KU LEUVEN

2022 BDA II KU Leuven Part 1

Master of Statistics & Data Science

Assignment – Part 1

Authors:
Luka Beverin (luka.beverin@student.kuleuven.be)
Nicolas Cozzi (nicolas.cozzi@student.kuleuven.be)
Marie Analiz April Limpoco (marieanalizapril.limpoco@student.kuleuven.be)
Khachatur Papikyan (khachatur.papikyan@student.kuleuven.be)

January 9, 2022

Contents

1 Objectives and Research Questions

To analyze the dataset containing the survival times of white grubs exposed to two different types of EPN species namely *S. sacchari* (SS) and *H. baujardi* (HB), we start by checking the Pseudo-Bayes Factor (PSBF) to select which distribution between Weibull and Lognormal best fits the data. Next, we build the model by including covariates such as the EPN species and grubs' size, as well as a random effect. Models with and without censoring will be explored. Convergence will be checked for each model, and model selection will be performed based on DIC. Posterior predictive checks will then be implemented to evaluate the chosen model. Finally, we interpret the model and use it to address the research questions.

2 Objective 1: To determine which of the grub worms' survival distributions, Log-normal and Weibull distributions, best fits the data

In this section we consider the midpoint of the lower and upper limit of the interval-censored survival time as the grub worms' survival time. We treat the data set as a one-sample data set with 140 observations and fit two models, namely the Log-normal and Weibull, to analyse which distribution best fits the data. Let T denote the time to death of a grub. Furthermore, we assume that censored data D are observed, unless otherwise specified.

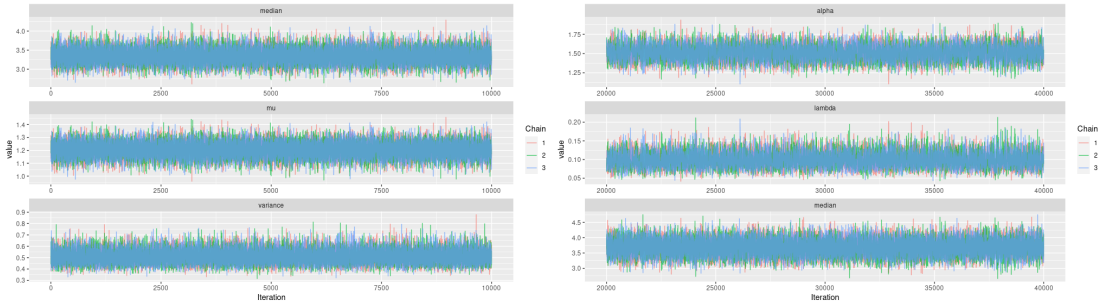
Fit the log-normal distribution Here we fit the log normal distribution with parameters μ and τ. Analysis of our log-normal model is very similar to the treatment of normal data – in this case, the logarithmic transformation of the response is assumed to be normally distributed. For this reason, suitable priors for this log-normal model are a normal distribution for $\mu \sim N(0, 0.0001)$ and an independent Gamma distribution for $\tau \sim Gamma(0.0001, 0.0001)$. These priors are referred to as the standard improper reference (SIR) priors and they are often used because of their ability to mimic the usual frequentist results seen for normal data. Three chains with 10 000 iterations each were instantiated and a burn in rate of 2000 iterations was used for this model.

$$\texttt{t[i]} \sim \texttt{dlnorm(mu,tau)} : f(t|\mu,\tau) = \frac{\sqrt{\tau}}{\sqrt{2\pi}t}\exp\{-\frac{\tau}{2}(\log(t)-\mu)^2\}$$

Fit the Weibull distribution In this part of the task we assume that $T_1, ..., T_{140}$ follows a Weibull distribution with parameters α and λ. For the Weibull model we consider a lognormal specification on α and a Gamma prior on λ. The priors are not necessarily definitive priors for this model – as assessed later regarding priors for parameters and coefficients. Here, three chains of 40 000 iterations each were instantiated with a burn in rate of 20 000 samples.

$$\texttt{t[i]} \sim \texttt{dweib(alpha,lambda)} : f(t|\alpha,\lambda) = \lambda\alpha t^{\alpha-1}e^{-\lambda t^{\alpha}}$$

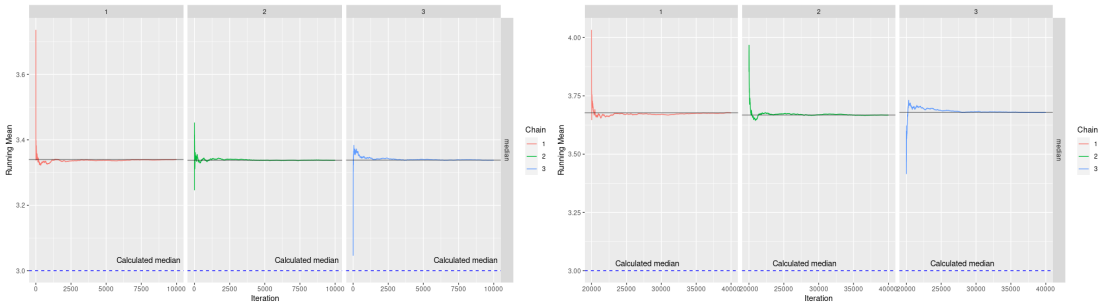
Check convergence using classical diagnostics In this section we apply various diagnostic techniques to check whether or not our models have converged. Figure 1 plots the parameter value at each time point against the iteration number. If our models converge, then the trace plot will move around the mode of the distribution. For all parameters of the Log-normal model the three chains mix well and stability of the parameter estimates across the iterations is visible. In contrast, the trace plot of the Weibull does not mix as well and requires much more iterations to show stability.



(a) Log-normal (b) Weibull

Figure 1: Trace plot

Figure 2 displays the running mean of the median (considered as a parameter in our model specifications) for both the Log-normal and Weibull model, which allows us to check whether the chain is slowly or quickly approaching its target distribution. The running mean plots show us similar effects to that of the trace plot. The calculated median of grubs' survival time is 3 days and both models overestimate this median, however, the Log-normal is considerably closer. The Weibull model also takes more iterations to reach stability when compared to the Log-normal model. The drawback of these plots are that they only look at the mean of the parameters and hence are inadequate as they oversimplify diagnostic checks.

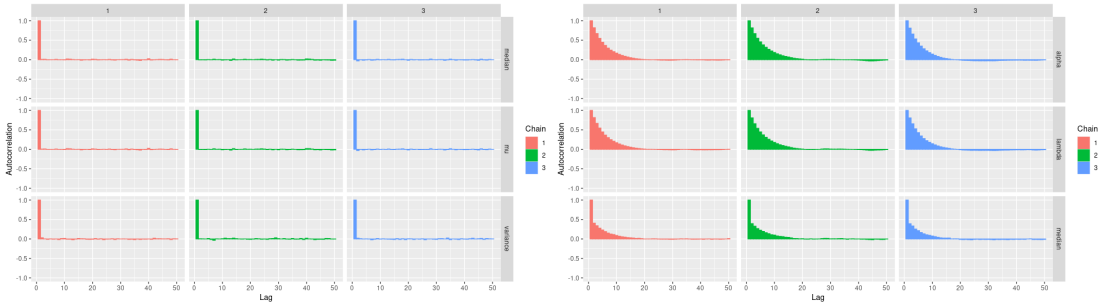


(a) Log-normal (b) Weibull

Figure 2: Running mean plot

The auto-correlation plot for both models is depicted in Figure 3. The Log-normal model goes to zero very quickly, suggesting that low correlation of the

