Prostate Cancer Survival Analysis

January 20, 2026

1 Introduction

Prostate cancer is one of the most prevalent malignancies among men, and understanding the factors influencing survival outcomes in advanced stages is critical for improving treatment strategies.

In this study we apply Bayesian survival analysis techniques to analyze time-to-event data from patients with stage 3 and 4 prostate cancer. By employing hierarchical and non-hierarchical Bayesian models, we aim to explore the relationship between survival outcomes and key covariates.

With the growing popularity of Bayesian methods due to new fast computing packages like brms, Stan, rstanarm that allow for ease of use model building, debugging, inference, and performance, decades-old problems especially in medicine are widely being revisited.

This project is motivated by the potential to enhance the interpretability and predictive power of survival models in a clinical context. By leveraging advanced Bayesian methods, we not only aim to provide a nuanced understanding of the factors affecting survival but also highlight the advantages and limitations of these approaches in modeling complex, real-world datasets. In doing so, we hope to contribute to the ongoing effort to refine treatment strategies for advanced prostate cancer and improve patient outcomes.

2 Data Description

The dataset for this study was sourced from Vanderbilt Biostatistics Datasets. The data consisted of patients with stage 3 or 4 prostate cancer, who were followed for up to 76 months. The dataset includes 502 datapoints, each consisting of 22 different variables. Key variables that were used in our study include:

Variable	**Description**
rx	Treatment / Medication dose
dtime	Follow-up time (months)
status	Patient status (alive or dead)
age	Age of the patient
sz	Tumor size
ap	Prostatic acid phosphatase level

Table 1: Summary of variables used in our study.

We excluded most of the variables to simplify our model and to filter out weakly or non-informative variables. Variables in table 1 were considered most relevant for our study, due to their reflective properties of cancer stage.

3 Methodology

3.1 Model specifications

3.1.1 Hierarchical Model

We chose a Bayesian hierarchical survival model to account for variability at multiple levels in our data. Specifically, the model analyzes time-to-event (death) data, with the response variable, survival time (*dtime*|*cens(died)*), modeled using the Weibull distribution. The Weibull distribution is particularly well-suited for survival analysis because it flexibly models hazard rates that can increase, decrease, or remain constant over time.

Predictors include scaled covariates of primary tumor size (*sz_scaled*), prostatic acid (*ap_scaled*), and age (*age_scaled*), which are used to explain individual-level differences in survival outcomes. Additionally, medication dosage (*rx*) is incorporated as a random effect to account for variability in survival times across different dosage levels. This inclusion of random effects makes the model hierarchical, as it accounts for both individual-level and group-level variation.

The formula for hierarchical model is specified as follows:

$$Time|Death \sim TumorSize + ProstaticAcid + Age + (1|Medication)$$

We used weakly informative priors in our model to regularize the model while allowing the data to inform the estimates:

- **Coefficients** (`class = "b"`): A normal prior (`prior(normal(0, 1))` is applied to the coefficients, assuming moderate effects centered around zero. This weakly informative prior prevents extreme estimates and lets the data drive the model.

- **Global Intercept** (`class = "Intercept"`): A Student's t distribution prior `prior(student_t(3, 0, 2.5))` is employed for the baseline parameter. This choice is motivated by: Degrees of freedom = 3 provides robustness while maintaining finite variance, scale parameter of 2.5 accommodates reasonable variation in baseline survival, heavier tails than normal distribution allow for outlying baseline values, center at zero reflects scale of the standardized predictors

- **Random Effects Scale (`class = "sd"`)**: A half-Cauchy prior `prior(cauchy(0, 1))` is specified for the group-level standard deviations. This specification is advantageous because of: Restricts to positive values, appropriate for scale parameter, heavy tails accommodate potential large between-group variation, scale of 1 is consistent with the expected magnitude of variation on standardized predictors, provides more stable inference than inverse-gamma alternatives

```
hierarchical_model <- brm(
    formula = dtime | cens(died) ~ sz_scaled + ap_
       scaled + age_scaled + (1|rx),
    family = weibull(),
    data = survival_data,
    prior = c(
        prior(normal(0, 1), class=”b”),
        prior(student_t(3, 0, 2.5), class=”Intercept”),
        prior(cauchy(0, 1), class=”sd”)
        )
)
```

3.1.2 Non-Hierarchial Model

The non-hierarchical model excludes random effects, assuming that observations are independent of each other. This simplifies the model structure, focusing solely on fixed effects. In this model, predictors are treated as explanatory variables with no group-level variation. The Weibull distribution is used again to model survival time, as it appropriately handles varying hazard rates.

