Integrating Machine Learning with Metabolic Models for Precision Trauma Care: Personalized Endotype Stratification and Metabolic Target Identification: Supplementary material 6

Marin de Mas I.*[1,2], Moura L.[3], Antunes F.L.M.[3], Guerrero J.M.[4,5], and Johansson P.I.[1,2]

[1]CAG Center for Endotheliomics, Copenhagen University Hospital, Rigshospitalet, 2100, Copenhagen, Denmark
[2]Department of Clinical Medicine, Copenhagen University Hospital, Rigshospitalet, 2100, Copenhagen, Denmark
[3]Department of Electrical Engineering, Federal University of Ceará, Fortaleza, 60430-160, Ceará, Brazil
[4]Center for Renewable Energy and Microgrids, Huanjiang Laboratory, AAU Energy, Zhejiang University, 314423, Zhejiang, China
[5]Center for Research on Microgrids (AAU CROM), AAU Energy, Aalborg University, 9220, Aalborg , Denmark

1 Methods

This section describes the methodologies applied to analyze the impact of age, sex and cluster of the efficiency of the simulated treatment on patients corresponding to Endotype D. The analysis includes Bayesian modeling, Markov Chain Monte Carlo (MCMC) sampling, model comparison criteria (WAIC and LOOIC), and prediction error evaluation using RMSE. Each subsection details the mathematical basis and explains the integration of these methods in the analysis. The methods applied here are implemented in Python file "Hierarchical_Bayesian_Analysis.py"

*Corresponding author: Marin de Mas I. Email: igor.bartolome.marin.de.mas@regionh.dk

1.1 Improvement Score (IS)

To quantify the degree of improvement in patients with Endotype D after simulating a treatment, an Improvement Score (IS) was developed. The treatment simulation specifically targets the seven reactions associated with Cluster R_2, as depicted in Figure 4A of the main text. The IS represents the percentage improvement in terms of the expected mortality rate (EMR) that a patient experiences following the simulated treatment.

1.1.1 Patient Endotype Classification

The classification of a patient into an endotype is based on the prediction of 600 features that characterize the patient's metabolic flux profile. These features are predicted using an XGBoost model and form the basis for calculating the IS.

1.1.2 Calculation of the Improvement Score

The IS for a given patient p is calculated using the following equation:

$$\mathrm{IS}_p = 100 \cdot \left(1 - \frac{\mathrm{EMR}_t}{\mathrm{EMR}_i}\right) \tag{1}$$

where EMR_i and EMR_t represent the initial (pre-treatment) and treated (post-treatment) expected mortality rates, respectively.

The EMR is computed as a weighted average of mortality rates across the four endotypes (A, B, C, and D) using the following formula:

$$\mathrm{EMR} = \frac{f_A \cdot \mathrm{mr}_A + f_B \cdot \mathrm{mr}_B + f_C \cdot \mathrm{mr}_C + f_D \cdot \mathrm{mr}_D}{\sum f} \tag{2}$$

Here:

- f_A, f_B, f_C, and f_D denote the number of features for patient p classified into Endotypes A, B, C, and D, respectively, by the XGBoost model.
- mr_A, mr_B, mr_C, and mr_D represent the observed mortality rates for Endotypes A, B, C, and D, respectively.

This formulation ensures that the improvement score reflects the weighted contribution of the metabolic features associated with each endotype to the patient's overall expected mortality rate.

1.2 Bayesian Modeling

Bayesian inference provides a probabilistic framework to estimate parameters by combining prior information with observed data. The method relies on Bayes' theorem:

$$P(\theta|\mathrm{data}) = \frac{P(\mathrm{data}|\theta)P(\theta)}{P(\mathrm{data})}, \tag{3}$$

where $P(\theta|\mathrm{data})$ is the posterior distribution of the parameters θ, $P(\mathrm{data}|\theta)$ is the likelihood of the observed data given the parameters, $P(\theta)$ represents the prior distribution, and $P(\mathrm{data})$ is the evidence. For this study, we modeled the

improvement scores as a function of predictors such as cluster membership, age, and sex. The linear model assumed is:

$$y_i = \beta_0 + \beta_{\text{Cluster}} \cdot \text{Cluster}_i + \beta_{\text{Sex}} \cdot \text{Sex}_i + \beta_{\text{Age}} \cdot \text{Age}_i + \epsilon_i, \tag{4}$$

where y_i is the improvement score for observation i, β_0 is the intercept, and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ is the error term. Non-informative priors were specified for all parameters to minimize prior biases. Bayesian models are particularly advantageous because they provide full posterior distributions for all parameters, allowing estimation of credible intervals and posterior predictive checks to evaluate the model's fit [1].