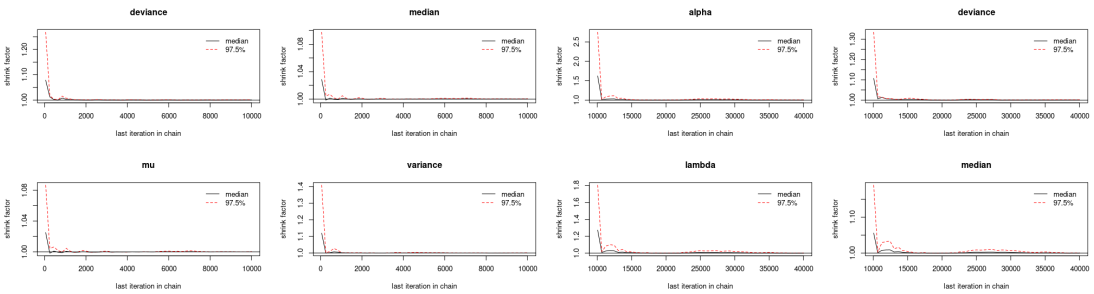
chain's successive lags. A case of consistent high serial correlation is visible in the Weibull model, indicating that the parameters in the model may be highly correlated. Since the Weibull model has high auto-correlation, it requires more iterations to explore the entire posterior distribution.



(a) Log-normal (b) Weibull

Figure 3: Autocorrelation plot

Figure 4 shows the Brooks-Gelman-Rubin test for both the Log-normal and Weibull models. The test is based on the 3 parallel chains and compares the within and between chain variances for each parameter. The parameters of the log-normal model has the better marginal posterior densities, as indicated by the chain being more stable around a shrink factor (R) of 1.



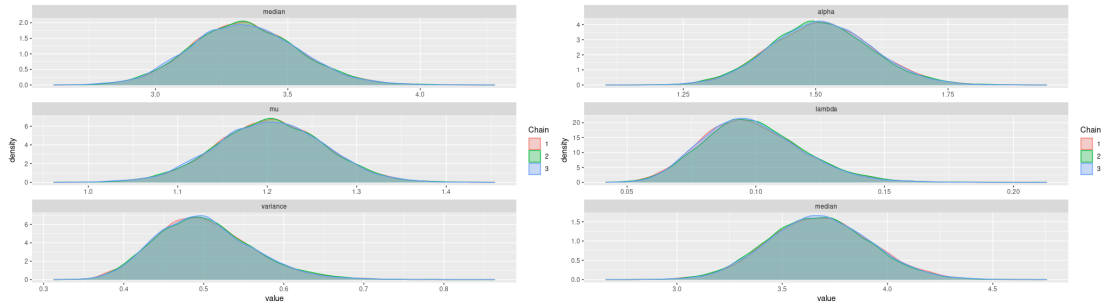
(a) Log-normal (b) Weibull

Figure 4: Brooks-Gelman-Rubin diagnostic plot

The Brooks-Gelman-Rubin test can also be used to assess how long the sampler should run until we produce decent approximation. The Weibull model requires almost triple the amount of iterations of the Log-normal model to reach a stable approximation.

Perform posterior predictive checks In this section we perform posterior predictive checks on both models to highlight whether or not our models provide valid predictions and estimations about the reality of grubs' survival data. A kernel density plot is a useful diagnostic tool that is used to check if our models make sense since they visualise non-convergence. In the case of non-convergence,

we would expect multi-modal distributions of the posterior. For both models, Figure 5 shows the kernel density across their respective chains. The density plots suggest that convergence has been met in both models.

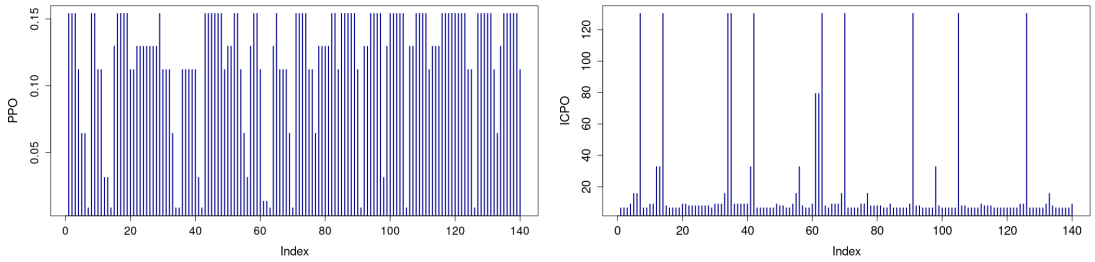


(a) Log-normal (b) Weibull

Figure 5: Density plots

A satisfactory kernel density plot would appear to be bell-shaped for which we do not require the shape to be symmetric.

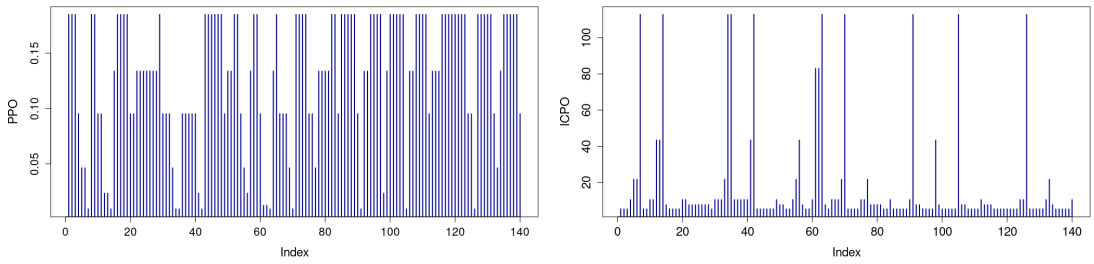
Outlying/influential Observation(s). The posterior predictive ordinate (PPO) is the density of the posterior predictive distribution evaluated at an observation and can be used as a tool for outlier detection. Low values of PPO indicate that the observation is at the tail area of the density function, thus suggesting potential outliers when the PPO are extremely low (relative to other evaluations).



(a) PPO (b) $1/\hat{\text{CPO}}$

Figure 6: Index plots of the Weibull model

Figures 6a and 7a display the PPO index plots of the Weibull and Lognormal models respectively. The PPO values in the Weibull model are slightly lower than that of the Lognormal model, suggesting that the Weibull may be a worse fitting model. The inverse of CPO_i is also plotted to highlight surprising observations. The index plots of $1/\hat{\text{CPO}}$ (Figures 6b and 7b) for both models indicate numerous outlying observations, which all correspond to a survival time of 12 days.



(a) PPO (b) 1/CP̂O

Figure 7: Index plots of the Lognormal model

The CPO_i can also be used as means for evaluating comparative adequacy among several models by taking the sum of the log conditional predictive ordinates, with the idea of selecting the models with the highest values. The CPOs can then be used to look at local fit, or one can define an overall score for each model, as we have done in Table 1.

	Log-Normal	Weibull
$\sum_{k=1}^{N} \log \text{CPO}_k$	**-320.1491**	-325.1017
DIC	**638.5**	648.3
Median	3.339	3.675

Table 1: Results

Table 1 displays the posterior mean of the median for both models, as well as the calculated deviance information criterion (DIC). The Log-normal model has the smallest DIC value, suggesting that the model would best predict a replicate data set which has the same structure as that currently observed. A common rule of thumb is to rule out the model with a lower DIC value, however, when comparing relatively simple models, pseudo-Bayes factor (PBF) are often favored. Disregarding interval-censored character of the survival times, the PBF suggests that the Lognormal distribution gives a better fit.