The formula for non-hierarchial model is specified as follows:

$$Time|Death \sim Age + TumorSize + ProstaticAcid + Medication$$

For this model we used different weakly informative priors:

- **Coefficients (`class = "b"`)**:
  A normal prior `prior(normal(0, 5))` is applied to the coefficients for the covariates. This prior allows for a wide range of possible effects while preventing extreme estimates. The larger standard deviation (5) reflects uncertainty about the exact size of the effects, providing flexibility in model fitting.

- **Intercept** (`class = "Intercept"`):
  A normal prior `prior(normal(0, 5))` is also used for the intercept. This prior allows for a flexible baseline survival time, accommodating variability in the intercept without imposing overly strict constraints.

- **Shape Parameter** (`class = "shape"`):
  A Gamma prior `prior(gamma(2, 1))` is assigned to the shape parameter of the Weibull distribution. The Gamma distribution ensures that the shape parameter is positive and can accommodate varying hazard rates. The chosen parameters imply a moderately increasing hazard rate but allow flexibility for the data to inform the shape.

```
non_hierarchical_model <- brm(
    formula = dtime | cens(died) ~ age_scaled + sz_
       scaled + ap_scaled + rx,
    data = data_clean,
    family = weibull(),
    prior = c(
        prior(normal(0, 5), class = ”b”),
        prior(normal(0, 5), class = ”Intercept”),
        prior(gamma(2, 1), class = ”shape”)
        ),
)
```

3.2 MCMC Parameters

For both models, we used Markov Chain Monte Carlo (MCMC) methods to estimate posterior distributions of our model parameters. These methods are implemented in the `brms` package, which interfaces with the `Stan` modeling language to perform Bayesian inference.

- **chains = 4**: We ran four independent chains.

- **iter = 4000**: The total number of iterations per chain, including both the warmup and sampling phases.

- **warmup = 2000**: Half of the iterations were used for ”warmup”.

- **adapt_delta = 0.95**: To improve convergence.

4 Model Diagnostics and Validation

4.1 MCMC Evaluation

To evaluate the convergence of our MCMC chains, we used Rhat Statistic, Effective Sample Size (ESS) and Divergences. The Rhat statistic assesses the convergence of the chain by comparing the variance between chains to the variance within chains. ESS quantifies the number of independent samples from the posterior distribution. It essentially tells how much effective information is gained from the MCMC sampling process. In practice a higher ESS indicates that the chains are efficiently exploring the parameter space and provide more reliable estimates of the true posterior distribution. Lastly, divergence is a diagnostic, which essentially tells where HMC algorithm, which MCMC uses, fails to explore the posterior distribution properly. These all were tested on both models separately.

- **Rhat Statistic**: Both models had R-hat values of 1.00 for all parameters, indicating that the chains converged to the same target distribution, making the results reliable for inference.

- **Effective Sample Size (ESS)**: ESS can be divided into bulk and tail ESS. Bulk ESS indicates the efficiency of the posterior sampling, which in the hierarchical model ranged from 1285 for the random intercept to 5082 for the scaled tumor size. These relatively high values indicate efficient sampling in the bulk of the posterior distribution, suggesting that the model's covariates were explored well. For tail ESS, which indicates the efficiency of tail exploring in the distribution, the values in hierarchical model ranged from 1011 to 2287. This indicates, that even if the values were lower than in bulk, they were high enough to suggest adequate sampling in the tails of the distribution.