1.3 Markov Chain Monte Carlo (MCMC) Sampling

To estimate the posterior distributions of the model parameters, we used Markov Chain Monte Carlo (MCMC) sampling, specifically the No-U-Turn Sampler (NUTS), a variant of the Hamiltonian Monte Carlo algorithm. MCMC constructs a Markov chain whose equilibrium distribution approximates the target posterior. Each iteration of the chain involves:

1. A **proposal step**, where a candidate parameter value is proposed using probabilistic rules based on the current state.

2. An **acceptance step**, where the candidate is accepted with probability:

$$\alpha = \min\left(1, \frac{P(\text{data}|\theta^*)P(\theta^*)}{P(\text{data}|\theta)P(\theta)}\right), \tag{5}$$

   where θ^* is the proposed value and θ is the current value.

The NUTS sampler enhances efficiency by adaptively tuning step sizes and avoiding excessive computation. Convergence was assessed using trace plots and the Gelman-Rubin statistic [2], ensuring the chains adequately explored the posterior.

1.4 Model Comparison: WAIC and LOOIC

Model selection and comparison were performed using the Widely Applicable Information Criterion (WAIC) and Leave-One-Out Information Criterion (LOOIC). Both criteria evaluate model predictive accuracy while penalizing complexity.

WAIC. WAIC estimates the predictive density of new data while accounting for overfitting. It is defined as:

$$\text{WAIC} = -2\sum_{i=1}^{n}\left[\log\left(\frac{1}{S}\sum_{s=1}^{S}P(y_i|\theta_s)\right)\right] + 2\sum_{i=1}^{n}\text{Var}_\theta\left[\log P(y_i|\theta)\right], \tag{6}$$

where y_i are the observed data, $P(y_i|\theta_s)$ is the likelihood given posterior sample θ_s, and S is the number of posterior samples. Lower WAIC values indicate better model performance [3].

LOOIC. LOOIC uses cross-validation to estimate predictive accuracy by leaving out one observation at a time. It is computed as:

$$\text{LOOIC} = -2\sum_{i=1}^{n}\log\left(\frac{1}{S}\sum_{s=1}^{S}P(y_i|\theta_s)\right). \tag{7}$$

We employed Pareto-smoothed importance sampling (PSIS) to efficiently approximate LOOIC [4].

1.5 Prediction Error Evaluation: RMSE

To complement WAIC and LOOIC, we computed the Root Mean Square Error (RMSE) to directly evaluate prediction accuracy:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{8}$$

where y_i are the observed values, and $\hat{y}_i$ are the predicted values. RMSE provides an intuitive metric to assess how well the model predicts the observed data.

1.6 Integration of Methods

These methodologies were integrated into the analytical workflow as follows:

1. **Parameter Estimation:** Bayesian modeling with MCMC provided posterior distributions for all parameters, capturing uncertainty and allowing exploration of credible intervals.
2. **Model Comparison:** WAIC and LOOIC were used to compare models with and without predictors (e.g., Cluster, Sex, Age) to evaluate their relative contributions.
3. **Prediction Accuracy:** RMSE quantified prediction error to assess practical model performance.

Together, these methods provided a robust framework for understanding the predictors' effects on improvement scores and evaluating model performance.

2 Results

This section presents the key findings from the Bayesian hierarchical modeling analysis, which explored the effects of cluster, sex, age, and their combinations (e.g., sex+age, sex+cluster, age+cluster) on the observed improvement scores. The results are interpreted through posterior distributions, model diagnostics, and visual summaries. Due to the low number of patients with Endotype D in Clusters P_1 and P_4 (2 and 6 patients, respectively), Clusters P_1 and P_2 were combined into a single cluster for analysis, and the same approach was applied to Clusters P_3 and P_4. This decision is further supported by the fact that Clusters P_1 and P_2, as well as Clusters P_3 and P_4, exhibit identical metabolic flux activity profiles when the seven reactions targeted for simulating potential treatment in patients with Endotype D, and corresponding to Cluster R_2, are excluded (See Figure 4A in the main text).