RQ1.1 Which model best fits the data?

To determine which distribution best fits the data when censoring is acknowledged, we compute the Pseudo-Bayes Factor (PSBF) by monitoring the inverse of CPO for each distribution and then performing post-processing to eventually calculate the $\log(PSBF_{WL})$. Here, we compare two classes of models before we propose covariates or clustering, so we only look at which distribution is well supported by the data. The results indicate evidence in support of a Weibull distribution such that $\log(PSBF_{WL}) = 4.371175$. In summary, the Weibull model is preferred for the interval-censored characteristics of the survival time while the Lognormal is a better fit when assuming the midpoint of the intervals as the actual survival time.

Although Weibull-based models might not necessarily be the most correct distribution since we only compared two among many, PSBF is still a useful and logical tool when choosing between classes of models based on the data even without parameterizations yet. Hence, we continue our analysis based on the Weibull distribution, although we will still mention some results for the Lognormal-based models.

3 Objective 2: To determine which of the EPN species acts the fastest in killing the white grubs

3.1 Initializations and Priors

The choice of starting values and priors is important in a Bayesian analysis. Poor initializations may lead to slow convergence, while carelessly choosing priors, albeit weakly informative, may lead to an improper posterior. Hence, for the initial values, we made use of estimates coming from the frequentist counterpart of the Bayesian models we use here. Meanwhile, we made use of weakly informative priors to represent our lack of expert knowledge. In particular, the coefficients of the covariates have diffuse normal priors; i.e. $\beta_i \sim N(0, 10000)$. On the other hand, the shape parameter of Weibull (ρ) and the precision parameter of Lognormal (τ) were given Gamma priors (Gamma(1, 0.0001)). However, the prior for the second level variance parameter which is associated with the random effect had to be chosen according to the recommendation of Spiegelhalter et al. [2004]; i.e., a uniform prior on $\sigma_\theta \sim U(0, c)$. This choice avoids having an improper posterior.

3.2 Convergence Tricks

To improve convergence, we employed tricks as suggested by Lesaffre and Lawson [2012]. For each model, we ran 3 chains instead of just a single chain to exploit mixing of chains and avoid getting stuck around a local mode. We chose starting values that are more plausible, as explained in the previous section. The covariate grubs' size was standardized as well. Stationarity was checked using both visual (trace plot, running mean plot, Q-Q plot) and formal (Geweke and BGR) diagnostics, but we will only highlight some results that are worth mentioning. We first run iterations until the trace plot appears satisfactorily stationary based on the thick pen test, before proceeding with the other visual and formal diagnostics. For the BGR diagnostic we considered the second version which does not assume normality for the variance parameter. Finally, we also assessed accuracy by ensuring that the Monte Carlo (MC) errors are at most 5% of the posterior standard deviation of a parameter.

3.3 Model with EPN Species as Covariate

First take the midpoint of the lower and upper limit of the interval-censored survival time and ignore the clustering. For the worms that survive, you take the

mid-point of the lower limit and day 12. Then you look at the median death time per EPN species to assess which of the two EPN species acts the fastest in killing the white grubs.

Entomopathogenic nematodes (EPNs) are a group of threadworms that can be used to effectively eliminate grub worms. In this section we compare two of these species, namely S. sacchari (SS) and H. baujardi (HB), in their ability to efficiently kill white grubs. For the time being, we disregard the effect of grub size and inter-plate variability. Figure 8 displays the frequency of survival time of the grubs for both species. The use of species SS to kill white grubs corresponds to a median survival time of 5 days, with most of the grubs dying between 2 and 6 days. However, there are 9 instances in which the grubs survived for 12 or more days. When using species HB, results suggest that most white grubs die on the third day. In contrast to species SS, there are much fewer grubs that survive past 6 days.

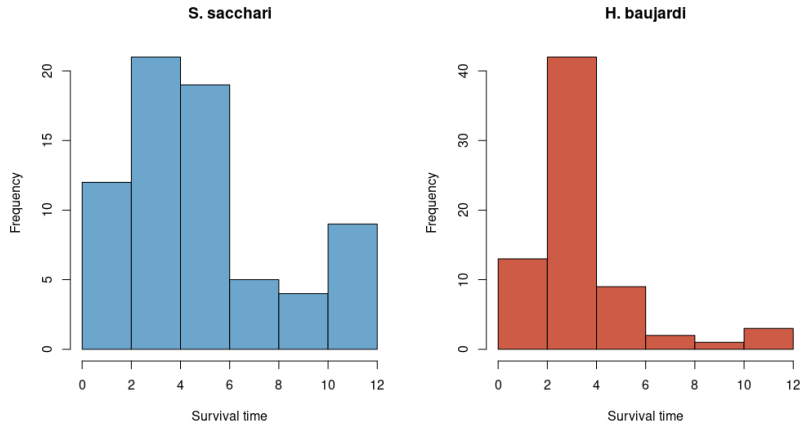


Figure 8: Histogram of survival times per species

The median survival times of white grubs when using species SS and HB are 5 and 3, respectively. The average survival time for species HB is also lower than that of species SS, with a value of 3.47, compared to a mean survival time of 4.96 for species SS.

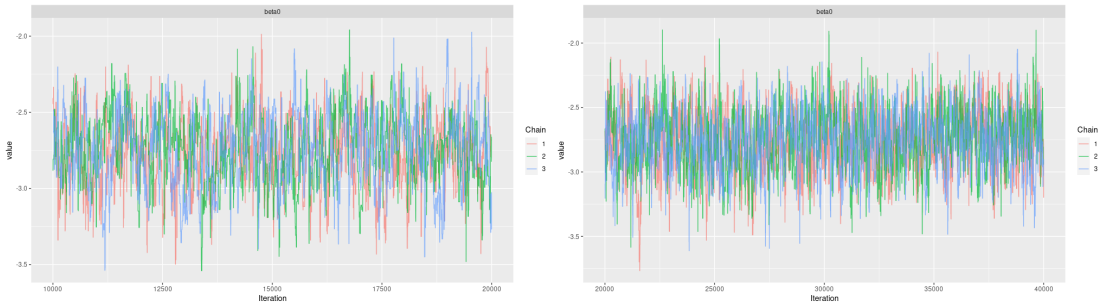
We next specify the model based on a Weibull distribution. Assuming no censoring and ignoring clustering, the model is

$$t_i \sim Weibull(\rho, \mu_i), \quad \log(\mu_i) = \beta_0^* + \beta_1^* epn_i, \quad Md_i = \left(\frac{\log(2)}{\mu_i}\right)^{(1/\rho)}$$

where t_i is the survival time of grub worm i, ρ is a shape parameter, μ_i is the nonbaseline hazard component of the hazard function $h(t_i) = h_0(t_i)\mu_i$ such that μ_i depends on the species epn_i (0 - SS; 1 - HB), while Md_i is the median survival time with respect to the species. Since the effects β_0^* and β_1^* pertain to the hazard, we note that the coefficients of $E[\log(t_i)] = \beta_0 + \beta_1 epn_i$ can be computed from

the hazard function by using $\beta_i = -\beta_i^*/\rho$.