  For non-hierarchical model, the bulk ESS ranged from 4878 for the intercept to 7229 for scaled age covariate. Therefore, the values of this model turned to be higher than in the hierarchical one, suggesting more efficient sampling in the bulk of the posterior. However, it could also indicate that the model is essentially simpler, which is true in this case. The tail ESS ranged had values from 4655 to 5762, which again indicates even more efficient sampling in the non-hierarchical model.

- **Divergences**: For both hierarchical models and for the non-hierarchical model, there were no divergences found, which indicates that for both models, the sampling process was overall successful and both models converged well. This might be a consequence of having adaptdelta parameter initially high (0.95), which makes the integrator to take smaller steps.

4.2 Posterior predictive checks

Posterior predictive checks (PPCs) were used to assess how well the model could reproduce the observed data. PPCs were performed for both models using the `pp_check()` function from the `brms` package.

- **Hierarchical model**: Using PPC, it can be seen from Figure 1 that the simulated data can partially agree with the true, observed data, especially the furthest right part of the distributions. However, it is quite clear that the observed data's distribution has some unique features that our predicted posterior model cannot describe. For instance, the true posterior seems to have two or three local maximums, which make it quite impossible to fit the true posterior perfectly, as we used Weibull distribution as the family value. Therefore, even though there seems to be some draws, where the true posterior and predicted posterior seems closer to each other, it cannot really find the second (or third) mode of the true distribution. Additionally, variability in the peak suggests some discrepancies in the model's ability to consistently capture the most frequent values. Furthermore, our descriptions seems to have heavier left tail, which could be simply because the Weibull distribution used assumes a certain form of skewness and tail behavior, but the observed data might exhibit more extreme values or different tail behavior that this model cannot replicate.

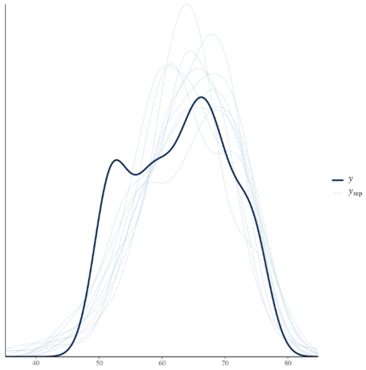

Figure 1: hierarchical model

- **Non-Hierarchical model**: The posterior predictive check for non-hierarchical model does show slightly improved results when comparing with the previous model, as the predicted draws seems to be able to find the second local maximum of the true posterior distribution (Figure 2). However, even with this model, when compared to true data, the PPC cannot really focus on the true peak, but has plenty of variation in this area. Moreover, the left tail seems to be again heavier than in the true data. Interestingly, some predictions would suggest that the second local maximum would appear in this area, as can be seen in the figure.

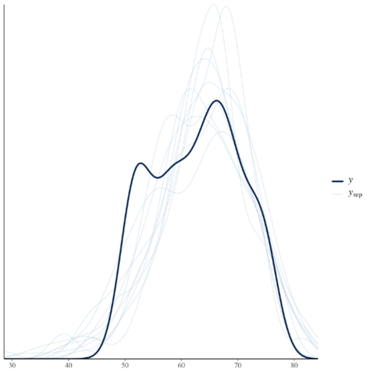

Figure 2: non-hierarchical model

4.3 Sensitivity analysis with respect to prior choices

Next, we implemented sensitivity analysis for both models to investigate their robustness and reliability with respect to varying priors. With each model, we ran the MCMC with two new set of priors, and then analyzed the posterior distributions of the covariates.

4.3.1 Changing priors for hierarchical model

First we analyze sensitivity for the hierarchical model with the following set of priors and compare the posterior distributions of the covariates with the original prior set (see Section 3.1.1):

```
prior = c(
          prior(normal(2, 1), class="b"),
          prior(student_t(3, 0, 2.5), class="Intercept"),
          prior(cauchy(0, 1), class=”sd”)
          )
```

where the covariate priors have been given more influence, i.e. their means have been increased, and the second prior set is:

```
prior = c(
        prior(normal(0, 1), class=”b”),
        prior(student_t(3, 0, 2.5), class=”Intercept”),
        prior(cauchy(0, 10), class=”sd”)
        )
```

where sd, the standard deviation of the random effect corresponding to (1|rx), is given an increased scale, essentially making the prior more uninformative. This allows for more flexibility in the model by enabling more extreme values for the random intercepts associated with the levels of rx.

