UNIVERSITY OF WISCONSIN-LA CROSSE

Graduate Studies

USING BAYESIAN STATISTICS TO ANALYZE ENVIRONMENTAL TOXICOLOGY DATA FROM NATIVE FRESHWATER MUSSELS

A Manuscript Style Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Applied Statistics

Laura Peterman

College of Science and Allied Health
Applied Statistics

May, 2021

USING BAYESIAN STATISTICS TO ANALYZE ENVIRONMENTAL TOXICOLOGY DATA FROM NATIVE FRESHWATER MUSSELS

By Laura Peterman

We recommend acceptance of this thesis in partial fulfillment of the candidate's requirements for the degree of Degree of Master of Science in Applied Statistics.

The candidate has completed the oral defense of the thesis.

Sherwin G. Toribio 05/05/2021

Sherwin Toribio, Ph.D. Date
Thesis Committee Chairperson

On behalf of the committee members named below:

Douglas Baumann, Ph.D.

Teresa Newton, Ph.D.

Thesis accepted

Meredith Thomsen 5/12/2021

Meredith Thomsen, Ph.D. Date
Dean of Graduate & Extended Learning

ABSTRACT

Peterman, L. L. Using Bayesian statistics to analyze environmental toxicology data from native freshwater mussels. MS in Applied Statistics, May 2021, 59 pp. (S. Toribio)

Scientists at the US Geological Survey are looking at the effects of the chemical combination 3-trifluoromethyl-4-nitrophenol (TFM) with 1% niclosamide on native freshwater mussels' behavior and reproduction. Mussels were exposed to the chemicals for 24 hours and reproductive and behavioral outcomes were measured throughout the exposure and 10-day recovery period. The objective of the study is to compare the treatments with respect to these outcome variables. This is usually done using traditional statistical methods like the analysis of variance (ANOVA) or the Kruskal-Wallis test. In this thesis, alternative methods are proposed, using Bayesian methodology. To compare the six treatments, a Bayesian linear regression procedure, a Bayesian logistic mixed effects procedure, and a Bayesian ordinal model were used. This thesis will illustrate how these Bayesian procedures can be implemented using two new R packages. These two R packages use MCMC methods to simulate from the posterior distributions. These simulated values allowed the construction of credible intervals, which can be used in comparing the treatments. Results obtained using the Bayesian methods applied to the native freshwater mussel toxicology data were similar to the results from traditional statistical methods. Some additional benefits of using Bayesian statistics are also discussed in this thesis.

ACKNOWLEDGEMENTS

First, I would like to thank Dr. Sherwin Toribio for the idea that blossomed into my master's thesis. I have really enjoyed learning about Bayesian statistics and how to apply it. I would also like to thank Dr. Toribio for our weekly meetings, keeping me on track, and his advice and guidance while supporting my decision to follow where my research took me. I would like to thank Dr. Teresa Newton for the use of her team's data and her helpful edits and answers to questions. I would also like to thank Dr. Douglas Baumann for taking part in my thesis committee and for all the fun R and statistics concepts I have learned from his classes over the last couple of years. I would like to thank my parents for their advice about the process of doing a master's thesis. I would like to thank my friends for listening to me talk about how cool I thought R was and statistics over the years and interesting things I learned about Bayesian statistics this year. I would like to thank Dr. Barbara Bennie for her advice and guidance throughout my time at UWL. I would like to thank Dr. Anita Davelos for her advice and mentorship throughout my time at UWL. I will forever be grateful for the summer research internship I did with her and Dr. Todd Osmundson. Finally, I would like to thank the UWL biology professors for incorporating statistics into their course curriculums and encouraging my application of statistics to biological data; as it was in their classes I first truly grasped the potential of what could be learned by applying statistics to biological data.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

INTRODUCTION

Aquatic invasive species are a growing global threat, causing changes in ecosystem structure, losses in biodiversity, and substantial economic costs. Sea lamprey (Petromyzon marinus) are an invasive aquatic species that were first found in the Laurentian Great Lakes in 1921 (in Lake Erie) and contributed, along with overfishing, to the decimation of commercial and recreational Great Lakes fish populations in 1950s (Wilkie et al., 2019). Sea lamprey have a multi-stage life cycle (Figure 1) that involves (1) larvae burrowing in sediments in the bottom of a river for 3-7 years, (2) spending about 3-4 months metamorphizing into parasitic juvenile sea lampreys, (3) moving downstream into the Great Lakes to feed on the blood of other fish for 12-20 months (killing many fish in the process), and (4) moving back upstream to spawn new larvae and die (Wilkie et al., 2019).