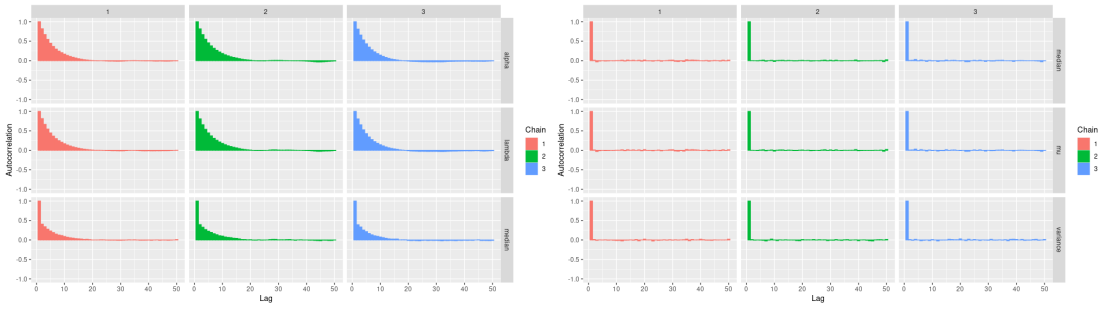
After initiating 3 chains for 20,000 iterations with 10,000 burn-in for the Weibull-based model, the trace plots of parameters suggested stationarity except for β_0^*, as shown in Figure 9a. Hence, we ran longer iterations of up to 40,000 and set 20,000 as burn-in. The trace plot in Figure 9b shows a more stable sampling of the posterior, unlike in the previous trace plot wherein we can still see upward and downward motion of the entire set of chains, suggesting slow exploration of the posterior. Since the mixing of the latter trace plot is adequate even though the individual chains appear nonstationary, we proceeded with other visual means of checking convergence and then performed formal diagnostics.



(a) 20k iter with 10k burn-in (b) 40k iter with 20k burn-in

Figure 9: Trace plot of β_0 for Weibull-based model

The autocorrelation plots as shown in Figure 10a further confirm the slow mixing rate of the Weibull model, implying that it takes time before the initial dependent values are forgotten, in contrast to the Lognormal model (10b). A case of consistent high serial correlation is visible in the Weibull model, indicating that the parameters in the model may be highly correlated; due to this, it takes more iterations for sampling methods to explore the entire posterior distribution.

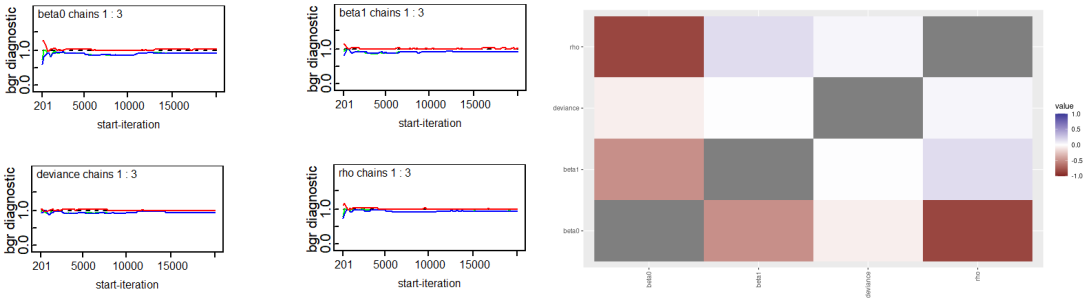


(a) Weibull (b) Lognormal

Figure 10: Autocorrelation plots of the model parameters

Although we also checked the global and dynamic Geweke diagnostics whose interpretation did not deviate much from stationarity except for a few z-scores outside

the [-1.96,1.96] interval, we focus on the BGR diagnostic since the Geweke diagnostic does not consider the advantage of having multiple chains that lead to better mixing, thus achieving stationarity faster. This is evident in the trace plots where we see that individually, the chains may not yet be stable but if taken altogether, stability can be observed. Indeed, the BGR diagnostics as displayed in Figure 11a show stable values around 1 for the interval-based potential scale reduction factor (PSRF) (red) while the ratios involving the length of total-chain interval (green) and the average length of the within-chain intervals (blue) have stabilized as well.



(a) BGR diagnostics

(b) Cross-correlation plot

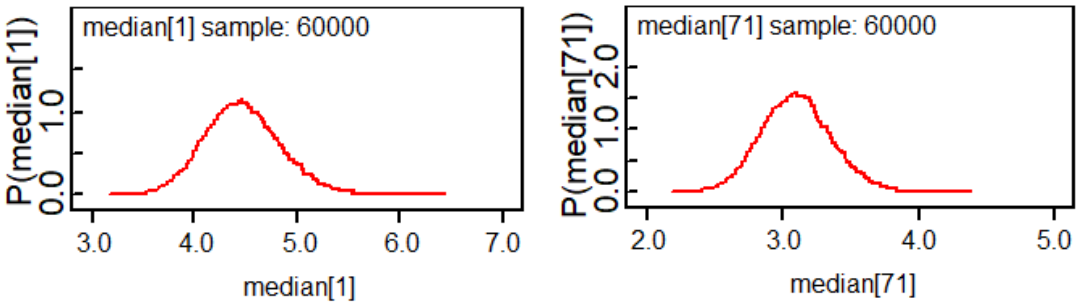
Figure 11: Weibull parameters

The cross-correlation plot displayed in Figure 11b reveals the possible reason for slow convergence of β_0^* in the Weibull model; i.e., a negative correlation (-0.52) between β_0^* and β_1^*.

After ensuring stationarity, we check the MC errors for accuracy; all errors are less than 5% of the posterior standard deviation. Table 2 shows the posterior summary measures and the effective sample size (ESS) per parameter obtained after reaching convergence while Figure 12 displays the marginal posterior density of the median per EPN species, where we see that HB has shorter survival times centered a little above 3 days compared to that of SS centered between 4 and 5 days. We find from the ESS that although the iterations after burn-in reached 20,000, only a small amount of information equivalent to independent elements is contained in the chains. For instance, the information contained in the chain used to estimate β_0^* is just equivalent to only 905 independent elements, and not 20,000.

Table 2: Posterior summary measures for the Weibull-based model

Parameter	mean	sd	MC error	95%CrLL	median	95%CrUL	ESS
β_0^*	-2.755	0.2332	0.006564	-3.226	-2.753	-2.305	905.3
β_1^*	0.5803	0.1728	0.003178	0.2474	0.5795	0.9249	2830.2
ρ	1.598	0.09964	0.002439	1.406	1.597	1.795	1283.4
Md_{SS}	4.465	0.3664	0.008526	3.787	4.451	5.223	1545.8
Md_{HB}	3.105	0.2556	0.003246	2.628	3.099	3.63	7019.3



(a) SS (b) HB

Figure 12: Marginal posterior density of the median survival time for SS (left) and HB (right) for the Weibull-based model

RQ2.1 Is EPN species a significant covariate in our model?

In addition to the Weibull, a model based on the Lognormal distribution was also formulated, and both models agree that grub worms applied with HB has a shorter median survival time compared to those applied with SS. Moreover, β_1 which is the effect of the EPN species is significant since the 95% credible intervals do not contain 0. However, it is positive for the Weibull-based model because the interpretation affects the hazard such that HB whose EPN species code is 1 imposes a *higher hazard* or *relative risk* than SS ($epn_i = 0$). If we perform the conversion for the Weibull coefficients, we have: $\beta_0 = -\beta_0^*/\rho = -(-2.755)/1.598 = 1.7240$ and $\beta_1 = -\beta_1^*/\rho = -(0.5803)/1.598 = -0.3631$. Hence, the Weibull survival time model is $E[\log(t_i)] = 1.724 - 0.3631 epn_i$ where the negative coefficient for the EPN species covariate now implies *shorter mean survival time* for worms exposed to HB compared to those exposed to SS.