The figures below reveal small to none effects of different priors on the posterior distributions. Regardless the prior set, distributions seem to be very similar. However, one thing that is quite noticeable is that the peak of every distribution have slight variations with respect to priors. This could be the reason that earlier in Section 4.2, we saw that the predictions could not find the true peak. By altering priors, this could be converged to the true mode, but this also requires more investigation, and not focusing only on experimental testing, which we did. Overall, we see that the prior alterations do not really affect the posterior distributions, suggesting that the model is relatively robust in this case.

Above, the blue distribution uses the original priors, red uses the prior set with modified class = ”*b*” prior, and green uses the prior set with modified class = ”*sd*” prior.

4.3.2 Changing priors for non-hierarchical model

For the second model, we implement similar alterations to the original prior set (see Section 3.1.2). The following prior sets are used to analyse the sensitivity in non-hierarchical model:

```
prior = c(
        prior(normal(0, 0.5), class=”b”),
        prior(normal(0, 5), class=”Intercept”),
        prior(gamma(2, 1), class=”shape”)
```

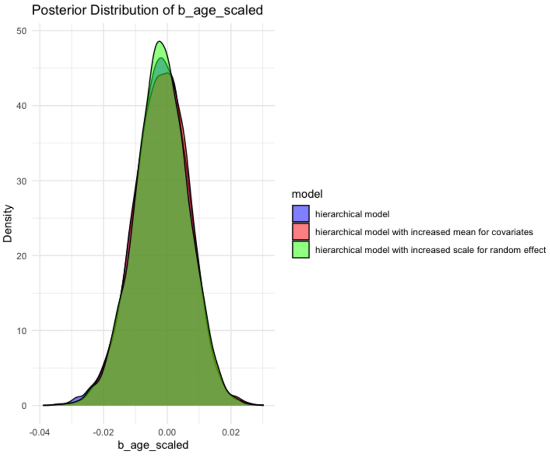

Figure 3: Posterior distribution of age (scaled)

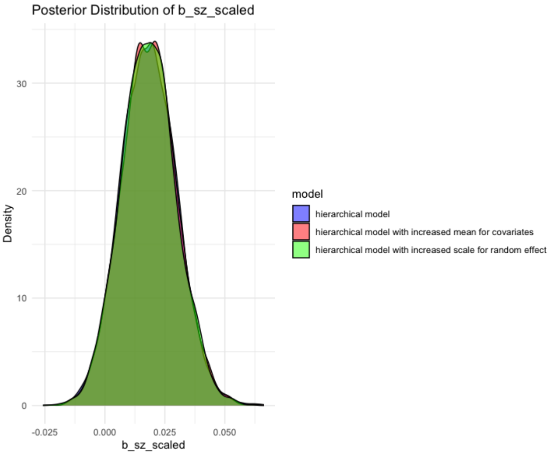

Figure 4: Posterior distribution of tumor size (scaled)

```
          )
```

where the covariate priors are now more informative, i.e. less variance. The second prior set for non-hierarchical model is:

```
prior = c(
          prior(normal(0, 5), class="b"),
          prior(normal(0, 5)), class="Intercept"),
          prior(gamma(2, 10), class="shape")
          )
```

Where the prior for the shape parameter is made more flexible by using a larger rate for the gamma distribution. Therefore, this allows the data to have more influence on the shape parameter.

For non-hierarchical models, the discrepancies between distributions with different priors is much more noticeable. Original non-hierarchical model's prior set has significantly higher peak, compared to the two other prior sets. Additionally, original model seems to be narrower. One thing that can be pointed out is that the original non-hierarchical model has covariate prior as $N(0, 5)$, which can be seen as an uninformative prior. For this reason the original prior set can cause the predicted model peak to vary more. The prior set with increased gamma rate parameter also had this same prior for covariates, but this change seems to have imposed stronger constraints on the model's variability.