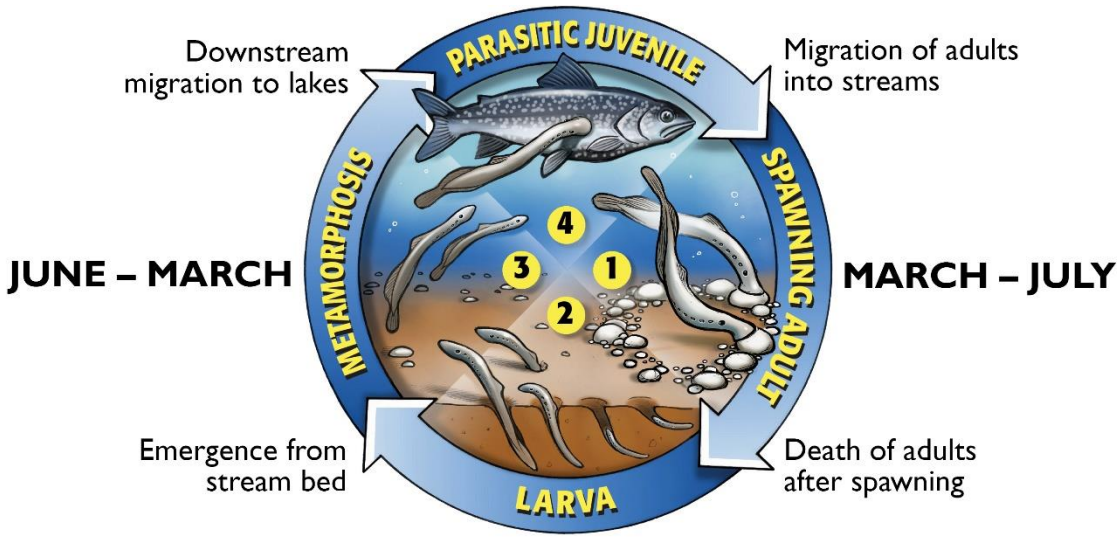

Figure 1. Sea Lamprey Lifecycle. Figure from http://www.glfc.org/sea-lamprey-lifecycle.php.

The international Great Lakes Fishery Commission (GLFC) was created in the 1950's and was tasked with developing a lampricide to kill sea lamprey. After testing thousands of chemicals, 3-trifluoromethyl-4-nitrophenol (TFM) was developed as a lampricide in the 1950's and niclosamide (2',5-dichloro-4'-nitrosalicylanilide) was developed in the 1960's (Figure 2, McDonald & Kolar, 2007). Applying TFM has been a remarkably effective method of controlling sea lamprey populations as sea lamprey populations are down to about 10% of what they were in the 1950's (Wilkie et al., 2019). Niclosamide (in the granular form of Bayluscide®) and TFM are often used together because niclosamide reduces the amount of TFM needed and it is a bottom formulation that targets sea lamprey in large and/or fast flowing waters (Wilkie et al., 2019). TFM and niclosamide work by uncoupling oxidative phosphorylation, decreasing the amount of ATP (adenosine triphosphate, a spark of energy for a cell) produced (Birceanu et al., 2011; Park et al., 2011). Most non-lamprey fish can quickly recover from TFM exposure using enzymes to biotransform TFM into water soluble compounds that can then be

excreted as urine (Wilkie et al., 2019). Despite the successful nature of the GLFC's sea lamprey control program, resource managers are concerned about the potential effects of lampricides on non-target organisms like other fish and native freshwater mussels.