RQ2.2 Which species corresponds to a lower median survival time?

From the posterior estimates, we can say that HB corresponds to a lower median survival time (3.105 days) compared to SS (4.465 days). To test whether there is at all a difference in the survival times between the two species, we make use of the **log rank test**. The test is non-parametric and suitable to use when the

data is right skewed and censored, as is the case in our experiments. The null hypothesis of the log rank test is that there is no difference between the species in the probability of an event (here the death of a white grub) at any time point. Applying the log rank test to our data, we observe a p-value of $p = 0.004$, suggesting that that there ***is a significant*** difference between the two species regarding the survival time of white grubs.

Since the log-rank test is based on the same assumptions as the Kaplan-Meier (KM) survival curve, we also plot the KM survival curve (see Figure 13) to obtain a better understanding of the survival probabilities at different time points.

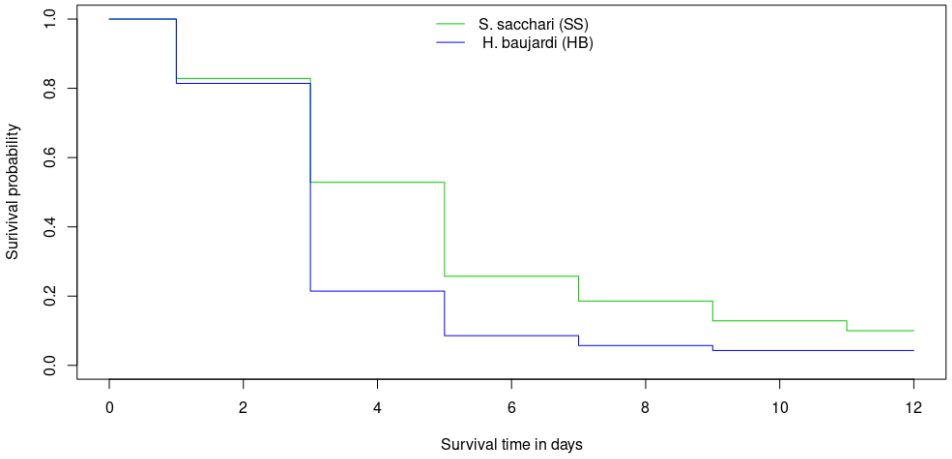


Figure 13: KM Survival curve

From day 1 to 2, there does not seem to be any significant difference between the two species with respect to the survival probabilities of white grubs. It is only at day 3 when the survival plots begin to diverge. On day 3, the survival probability of a grub under species SS is just above 50%, while at the same time point and under species HB, the survival probability reduces to around 20%. This noticeable difference in survival probabilities continues until day 12, the last observed time point. In other words, at each time point after day 3, the survival probabilities of grubs attached to species SS is higher than that of species HB.

Based on the analyses of Section 3, the results of the log rank test and interpretation of the KM survival plot, we conclude that species H. baujardi acts the fastest in killing white grubs. These conclusions do not take into account the effect of other factors such as grub size and inter-plate variability.

4 Objective 3: To determine whether or not the grubs' size predicts the survival of the worm

RQ3.1 Is grub size significant in our models?

4.1 Model with EPN Species and Grubs' Size as Covariates

Evaluate the effect of the covariate grubs' size.

To evaluate the effect of grubs' size on the survival of grub worms, we included the standardized values as a covariate to the previous model. The standardization was performed to avoid convergence problems. The same number of iterations and burn-in was used and visual and formal tests were employed to ensure convergence. MC errors are also less than 5% of the posterior standard deviations. Table 3 provides the posterior summary measures for the Weibull-based model where we find that grubs' size effect (β_2^*) is not significant.

Table 3: Posterior summary measures for the Weibull-based model with grubs' size

Parameter	mean	sd	MC error	95%CrLL	median	95%CrUL	ESS
β_0^*	-2.781	0.2303	0.006787	-3.239	-2.775	-2.34	962.0
β_1^*	0.5965	0.1727	0.003214	0.2584	0.5957	0.936	2875.8
β_2^*	0.1315	0.09148	8.659E-4	-0.04792	0.1313	0.3098	13214.0
ρ	1.61	0.09966	0.00254	1.42	1.608	1.809	1430.8

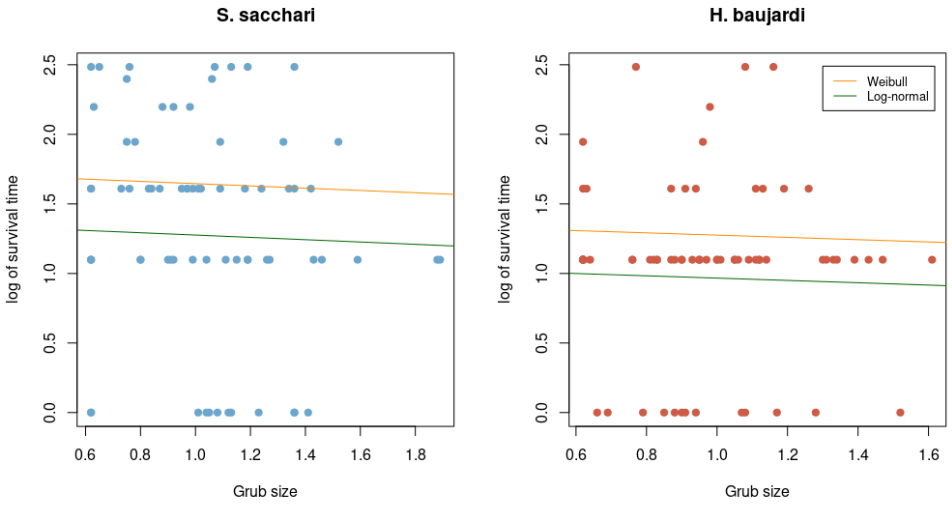


Figure 14: Fitted models per EPN Species with Grub size as a covariate

Figure 14 displays fitted linear models on grub size using the estimated parameters obtained from Table 3. The straight line is defined by $E[\log(t_i)] =$

$\beta_0 + \beta_1 epn_{ij} + \beta_2 size_{ij}$, where *epn* is a dummy variable (either 0 or 1) and $\{\beta_0, \beta_1, \beta_2\}$ are parameters estimated by the Weibull (orange) and Log-normal (green) models. This confirms that the expected log survival time is not affected even if grubs' size is increased.

4.2 Model with Random Effect

Consider the clustering by including a random effect, i.e., the variation of location parameters across replicate plates. What does the random effect's variance component estimate suggest about the "clustering" of responses?