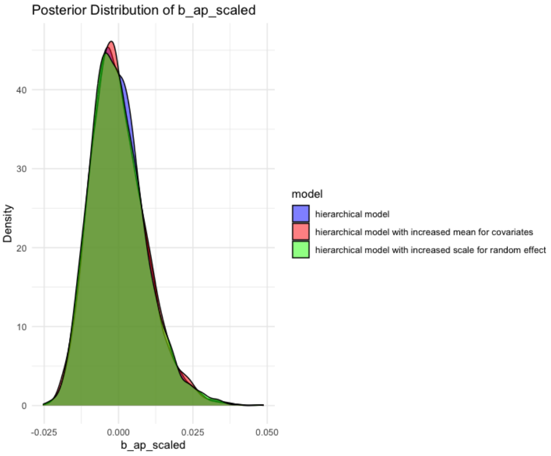

Figure 5: Posterior distribution of ap (scaled)

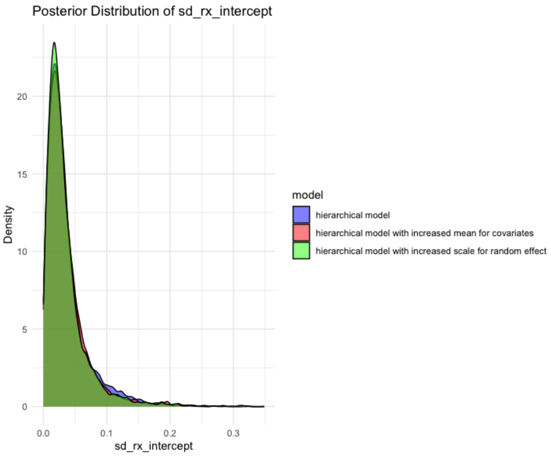

Figure 6: Posterior distribution of sd_rx random effect

When looking at the figures from hierarchical and non-hierarchical model, it seems that the hierarchical model exhibits more stability across different prior sets. In addition, the uninformative prior for original non-hierarchical model is enabling possible overfitting, as seen by the narrower posterior distribution below. Overall, these findings emphasize the importance of carefully selecting prior distributions, to ensure that they are not too informative nor too loose. It is important to mention that these priors were selected by experimenting, and thus it is entirely possible that other prior selections might show differing results with respect to priors, also for the hierarchical model.

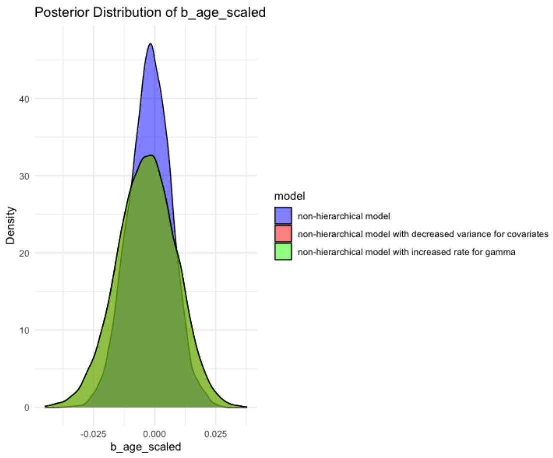

Figure 7: Posterior distribution of age (scaled)

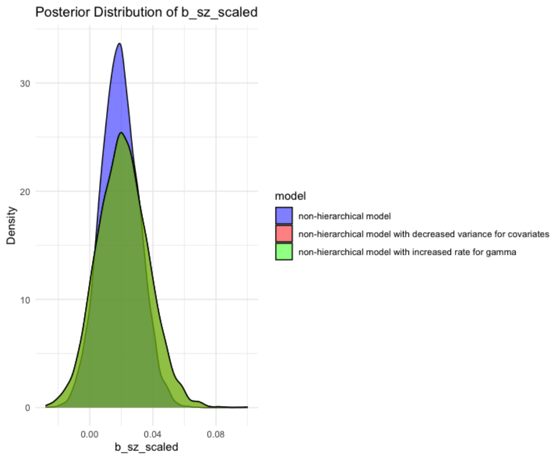

Figure 8: Posterior distribution of tumor size (scaled)