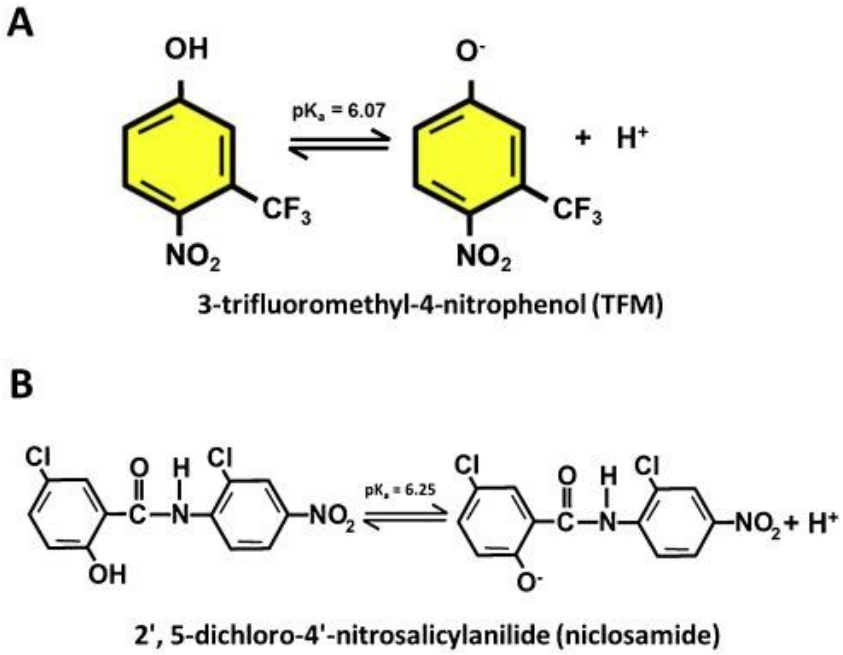

Figure 2. Chemical Structure of (A) TFM and (B) Niclosamide. Figure from Wilkie et al., 2019.

North America has the most freshwater mussel species, with at least 300 species recorded (Haag and Williams, 2013). Mussels perform important ecosystem services in aquatic systems. Mussels are strong filter feeders that filter the water, recycle and store nutrients like nitrogen, provide habitat for other species, and are eaten by people in some Asian countries (Vaughn, 2020). Mussels have a complex lifecycle. Adult mussels release glochidia (larval mussels) that must physically connect with a host (likely a fish) to complete their lifecycle (Figure 3, Barnhart et al., 2008). Glochidia typically encapsulate onto the gills or fins of fish for a period of several weeks to several months before they disconnect from the host and begin a free-living life stage as juveniles

(Barnhart et al., 2008). About 46% of the native mussel species in Wisconsin are endangered, threatened, or a 'species of concern' (*Species List- Wisconsin Mussel Monitoring Program*, n.d.; *Wisconsin's Rare Mussels and Clams -Wisconsin DNR*, 2016). One endangered species is the federally endangered snuffbox (*Epioblasma triquetra*). Boogard et al. (2015) evaluated the effects of TFM exposure on snuffbox mussels, ellipse mussels (*Venustaconcha ellipsiformis*), and the host fish for the snuffbox, log perch (*Percina caprodes*). They found that TFM did not negatively affect survival of either mussel species, suggesting that when applied, TFM may not harm these species at the concentrations tested. Newton et al. (2017) explored the lethal and sublethal effects of Bayluscide© exposure to sub-adult (1-2 years of age) and adult life stages of eight species and generally found that sub-adults were more sensitive to Bayluscide© than adults. Sub-lethal effects included whether the valves were gaped or not, the degree of mucus production, siphoning activity, and foot extension (Newton et al., 2017).