The model including a random effect based on the Weibull distribution is

$$t_{ij} \sim Weibull(\rho, \mu_{ij}), \qquad \mu_{ij} = \exp\left(\theta_i + \beta_1^* epn_{ij} + \beta_2^* size_{ij}\right)$$

where $i = 1, ..., 20$ represents the replication plates each with grub worms $j = 1, ..., 7$ and $\theta_i = \beta_0 + b_i$ is the random effect per replication plate i. Here, we implement hierarchical centering by including in the first level of the model $\theta_i \sim N(\beta_0, \sigma_\theta^2)$ centered on the overall effect β_0 instead of an alternative version where there is an overall effect β_0 with its own weakly informative prior and a random effect $b_i \sim N(0, \sigma_b^2)$ centered on 0. By doing this, we expect to have improved mixing that leads to smaller MC errors. Indeed, when we tried both versions of the model, the MC error for say β_0^* in the uncentered is 0.01719 while in the centered model, it is 0.01134.

Stationarity for the Weibull-based model was achieved after 40,000 iterations with 20,000 burn-in. Similarly as with the previous Weibull-based models, mixing is slow as observed from the autocorrelation plots. Nevertheless, the other plots suggest convergence. The dynamic Geweke diagnostic as well as the non-parametric version of the BGR diagnostic is also consistent with the visual checks. MC errors are less than 5% of the standard deviations. Table 4 presents the posterior estimates of this model.

Table 4: Posterior summary measures for the Weibull-based model with random effect

Parameter	mean	sd	MC error	95%CrLL	median	95%CrUL	ESS
β_0^*	-3.049	0.3671	0.01134	-3.809	-3.034	-2.373	661.4
β_1^*	0.4876	0.4127	0.01581	-0.3254	0.4819	1.323	379.1
β_2^*	0.101	0.09787	0.0009744	-0.09334	0.1019	0.2906	10769.5
ρ	1.875	0.1323	0.003728	1.621	1.872	2.143	1091.4
σ_θ^2	0.784	0.4367	0.007734	0.2073	0.6972	1.862	3462.9

Contrary to the previous models without random effects, the results now indicate that EPN species does not significantly influence the survival time of grub worms. The variance of the random effect σ_θ^2 does not differ much with (0.784) or without (0.8316) covariates, implying that although the covariates explain a portion of

the variability, there is still a larger proportion explained by the clustering. These results will be investigated further since it is a bit strange for the EPN species to become uninfluential in the presence of random effects. We will say more about the variance component estimate after we have investigated outlying observations.

4.3 Outlying/Influential Observations

Check whether there are outlying/influential observation(s). If so, give possible reasons for the outlying observation(s)

Once again, likelihood-related computations for outlier detection are implemented (PPO and CPO). In case of the lognormal model, for clearer illustration, results are displayed both for a sub-optimal version – lognormal distribution including *grubsize*, excluding random effects - and for one with lower deviance – lognormal distribution with covariates, plus random effects -; best-performing models are a topic to be addressed later. In this context, the detection of outliers can be altered when disregarding variance between plates in the distribution. Values of 1/CPO confirm the previous statements.

Plots of PPO values for each of the two models are displayed below in Figure 15; a simple standardization of $\frac{\text{PPO}_i}{\max\{\text{PPO}_i\}}$ was performed for interpretability. As seen, the model without random effects has a few variables for which the model seems very suitable, while most variables would be located closer to the tail end of the distribution. In contrast, the model with random effects – which has lower deviance – displays a more even posterior prediction performance across the sample dataset.

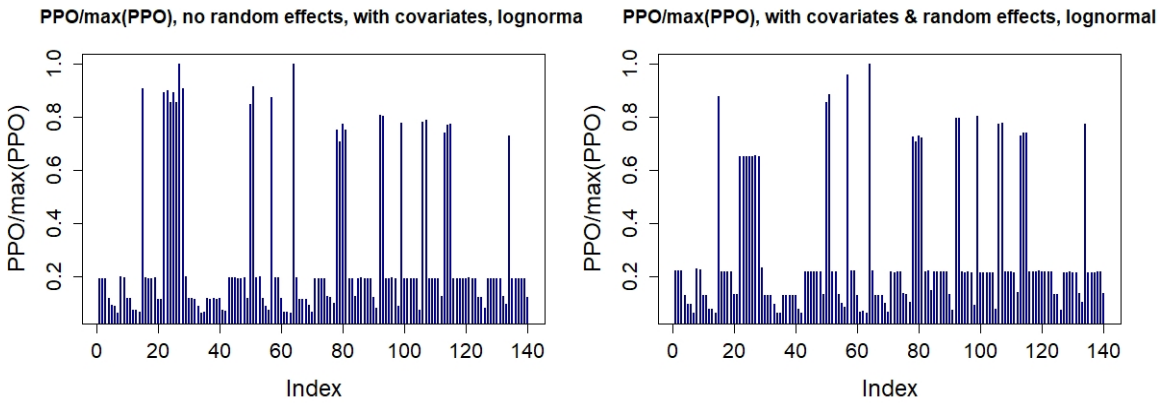


Figure 15: Standardized PPO values of lognormal models with covariates, without random effects (left) vs. with random effects (right)

For determining the cause of lower PPO and higher 1/CPO values of the lognormal model with random effects, the sample is ordered according to the respective

1/CPO values from the MCMC simulations. After performing this task, it is observed that 1/CPO is highest for observations with higher values of (midpoint) mortality time; thus, the model would be better fit for lower values, and decreasing its performance for higher values.

When assessing PPO and 1/CPO values of the Weibull (plotted in Figure 16), one observes that it detects as outliers only those few observations which responses deviate from other samples within a plate – that means, high survival times (12 or 9) among other samples with a much lower response within the plate (mainly 1) -; the lowest 1/CPO values correspond to the only plate with all samples having the same outcome (of 1 in that case). Thus, the Weibull model seems to detect within-group variance more strongly.

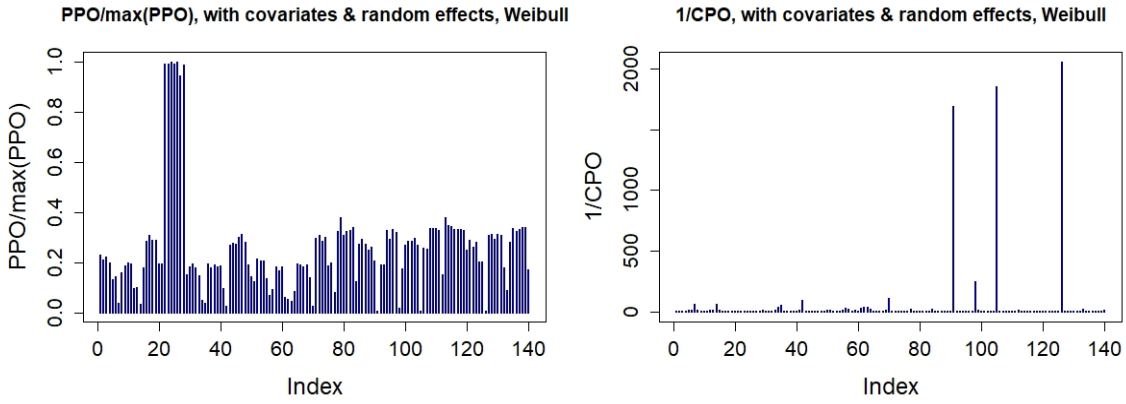


Figure 16: PPO (left) and 1/CPO (right) values of Weibull model, with covariates and random effects

What seems to be a common pattern, however, is that neither model seems to account for grub size in outlier detection. This statement seems to support the hypothesis that this covariate would not have statistically significant explanatory power in predicting the survival time of white grubs.