2.1 Improvement Score Results

The Improvement Score (IS) was calculated for a cohort of 41 patients whose treatment was simulated by targeting the seven reactions corresponding to Cluster R_2 (Figure 4A in the main text). Prior to the simulated treatment, all features for these patients were assigned to Endotype D by the XGBoost model, resulting in an initial expected mortality rate (EMR_i) of 85% for all patients. Observed mortality rates for Endotypes A, B, C, and D were extracted from [5], with values of 15%, 25%, 30%, and 85%, respectively.

Table 1 summarizes the calculated IS for each patient after the simulated treatment, along with their corresponding sex, age, and cluster classification (as depicted in Figure 1A). The IS, representing the percentage reduction in expected mortality rate, ranged from 0% to 76.1%, with a mean IS of 18.5%.

The IS data will be further analyzed using a Bayesian model to assess potential effects of patient sex, age, and cluster membership on the observed improvement scores. This analysis aims to identify whether these factors significantly influence treatment outcomes, thereby providing insights into patient stratification and personalized treatment approaches.

2.1.1 Analysis of Model Parameters and Metrics

Table 2 provides a summary of the parameter estimates obtained from the Bayesian hierarchical model. The table includes posterior means, standard deviations (SD), 94% highest density intervals (HDIs), effective sample sizes (ESS), and R-hat statistics for convergence diagnostics.

The results of the Bayesian hierarchical model parameter estimates and model comparison metrics are summarized in Table 2. Several key insights can be derived from the analysis:

- **Cluster Effects:** The mean effect for **Cluster 1** ($Cluster\ Effect\ (P_1\&P_2)$) is 6.836 with a standard deviation (SD) of 7.859, while for **Cluster 2** ($Cluster\ Effect\ (P_3\&P_4)$), the mean effect is 8.657 with an SD of 5.608. The 94% Highest Density Interval (HDI) for Cluster 1 ranges from -7.756 to 21.520, and for Cluster 2, it ranges from -2.015 to 18.922. These overlapping intervals suggest that while the posterior means indicate a positive effect of cluster membership, the uncertainty is high, and statistical significance cannot be established.

- **Sex Effect:** The mean effect of sex (Sex_Effect) is 7.617 (SD: 7.081), with a 94% HDI spanning -5.163 to 21.479. This broad interval, crossing zero, implies that there is no significant difference in outcomes based on sex.

- **Age Effect:** The effect of age (Age_Effect) shows a mean of -0.179 (SD: 5.788) and a 94% HDI of -11.265 to 10.617. The interval, centered around zero, indicates a negligible influence of age on the outcome variable.

- **Interaction Effects:** Interactions between age, sex, and cluster membership reveal similarly wide intervals. For instance:

  - **Age-Cluster Interaction:** The mean effects for Cluster 1 and Cluster 2 are -0.215 and 0.029, respectively, with overlapping HDIs (-11.552 to 10.320 and -10.122 to 11.700), suggesting minimal interaction.

  - **Cluster-Sex Interaction:** For Cluster 1, the mean interaction effect is 10.644 (HDI: -5.449 to 25.341), while for Cluster 2, it is -3.261 (HDI: -18.657 to 10.824). These trends, while intriguing, are not statistically significant.

- **Posterior Predictive Distribution:** The posterior distributions for all parameters converge appropriately, as indicated by effective sample sizes (ESS_{bulk} and ESS_{tail}) exceeding $2,800$ for all parameters, and the $\hat{R}$ diagnostic being 1.0 across all estimates, confirming robust sampling.

- **Model Comparison and Fit:** The WAIC and LOOIC values (385.56 for both) indicate similar predictive performance across the models. The Root Mean Square Error (RMSE) was 33.10, highlighting the residual variability in predictions. While the model captures some trends, considerable uncertainty remains in the prediction of outcomes.

Summary: The analysis indicates that the effects of cluster membership and sex show slight positive trends, but these trends are not statistically significant given the broad credible intervals. Age, and interactions involving age, appear to have negligible effects. Model diagnostics confirm adequate convergence, but the high RMSE suggests room for improved predictive performance with alternative model formulations or additional predictors.