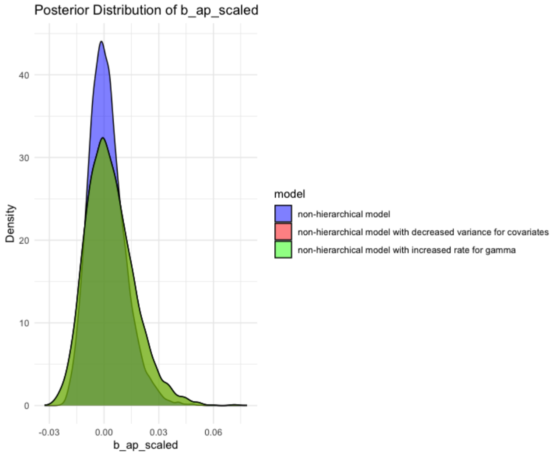

Figure 9: Posterior distribution of ap (scaled)

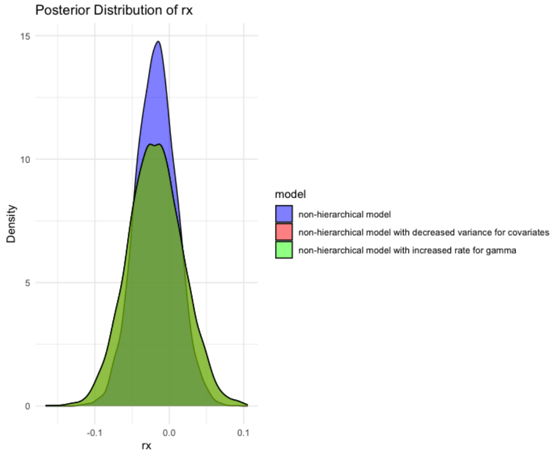

Figure 10: Posterior distribution of rx

Above, the blue distribution uses the original priors, red uses the prior set with modified class = ”*b*” prior, and green uses the prior set with modified class = ”*shape*” prior. Red distribution is not visible because it lies under green distribution, but it has similar distribution as the green.

5 Results and Analysis

5.1 Model comparison

For model comparison, we utilized the suggested version of cross-validation: Leave-One-Out cross-validation (LOO-CV). LOO-CV works by removing one observation from the dataset and then fitting the model to the remaining data, and then evaluating the model's accuracy to predict the left-out observation. This is then repeated for each observation in the dataset, and the model's overall performance is the average of all iterations. This is usually very computationally heavy, but because our dataset is relatively small, it applies very well.

Here we used the $loo\_compare()$ function from package loo in R, which measured the expected log pointwise predictive density (ELPD) between the models. Higher ELPD indicates better predictive performance. The results between the two types of models can be seen in Table 2 below.

Model	**ELPD difference**	**SE difference**
Hierarchical model	0.0	0.0
Non-Hierarchical model	-0.7	1.0

Table 2: Comparison of hierarchical and non-hierarchical models with LOO-CV

In the Table 2 the Hierarchical model functions as the baseline, with no difference in terms of predictive accuracy relative to itself. Thus, we want to investigate ELPD difference and SE difference, which represents the standard error of the difference in ELPD, in the non-hierarchical model row. Non-hierarchical model has -0.7 as the ELPD difference, which suggests that the non-hierarchical model has a worse predictive performance between these two, which would be against our speculations in the earlier sections. However, it is important to notice that as the difference is frankly small, and the relative standard error even larger (1.0) than the ELPD difference, it is not statistically significant. Therefore, we cannot really say which model is better based on these LOO-CV values alone. Especially, when the posterior predictive checks seemed to be against this.