4.4 Robust Distribution for the Random Effect

Check the distribution of the random effect and propose a robust distribution for the random effect

The initial prior for the random effect was specified as a normal centered on the overall effect β_0^*. However, the presence of outliers as seen in the previous section requires a more robust distribution especially since Weibull is sensitive to outliers. Hence we propose the t_2-distribution; i.e. $\theta_i \sim t_2(\beta_0^*, \sigma_\theta^2)$. The family of t-distributions is a robust alternative to the normal by virtue of the thicker tails, making them less sensitive to outlying observations. Table 5 presents the results of running the model with EPN species, grubs' size, and random effects, where we find that EPN species is now significant, which is much more sound with original descriptive statistics. Grubs' size is still insignificant, thus maintaining

our statement about the inability of grubs' size to predict the survival of the worm. Meanwhile, the variance component estimate σ_θ^2 is now smaller (0.1904) compared to the previous result (0.784), indicating less variability in survival times across plates. Moreover, σ_θ^2 without covariates is 0.3388(sd = 0.248) with a 95% CrI of [0.06077,0.9932], which is a little less than twice the value when there are covariates, implying that the covariate (in this case EPN species only since grubs' size is not significant) explains a considerable proportion of the variability.

Table 5: Posterior summary measures obtained using a robust distribution for the random effect

Parameter	mean	sd	MC error	95%CrLL	median	95%CrUL	ESS
β_0^*	-3.357	0.3415	0.01254	-4.051	-3.347	-2.716	387.4
β_1^*	0.7713	0.3181	0.01074	0.116	0.7723	1.394	641.6
β_2^*	0.1064	0.09653	9.314E-4	-0.08238	0.106	0.2972	11272.2
ρ	1.874	0.128	0.004178	1.632	1.87	2.131	803.6
σ_θ^2	0.1904	0.1702	0.00328	0.01774	0.1431	0.6364	2766.4

4.5 Models with Censoring

Secondly, you repeat the analyses considering the interval-censored character of the survival times. Are your conclusions still the same?

We repeat the analyses but this time considering the presence of interval- and right-censored events. We begin by grouping together observations with interval censoring and then those with right censoring for easier code implementation in Openbugs. We also use the robust distribution for the random effect as proposed in the previous section. Table 6 presents the posterior summary measures for the Weibull-based models. In comparison to the previous analyses where the midpoint was used as exact observed survival times, the conclusions when considering censoring are not different. In particular, EPN species is significant for all models such that HB species kills grub worms faster than SS. Grubs' size does not affect grub worms' survival. Lastly, the variance component estimates are close to the values estimated when censoring was ignored, still implying that the variation of responses across plates ($\sigma_\theta^2 = 0.2089$) is less than the variation within each plate, which is equivalent to $(1/\rho)^2 = \sigma^2 = 0.3530$.

Table 6: Posterior summary measures for the Weibull-based model with censoring

Parameter	EPN only		EPN & size		EPN, size, & random effect	
	mean(sd)	95% CrI	mean(sd)	95% CrI	mean(sd)	95% CrI
β_0^*	-2.534(0.2402)	[-3.007,-2.077]	-2.565(0.2486)	[-3.06,-2.09]	-3.118(0.3732)	[-3.909,-2.432]
β_1^*	0.5911(0.1819)	[0.2348,0.9552]	0.6108(0.1843)	[0.2487,0.9626]	0.8306(0.3261)	[0.197,1.516]
β_2^*	–	–	0.132(0.09574)	[-0.05642,0.318]	0.1064(0.1035)	[-0.0962,0.3081]
ρ	1.432(0.1045)	[1.232,1.43]	1.446(0.1072)	[1.241,1.661]	1.683(0.1408)	[1.42,1.976]
σ_θ^2	–	–	–	–	0.2089(0.1815)	[0.02434,0.6856]

5 Model Selection and Checking

5.1 Selection using DIC

We utilized DIC to select the final model among those that recognize censoring. The computation included only the second half of each chain; i.e., after the burn-in was removed. Furthermore, for the models with random effects, we note that the DIC generated by Openbugs is a conditional one and thus, measures the predictive ability of the model for future observations of the same plate. Table 7 presents the results.

Table 7: Information theoretic measures for the models

Model	Weibull			
	Dbar	Dhat	p_D	DIC
EPN only	449.1	446.2	2.976	452.1
EPN & size	448.3	444.3	3.998	452.3
EPN, size, & random effect	409.5	393.3	16.27	425.8
EPN & random effect	409.3	393.6	15.64	424.9

From the results, we find that the models with random effects have the lowest DIC values. Even though the difference in DIC is not substantial, we select the model with EPN species and random effects only since grubs' size is not significant.

Table 8: Posterior summary measures for the selected model

Parameter	mean	sd	MC error	95%CrLL	median	95%CrUL	ESS
β_0^*	-3.071	0.3963	0.01541	-3.902	-3.056	-2.327	294.9
β_1^*	0.7896	0.3697	0.01384	0.0949	0.7754	1.533	484.8
ρ	1.67	0.1403	0.004807	1.399	1.667	1.951	638.9
σ_θ^2	0.2244	0.1995	0.004209	0.02429	0.1711	0.745	2334.5

5.2 Sensitivity Analysis

5.2 Varying the priors Here, we vary the priors and check whether the conclusions change. In particular, we consider weakly informative Cauchy priors instead of vague normal priors for the regression coefficients, and an inverse gamma prior for the level-2 variance parameter instead of a flat uniform prior. Since Openbugs does not provide a Cauchy distribution, we make use of the t_2-distribution but following the suggestion from Gelman et al. [2008] for a logistic regression model; i.e. centered on 0 with scale equal to 10 for the constant term β_0 and 2.5 for the regression coefficient. The posterior estimates are shown in Table 9 where we find that although the estimates have changed, the inference remains the same; i.e. EPN species significantly affects survival time, and there is a clustering effect by virtue of the replication plates.

Table 9: Posterior summary measures after varying the priors

Parameter	Priors 1		Priors 2	
	mean(sd)	95% CrI	mean(sd)	95% CrI
β_0^*	-3.071(0.3963)	[-3.902,-2.327]	-2.821(0.32)	[-3.477,-2.223]
β_1^*	0.7896(0.3697)	[0.0949,1.533]	0.5977(0.2846)	[0.03914,1.173]
ρ	1.67(0.1403)	[0.02429,0.745]	1.596(0.1285)	[1.351,1.858]
σ_θ^2	0.2244(0.1995)	[0.02429,0.745]	0.1444(0.1388)	[0.007556,0.5136]

5.2 Sensitivity to outliers The purpose of this section is to check how robust the models' conclusions are for each of the two complete models (Weibull and lognormal, with covariates and random effects), when one perturbs the response of a subset of variables. For this purpose, the implemented perturbation function was changing the lower bound of the response of observation #105 (corresponding to plate 15) to 20 – a value that would make it an evident outlier in the sample.

For illustration purposes, we first plot the global influence, considering this new perturbation function, for both the lognormal and Weibull models (with species plus random effects), with the original MCMC iterations and burn-in. Results in Figure 17 show that the Weibull model does a much better job at detecting outliers.