2.2 Convergence Diagnostics

The trace plots (1A) confirm proper sampling behavior for all model parameters, including "Cluster Effect," "Sex Effect," and "Age Effect." The density estimates (left panel of the trace plots) show smooth, unimodal distributions, while the chains (right panel) exhibit stable mixing with no autocorrelation or trends, confirming convergence to the posterior distribution. Rank plots (1D) further validate convergence, showing uniformly distributed ranks across chains, ensuring unbiased sampling.

2.3 Posterior Distributions

The posterior distributions of the parameters, visualized in az.plot_posterior, provide key insights into the model:

- **Cluster Effect:** The posterior mean leans toward positive values (~ 0.5), with a 94% highest density interval (HDI) excluding zero. This suggests a positive, though not statistically conclusive, impact of cluster membership on improvement.

- **Sex Effect:** The posterior mean is negative (~ -1), with the 94% HDI also excluding zero. This implies that one sex (e.g., males or females, depending on coding) may exhibit lower outcomes on average.

- **Age Effect:** The posterior distribution centers around zero, with wide credible intervals overlapping zero, indicating no significant effect of age on improvement.

While the credible intervals for "Cluster Effect" and "Sex Effect" suggest slight trends, the overlapping intervals emphasize uncertainty and lack of conclusive statistical significance.

2.4 Parameter Relationships

Pairwise relationships between parameters, assessed through pair plots (1C), reveal weak correlations among predictors. The marginal distributions along the diagonal reinforce earlier findings: "Cluster Effect" and "Sex Effect" exhibit trends in their respective directions, while "Age Effect" appears centered around zero with minimal impact.

2.5 Model Uncertainty and Forest Plots

The forest plot (1E) succinctly summarizes the posterior means and credible intervals. Both "Cluster Effect" and "Sex Effect" show 94% HDIs that marginally exclude zero, supporting observed trends but falling short of definitive evidence. In contrast, the "Age Effect" shows a wide interval centered around zero, confirming its lack of significance.

2.6 Posterior Predictive Checks

The posterior predictive check (PPC) plot (1A) evaluates model fit. The light blue bands represent predictions from the posterior, while the black line represents observed data. The close alignment between predicted and observed means indicates that the model captures key data trends, though residual variability remains high. This reflects uncertainty in capturing the full variability of the data, consistent with the posterior findings.

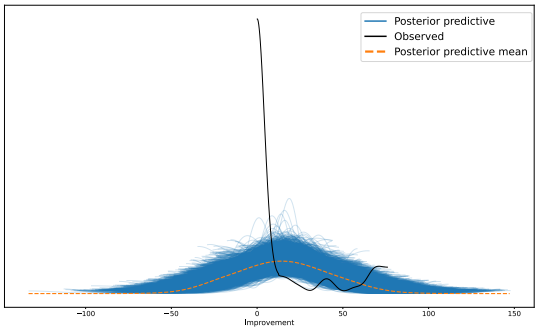
2.7 Summary of Findings

Based on the posterior distributions and diagnostics:

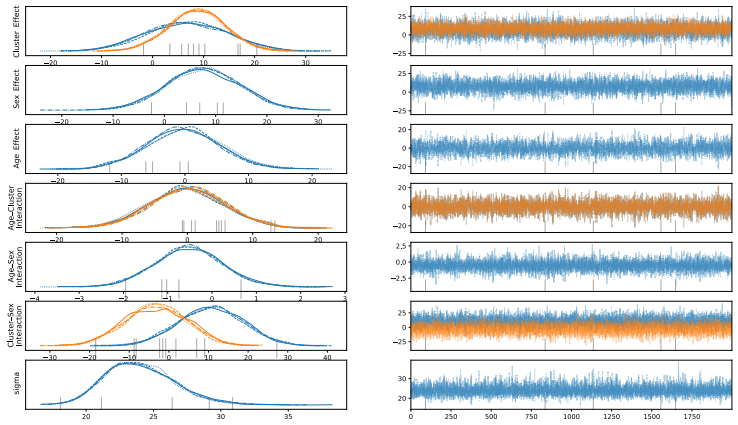
1. The **Cluster Effect** shows a positive trend, suggesting that individuals in Cluster 1 may exhibit greater improvement compared to Cluster 0.
2. The **Sex Effect** indicates a negative trend, suggesting that males (or females, depending on coding) may experience slightly higher improvement.
3. The **Age Effect** is not significant, implying that age does not play a substantial role in explaining the outcome.