6 Discussion

6.1 Problems encountered during analysis

There were few problems that arised when working on this analysis. First of all, as it was seen, the two models, hierarchical and non-hierarchical, fit quite similarly, making direct comparisons challenging. These models also struggled to accurately capture the true posterior distribution. This is mainly because our MCMC setup used unimodal distribution as the family, even though the target distribution showed multimodal characteristics. To address this problem, we tried to change the family to Weibull and lognormal mixture models, but even with these approaches, the models could not converge any better compared to our current models. The models produced with these families generated nearly indistinguishable posterior predictive checks. Mixture model with Weibull did show some improvements in the posterior predictive checks, but simultaneously got significantly worse R-hat statistic and ESS values, indicating that the model could not efficiently sample from the true data distribution, which is why we did not include it. Next, we tried to use Gaussian mixture models, which would allow essentially for more flexibility in capturing the multimodal characteristics of the data. Nonethless, this was not possible since it fails with the zero values from our censoring function (e.g., death events), as it calculates the logarithm of zero, which leads to negative infinity. Furthermore, mixture models are computationally intensive in an MCMC framework, making it impractical to experiment extensively with different priors or configurations. Therefore, due to simplicity but similar results, we decided to use simpler models.

6.2 Potential improvements

To address the challenges, employing more informative or tailored priors could enhance the model's ability to capture the unique shape of the data distribution. For instance, priors derived from exploratory data analysis may guide the posterior towards better parameter estimates. Additionally, the utilization of the mixture models could be investigated more thoroughly, to improve the generalizability of the posterior prediction.

7 Conclusion

This study applied both hierarchical and non-hierarchical Bayesian survival models to analyze time-to-event data from patients with stage 3 or 4 prostate cancer. The hierarchical model, which included random effects for medication dosage, accounted for group-level variations in survival outcomes, while the non-hierarchical model assumed independent observations. Despite the similarities in model fits, the models faced challenges in accurately capturing the true posterior distribution, particularly due to the unique and potentially multimodal nature of the data.

Model diagnostics, including R-hat statistics, effective sample size (ESS), and posterior predictive checks (PPCs), indicated that both models showed reasonable convergence and effective sampling. However, both models struggled to fully replicate the complex shape of the data, with issues such as failure to capture bimodal characteristics and discrepancies in the tail behavior.

Model comparison through Leave-One-Out Cross-Validation (LOO-CV) revealed no significant difference in predictive performance between the two models, suggesting that the complexity added by the hierarchical structure did not substantially improve model accuracy in this case. Hierarchical model however, seemed to have more robust nature, as was seen in the sensitivity analysis. Future work could focus on refining prior specifications or exploring mixture models to better capture the bimodal nature of the data. Tailoring priors based on exploratory data analysis may also help improve model performance and the ability to capture the true distribution more effectively.

8 Reflection

This project deepened our technical understanding of Bayesian survival models and highlighted the complexities of real-world data analysis, especially in a healthcare context where data irregularities are common. Through trial and error, we realized how crucial it is to carefully select priors that align with the underlying data and how to interpret the results critically when the model's assumptions do not perfectly match the data's structure.

References

[1] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. R. (2013) *Bayesian Data Analysis, Third Edition*

[2] Ibrahim, J. G., Chen, M. H., & Sinha, D. (2001). *Bayesian Survival Analysis.* Springer Series in Statistics.

[3] Sinha, D., & Dey, D. K. (1997). Semiparametric Bayesian Analysis of Survival Data. *Journal of the American Statistical Association*, 92(439), 1195-1212.

[4] Byar, D. P., & Green, S. B. (1980). The Choice of Treatment for Cancer Patients Based on Covariate Information: Application to Prostate Cancer. *Bulletin du Cancer*, 67(4), 477-490.

[5] Berry, D. A. (2006). Bayesian Clinical Trials. *Nature Reviews Drug Discovery*, 5(1), 27-36.

[6] Brawley, O. W. (2012). Prostate Cancer Epidemiology in the United States. *World Journal of Urology*, 30(2), 195-200.