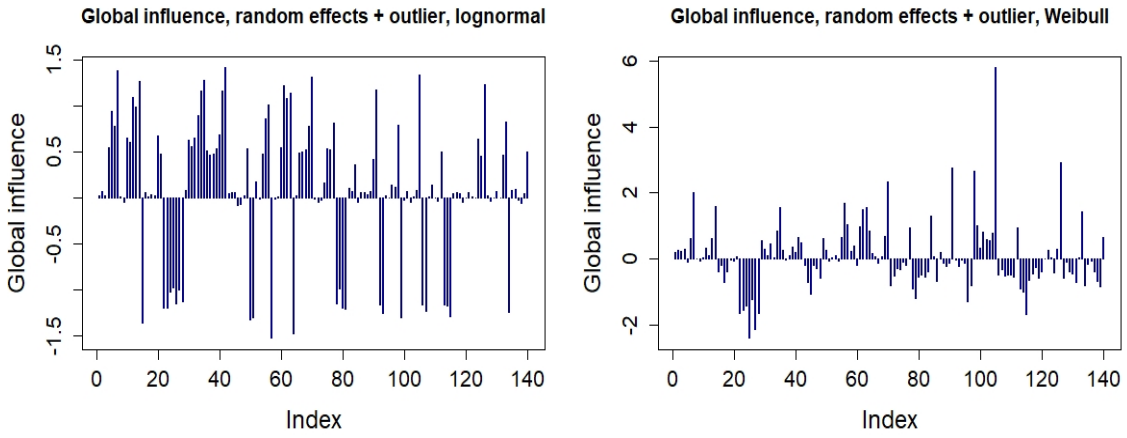


Figure 17: Global influence with outlier, lognormal (left) vs. Weibull (right)

Regarding stability of conclusions for the selected model – Weibull with species plus random effects -, one observes that, although there is variation in the magnitudes of coefficients, conclusions still hold – i.e., species baujardi would have a statistically significant and higher survival time than sacchari -, as illustrated in Table 10. Although not conclusive, these results favor the selected model as a rather robust one.

Table 10: Posterior summary measures, selected model with outlier

Parameter	mean	sd	MC error	95%CrLL	median	95%CrUL
β_0^*	-2.899	0.3626	0.01049	-3.515	-2.883	-2.336
β_1^*	0.7065	0.3532	0.009922	0.1394	0.6992	1.3
ρ	1.586	0.1322	3.32E-03	1.377	1.581	1.811
σ_θ^2	0.2417	0.2109	0.002996	0.04136	0.1852	0.6288

5.3 Posterior Predictive Checks

Due to the censored nature of the data, it is difficult to compute the usual summary measures such as the mean or median survival time. Thus, we decided to use $T_{min}(\mathbf{t})$ as our GOF test statistic since this also gives an idea of how well the model fits the lower tail of the distribution. For our dataset, the minimum survival time is between 0 and 2 days, so we set $T_{min}(\mathbf{t}) = 0$, although the interval is still wide to exactly pinpoint the minimum value. We then take the converged chain for the posterior predictive distribution of future survival times and for each iteration, we compare the generated $T_{min}(\tilde{\mathbf{t}}^k)$ to the value observed from the data. The resulting PPP value $\bar{p}_T$ is 1.0 with a posterior mean $\bar{T}_{min}(\tilde{\mathbf{t}}^k) = 0.109$ day, indicating that according to the selected model, there are grub worms that died after being exposed to the EPN species within 2-3 hours. Figure 18 displays the PPP plot for T_{min} where the red vertical line shows the observed value while the histogram shows the replicated values.

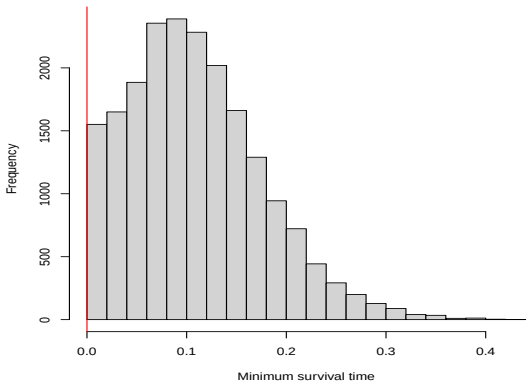


Figure 18: PPP plot for $T_{min}(\mathbf{t})$

From the results, we find that the observed minimum value is not too far from the bulk of the minimum survival times, and due to the uncertainty of the observed minimum value, the observed value appears to the leftmost of the distribution. Nevertheless, the proportion of minimum survival times towards 0 is high implying that the minimum survival times are more likely near 0. Hence, we can say that the selected model is able to fit the data well in the lower tail of the distribution. The distribution of future predicted observations $\tilde{t}_i$ also resembles the distribution of lower limits from the observed intervals, as shown in Figure 19.

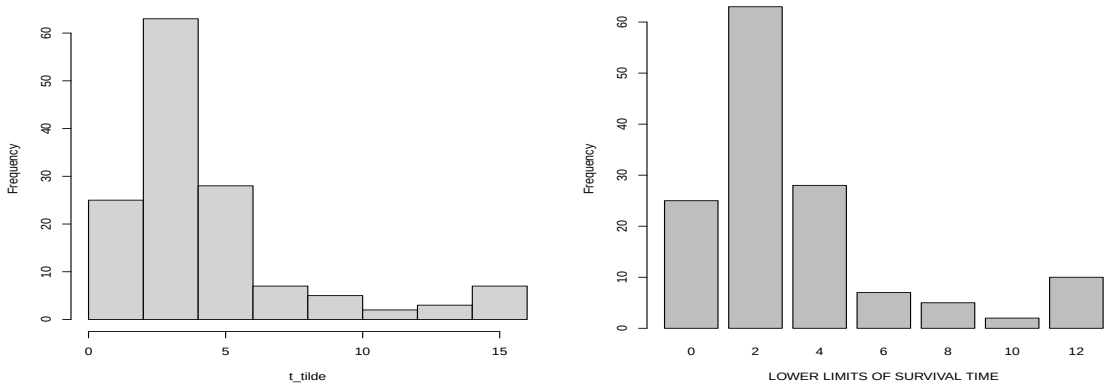


Figure 19: Comparison of PPD and observed data

6 Conclusion

Considering the survival time of the white grubs' exclusively, the Weibull model was shown to be preferred over the Lognormal model when taking into account the interval-censored characteristics of survival time. Section 3 focused on determining which of EPN species acts the fastest in killing white grubs. In this regard, our analysis of models including EPN species as a covariate showed that EPN species has a significant impact on the surival times of white grubs, with EPN species H. baujardi acting the fastest in killing white grubs. Futhermore, grub size as a covariate was added to the models and it was subsequently shown that grub size is not statistically significant.

Including random effects in the Weibull model with EPN species and grub size as covariates resulted in rather contradictory results at first, in which EPN was found to be non-significant. To this end, we proposed a more robust distribution for the random effect in the presence of outliers. This in turn lead to EPN species once again being significant, while grub size remained uninfluential. The analyses of models was also replicated with the consideration of the censoring effect, which is displayed in Section 4.5.

Comparing calculated DIC values among models whereby censoring was recognized, we observed that the models with random effects had the lowest values. In particular, the Weibull model with EPN species as a covariate and a random effect appeared to be best fitting. The models under consideration were also shown to be robust with respect to varying priors and the presence of outliers.

References

David J Spiegelhalter, Keith R Abrams, and Jonathan P Myles. *Bayesian approaches to clinical trials and health-care evaluation*, volume 13. John Wiley & Sons, 2004.

Emmanuel Lesaffre and Andrew B Lawson. *Bayesian biostatistics*. John Wiley & Sons, 2012.

Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. A weakly informative default prior distribution for logistic and other regression models. *The annals of applied statistics*, 2(4):1360–1383, 2008.