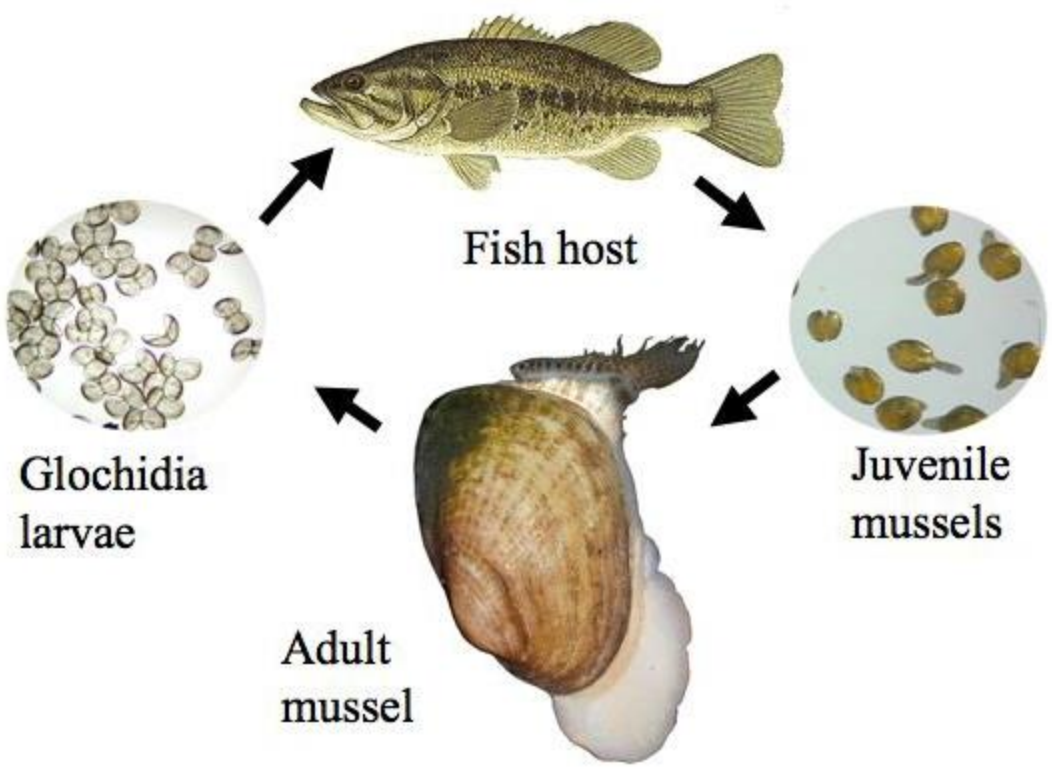

Figure 3. Freshwater Mussel Lifecycle. Figure from https://molluskconservation.org/MUSSELS/Reproduction.html.

METHODS

Study Methods

Currently, Dr. Teresa Newton and her team at U.S. Geological Survey's (USGS) Upper Midwest Environmental Sciences Center (UMESC) are exploring the potential for the lampricides TFM and TFM with 1% niclosamide to affect behavioral and reproductive endpoints of mussels, specifically the premature release of glochidia due to chemical exposure. This phenomenon has been reported in native mussels when exposed to fluoxetine, Prozac© (Bringolf et al., 2010). Prematurely released glochidia may be unable to attach and metamorphose into juvenile mussels on host fish and could affect the population size (Bringolf et al., 2010). The goals of the current USGS work are to

determine the acute toxicity of TFM and TFM: 1% Niclosamide (TFM: N) to free glochidia and glochidia still inside the marsupial gills of adult mussels and to look for potential behavioral and reproductive effects to TFM and TFM: N in pregnant (gravid) adult mussels (Newton et al., 2020). The pocketbook mussels (*Lampsilis cardium)* collected from the Upper Mississippi River (as a surrogate for the federally endangered Higgins eye pearlymussel, *Lampsilis higginsii*) in the fall of 2018 were exposed to TFM and the pocketbook mussels collected in the fall of 2019 were exposed to TFM: N (Newton et al., 2020).

Study Data

Tests with adult mussels conducted in the fall of 2018 and 2019 contain both behavioral and reproductive endpoints. Reproductive endpoints included the number of released glochidia at the end of the 24-hour exposure and every two days during the recovery period along with other parameters. Behavioral endpoints assessed every 2 hours during the first 12 hours of exposure, at 24-hour exposure, and every two days during a 10-day recovery period, include parameters like the degree of mantle lure display (used to attract fish and continue the lifecycle) and the degree of foot extension (Newton et al., 2020). The treatment levels included a negative control with no TFM: N, a positive control with serotonin (20mg/mL, as serotonin induces spawning in bivalves, (Cunha & Machado, 2001)), 2.3 mg/L TFM: N, 3.9 mg/L TFM: N, 5.7 mg/L TFM: N, and 8.1 mg/L TFM: N. In this thesis, the negative control was labeled as the control, while the different concentrations of TFM: N were labeled as TN1, TN2, TN3, and TN4.

