to match all relevant
      improvement columns
      names_to = "ImprovementType",
      values_to = "Improvement")

# Remove rows with missing values
data_long <- data_long[!is.na(
  data_long$Improvement), ]

# Define priors
priors <- c(
  set_prior("normal(0, 5)", class = "
    Intercept"), # Prior for the overall
    intercept
  set_prior("normal(0, 2)", class = "b"),
    # Prior for group-level
    coefficients
  set_prior("student_t(3, 0, 10)", class = "
    sd") # Prior for random effects
    standard deviation
)

# Hierarchical Bayesian logistic regression
  model
model <- brm(
  Improvement ~ Group + (1 + Group |
    ImprovementType), # Group-level
    effects for ImprovementType
  data = data_long,
  family = gaussian(), # Use Gaussian if
    Improvement is continuous; switch to
    Bernoulli for binary outcomes
  prior = priors,
  control = list(adapt_delta = 0.95),
  cores = 4
)

# Summary of the model
summary(model)

# Posterior predictive checks
pp_check(model)

# Hypothesis testing
hypothesis(model, "GroupFullROM < 0")

```

Listing 3: Code for the Hierarchical Model