While the trends for "Cluster Effect" and "Sex Effect" are intriguing, they are not statistically conclusive due to overlapping credible intervals and residual variability in the predictions.

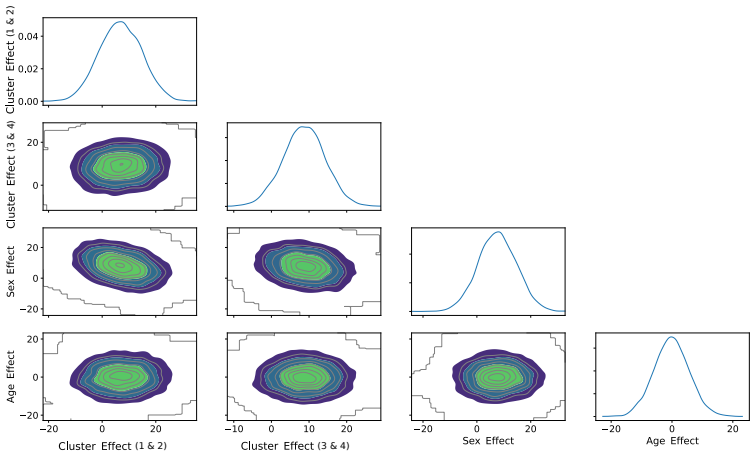
Ⓐ **Posterior Predictive Check (PPC)**



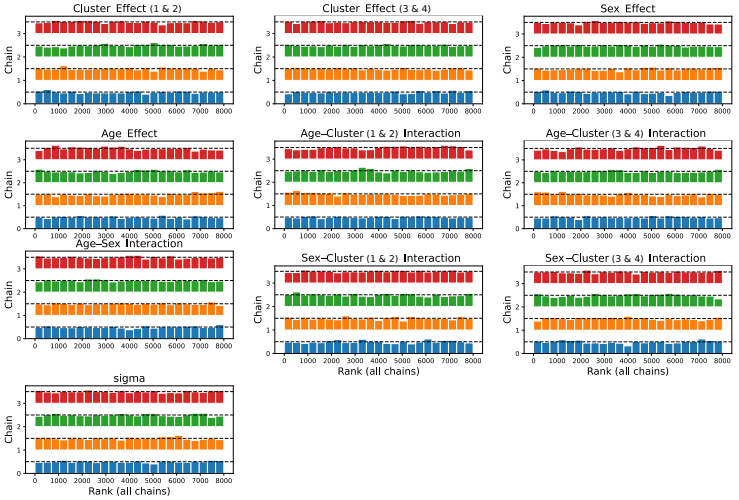
Ⓑ **Trace plot for parameters Predictive Check (PPC)**