Only a subset of the variable from the fall 2019 data were analyzed in this thesis. The statistical programming language, R (version 3.6), was used to analyze the data (R Core Team, 2019).

The scientists at UMESC looked at the effect of the treatment on the percentage of free glochidia that are viable after 12 hours of TFM: N exposure, after 24 hours of TFM: N exposure, and the percentage of glochidia still in the mussel marsupiums that are viable on the 10th day of recovery (Newton et al., 2020). The viability of glochidia was calculated by obtaining the proportion of glochidia that closed their shells after being exposed to NaCl. The scale of the 12-hour and 24-hour data was on a zero to 100 scale while the 10^{th} day of recovery data was on a zero to one scale. For the glochidia toxicity tests at 12 and 24 hours, in place of serotonin there was an additional higher TFM: N treatment TN5.

In the original analysis using traditional (frequentist) statistical methods, a Kruskal-Wallis test was run to look at the effect of the treatment on the percentage of free glochidia that are viable after 12 hours of TFM: N exposure. The Kruskal-Wallis test was used instead of an analysis of variance due to the left skew of the data (Figure 4b) and the potential outlier in treatment four (Figure 4a).

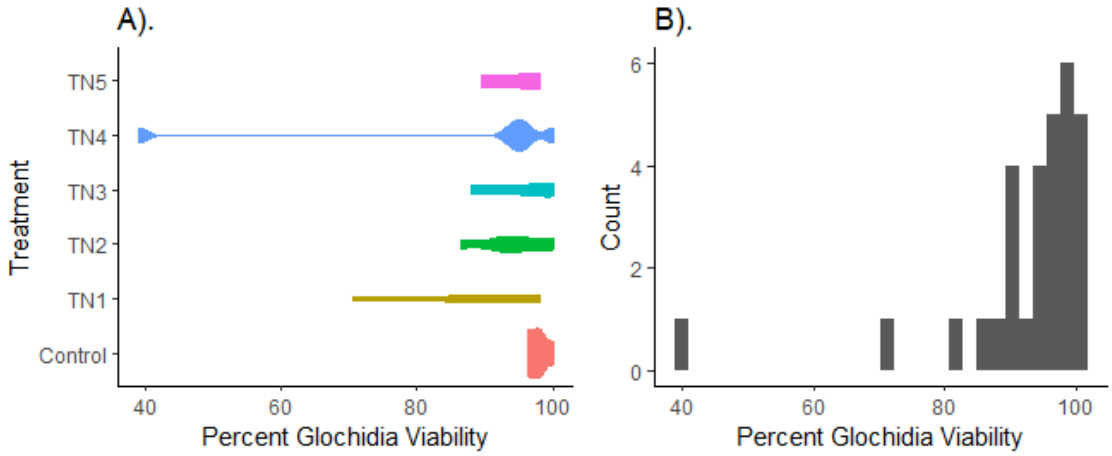

Figure 4. Percentage of Viable Glochidia after 12 Hours of TFM: Niclosamide Exposure. A). Percentage of glochidia that are viable by treatment. B). Histogram of percentage of viable glochidia.

A Kruskal-Wallis test was originally run to look at the effect of the treatment on the percentage of free glochidia that are viable after 24 hours of TFM: N exposure. The Kruskal-Wallis test was used instead of an analysis of variance (ANOVA)[1] due to the left skew of the data (Figure 5b) and the potential outlier in treatment four (Figure 5a).

[1] The term ANOVA was coined by John Tukey along with exploratory data analysis (McGrayne, 2011).

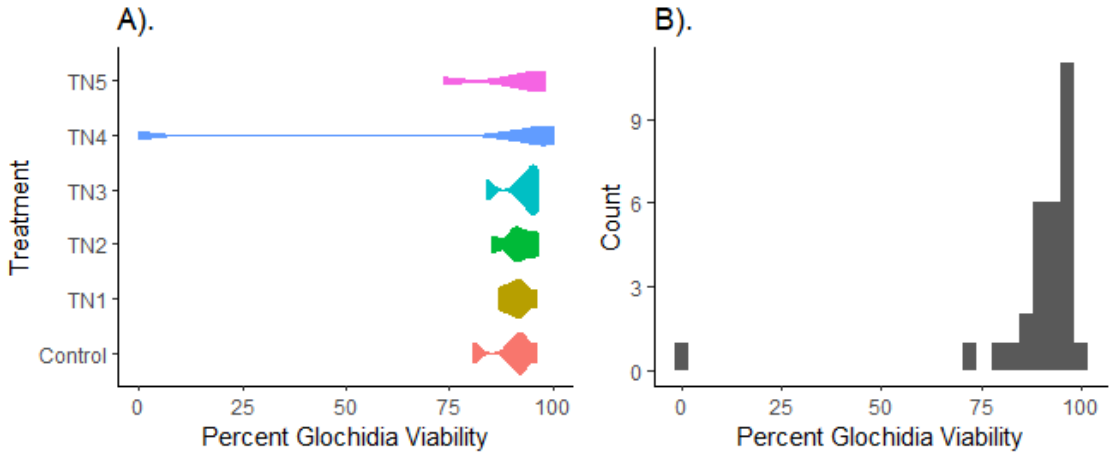

Figure 5. Percentage of Viable Glochidia 24 Hours after the Start of TFM: Niclosamide Exposure. A). Percentage of glochidia that are viable by treatment. B). Histogram of percentage of viable glochidia.

An ANOVA (since there was no noticeable outliers and the data did not have a noticeable skew, Figure *6*) was originally run to look at the effect of the treatment on percentage of glochidia still in the mussel marsupiums that are viable on the 10th day of recovery.

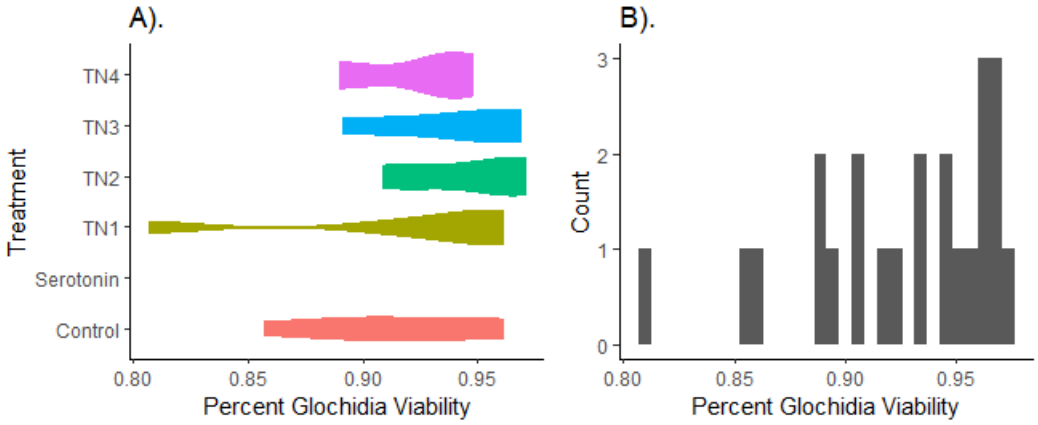

Figure 6. Percentage of Glochidia Still in the Marsupium that are Viable on the 10th Day of Recovery. A). Percentage of glochidia that are viable by treatment. B). Histogram of percentage of glochidia that are viable.

The scientists at UMESC looked at the effect of treatment on the level of mantle lure display where stage 1 (Figure 7a) is when the shell is closed and there is no visible display of mantle lure, stage 2 (Figure 7b) is when the shell is gaped and the mantle tissue is partially extended, and stage 3 (Figure 7c) is when the shell is gaped with the mantle lure fully extended and the marsupial gills extended beyond the shell margin (Leonard et al., 2014). They recorded the level of mantle lure display every two hours from two to 12 hours, at 24 hours, and then at two, four, six, eight, and 10 days (Newton et al., 2020).