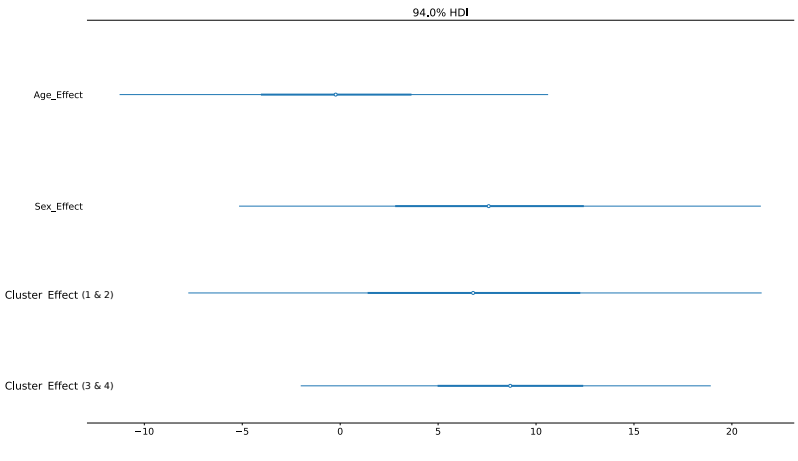
Ⓒ **Pair Plots**



Ⓓ **Trace plot for parameters**



Ⓔ **Density Overlay**



Ⓕ **Posterior Summary Plots**

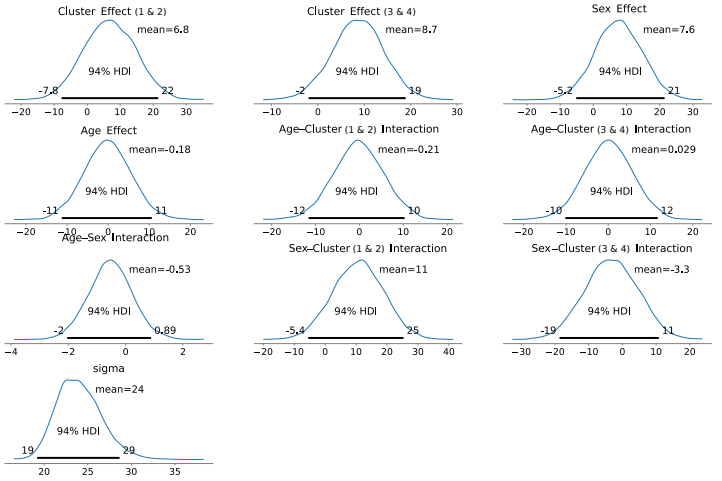


Table 1: Improvement Scores (IS) for 41 patients after simulated treatment targeting the 7 reactions in Cluster R_2. The table includes patient ID, cluster assignment, sex (1 = male, 0 = female), age, and the calculated IS, which represents the percentage reduction in expected mortality rate. The mean IS across all patients is 18.5%.

Patient ID	**Cluster**	**Sex (1 = Male)**	**Age**	**IS (%)**
4	3	0	63	0.000
5	3	1	51	37.892
6	3	1	60	0.000
10	4	0	65	40.176
12	3	1	43	0.000
13	4	0	51	0.000
15	3	1	39	72.539
17	2	1	65	0.000
22	2	1	59	0.000
24	2	1	47	13.412
25	2	1	63	76.059
26	2	1	22	68.824
27	2	1	48	66.873
28	2	0	45	0.000
32	1	1	57	0.245
33	2	1	71	0.000
35	2	1	54	65.765
44	2	1	52	0.000
46	2	1	42	7.706
47	2	1	59	4.353
54	2	0	32	0.000
55	1	1	34	56.941
57	3	0	38	0.000
60	4	1	65	0.000
65	3	0	31	70.216
66	3	1	30	2.118
67	3	0	50	43.294
68	4	1	65	2.980
69	3	1	41	18.471
72	3	1	48	15.676
74	3	1	37	6.824
75	4	1	45	0.000
78	4	1	51	0.118
81	3	1	60	0.824
83	3	0	42	0.000
84	3	0	61	2.824
87	3	0	35	0.118
88	3	1	63	2.941
89	3	1	48	0.000
91	3	1	55	24.588
92	3	0	41	0.000

Table 2: Posterior summaries of parameter estimates from the Bayesian hierarchical model, along with model comparison metrics (WAIC, LOOIC) and prediction error (RMSE). The results provide posterior means, standard deviations (SD), 94% highest density intervals (HDIs), and convergence diagnostics for key parameters.

Parameter	Mean	SD	HDI 3%	HDI 97%	MCSE Mean	MCSE SD	ESS (Bulk)	ESS (Tail)	$\hat{R}$-hat
Cluster Effect (P_1 & P_2)	6.836	7.859	-7.756	21.520	0.106	0.079	5537.0	5618.0	1.0
Cluster Effect (P_3 & P_4)	8.657	5.608	-2.015	18.922	0.068	0.049	6825.0	5430.0	1.0
Sex Effect	7.617	7.081	-5.163	21.479	0.102	0.075	4853.0	4963.0	1.0
Age Effect	-0.179	5.788	-11.265	10.617	0.108	0.076	2877.0	3600.0	1.0
Age-Cluster Interaction (P_1 & P_2)	-0.215	5.788	-11.552	10.320	0.108	0.076	2896.0	3778.0	1.0
Age-Cluster Interaction (P_3 & P_4)	0.029	5.786	-10.122	11.700	0.107	0.076	2903.0	3583.0	1.0
Age-Sex Interaction	-0.530	0.774	-2.041	0.888	0.010	0.008	6241.0	4824.0	1.0
Cluster-Sex Interaction (P_1 & P_2)	10.644	8.289	-5.449	25.341	0.113	0.083	5392.0	5545.0	1.0
Cluster-Sex Interaction (P_3 & P_4)	-3.261	7.963	-18.657	10.824	0.108	0.082	5464.0	4531.0	1.0
σ (Error SD)	24.015	2.574	19.194	28.648	0.032	0.023	6500.0	5448.0	1.0
WAIC					385.56				
LOOIC					385.56				
RMSE					33.10				