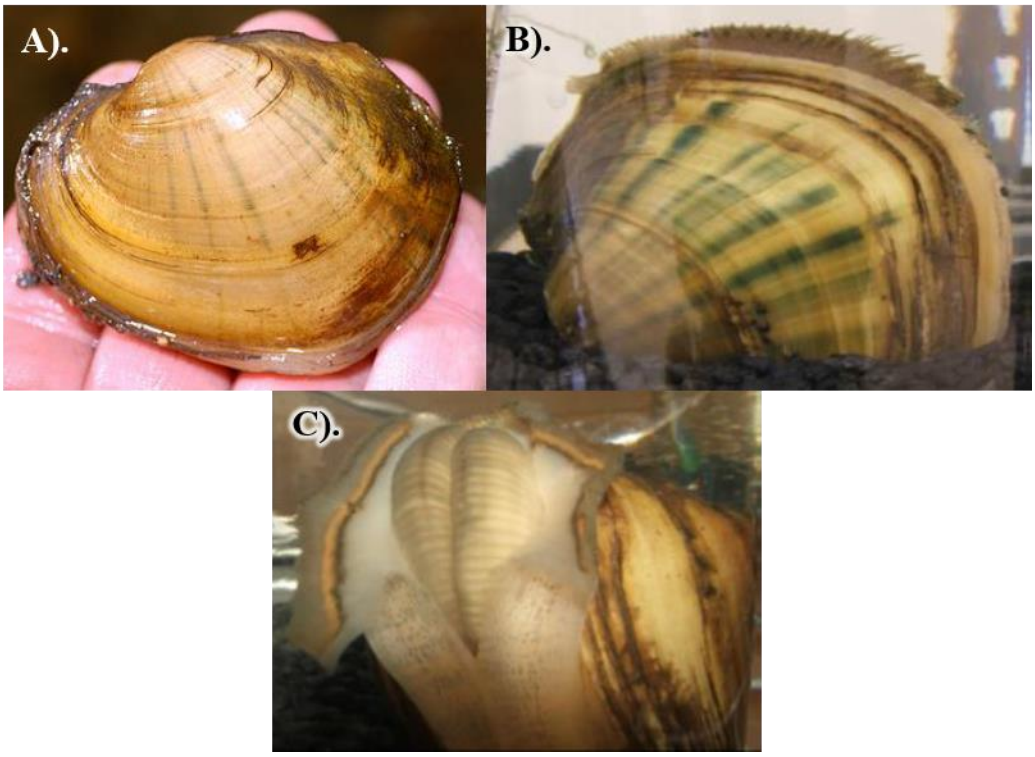

Figure 7. Level of Mantle Lure Display. A). Stage 1, B). Stage 2, C). Stage 3. Pictures from Newton et al. (2020).

The scientists at UMESC looked at the effect of treatment on the level of foot protrusion where stage 1 (Figure 8a) is when the foot is not extended, stage 2 (Figure 8b) is when the foot is partially extended, and stage 3 (Figure 8c) is when the foot is fully

extended (Newton et al., 2020). They recorded the level of foot protrusion at the same time increments as the level of mantle lure display (Newton et al., 2020).