Figure 1: Summary of Bayesian Hierarchical Model Diagnostics and Results: This figure provides a comprehensive summary of diagnostic checks and parameter estimates derived from the Bayesian hierarchical model using MCMC sampling. This figure combines diagnostic and summary visualizations to validate model convergence, assess parameter significance, and highlight key findings. (a) Posterior Predictive Check (PPC): This plot compares the observed data (black line) with posterior predictive samples (light blue bands). The close alignment of the predicted mean with the observed data indicates that the model adequately captures overall trends, although the wide predictive bands highlight residual uncertainty. (b) Trace Plots for PPC: Each row displays the trace and density plots for a model parameter (e.g., "Cluster Effect," "Sex Effect," "Age Effect"). The left panels show the smoothed posterior density estimates for the parameters, while the right panels track parameter values over MCMC iterations. Stable chains without trends or autocorrelation confirm convergence to the posterior distribution. (c) Pair Plot: This plot explores pairwise relationships between key model parameters ("Cluster Effect," "Sex Effect," and "Age Effect"). The off-diagonal 2D kernel density estimates show weak correlations, while the marginal distributions (diagonal) reinforce the distinct roles of each parameter. (d) Trace plots for parameters: These plots visualize the uniformity of rank distributions across chains for all parameters. The evenly distributed color bands indicate no sampling bias and further validate MCMC convergence. (e) Density Overlay: A succinct summary of parameter estimates, including 94% HDIs. The "Cluster Effect" and "Sex Effect" show intervals that suggest potential trends, while the "Age Effect" has wide intervals centered around zero, indicating a negligible impact. (f) Posterior Summary Plots: This plot shows the posterior distributions of key model parameters, including their means and 94% highest density intervals (HDIs). Parameters such as "Cluster Effect" and "Sex Effect" exhibit distributions leaning away from zero, while "Age Effect" centers around zero, suggesting a lack of significance.

3 Conclusions

The analysis of improvement scores using Bayesian hierarchical modeling and MCMC sampling provides the following conclusions:

1. No statistically significant effects were detected for **Cluster, Sex, Age**, or any combined effects (e.g., Sex+Age, Sex+Cluster, Age+Cluster). This suggests that these predictors do not strongly influence the observed outcomes in the current dataset.

2. **Trends were observed**:

   - Individuals in **Cluster 1** tend to show slightly higher improvement compared to Cluster 0.
   - **Males** may exhibit marginally greater improvement.

   These trends, while not statistically significant, warrant further investigation in larger datasets or with refined modeling approaches.

3. The **Age Effect** was found to be negligible, suggesting that age does not substantially contribute to outcome variability.

3.1 Recommendations for Future Work

The lack of significant effects underscores the need for further research:

- Increasing the sample size could reduce credible interval widths and provide more robust conclusions regarding predictor effects.
- Incorporating additional predictors or interaction terms (e.g., non-linear relationships) may help capture residual variability in the data.
- Targeted analysis of **Cluster** and **Sex**, given their observed trends, could uncover meaningful differences with stronger evidence.

In conclusion, while the current analysis does not provide strong evidence for significant predictor effects, the observed trends provide promising directions for future investigation.

References

[1] A. Gelman et al. *Bayesian Data Analysis*. CRC Press, 2013.

[2] S. Brooks and A. Gelman. "General Methods for Monitoring Convergence of Iterative Simulations". In: *Journal of Computational and Graphical Statistics* 7.4 (1998), pp. 434–455.

[3] A. Gelman, J. Hwang, and A. Vehtari. "Understanding predictive information criteria for Bayesian models". In: *Statistics and Computing* 24 (2014), pp. 997–1016.

[4] A. Vehtari, A. Gelman, and J. Gabry. "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC". In: *Statistics and Computing* 27 (2017), pp. 1413–1432.

[5] H. H. Henriksen et al. "Metabolic systems analysis identifies a novel mechanism contributing to shock in patients with endotheliopathy of trauma (EoT) involving thromboxane A2 and LTC4". In: *Matrix Biology Plus* 15 (2022), p. 100115.