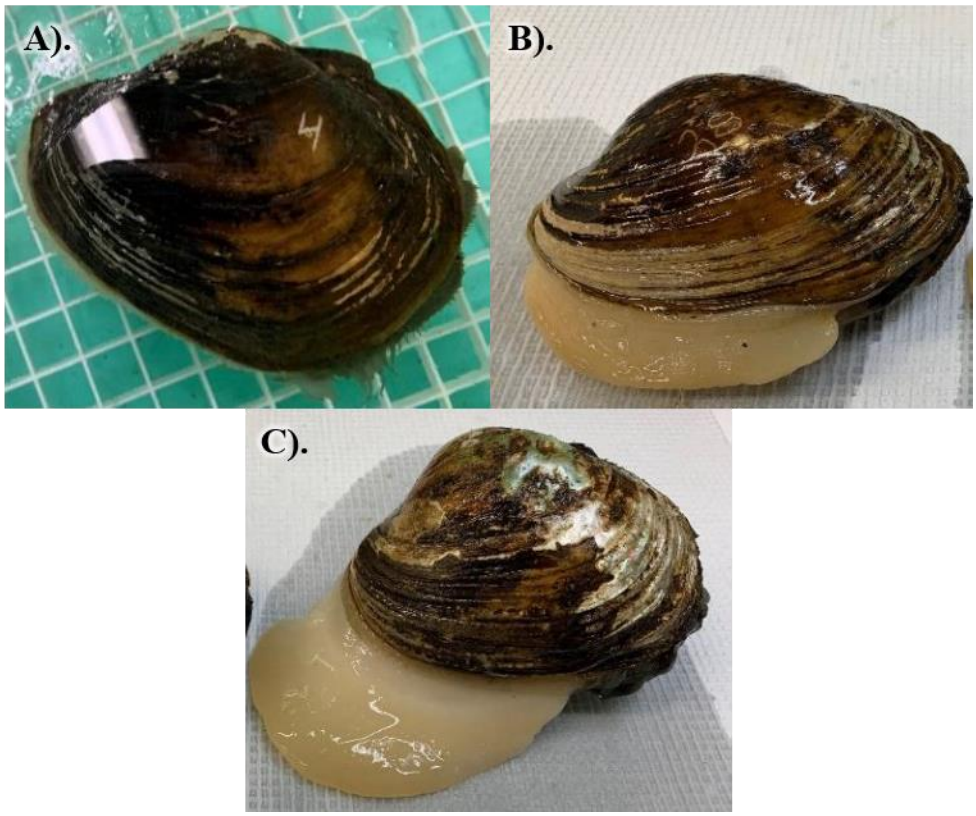

Figure 8. Level of Foot Protrusion. A). Stage 1, B). Stage 2, C). Stage 3. Pictures from Newton et al. (2020).

Reasoning for Chosen Statistical Analysis

As stated in the previous section, the data were originally analyzed by traditional statistics including ANOVA, Kruskal-Wallis, and Dunnett tests. Given the complexity of the data, the small sample sizes (because of the imperiled status of mussels), and the combination of categorical and continuous data, traditional statistical analyses may be inadequate to accurately describe the data. Thus, the main objective of this work is to explore the possibility of analyzing these types of data using Bayesian statistical methods.

In recent years, more people have felt the need to author papers to clarify how null hypothesis significance testing can be used and how p-values should be reported. Wasserstein and Lazar have put out multiple articles in *The American Statistician* about limitations of p-values (Wasserstein & Lazar, 2016; Wasserstein et al., 2019). The American Statistical Association (ASA) put out a statement in 2016 about statistical significance and p-values that also mentioned other approaches to complement or replace the p-value including adding confidence intervals or using alternate methods of testing like Bayesian methods or likelihood ratios (Wasserstein & Lazar, 2016). In 2019, the ASA went a step further and put out a statement that said that one should not say something is statistically significant and gave a thorough list of approaches to complement or replace the p-value, including multiple using Bayesian ideas and multiple using Bayesian statistics (Wasserstein et al., 2019). A recent article in *Environmental Toxicology and Chemistry* gives guidance on null hypothesis significance testing, how p-values should be used, and provides alternative methods like using Bayesian statistics along with worked examples relevant to environmental toxicology (Erickson & Rattner, 2020).

Many ecological datasets contain non-normal data and having small sample sizes does not help. There are two main approaches to deal with non-normal data: transform the data to follow a normal distribution or use non-parametric statistical tests like the Kruskal-Wallis. While both approaches are valid, if you also have small sample sizes like the data described above (six treatments with only five replicates each), neither approach is powerful enough to consistently detect treatment effects (Hackshaw, 2008; Newton et al., 2020). While larger sample sizes would certainly benefit statistical analyses, it can be

expensive, time-consuming, and in the case of rare species do more harm to the species than good to collect many samples. Weather can also delay or prohibit the collection of field samples.

John Tukey, the man who developed the post-hoc multiple comparison procedure Tukey's HSD (honestly sufficient difference) (Tukey, 1949), said that "discarding Bayesian techniques would be a real mistake" (McGrayne, 2011). It is important to consider all potential statistical methods (Bayesian and frequentist) when picking the right method for the job. It is also important to consider differences between the two branches of statistics (

Table *1*). Bayesian statistics finds the probability of hypothesis 1 being true given the observed data and frequentist statistics finds the probability of the data being observed given the null hypothesis is true (Ellison, 2004). Asking for the probability of it raining given that is noon ($P(H_1 \mid Data)$) is a different question than asking for the probability of it being noon given that it is raining ($P(Data \mid H_1)$), Kruschke & Liddell, 2018b). Frequentist confidence intervals specify uncertainty about the interval itself while the Bayesian version, credible intervals, specify uncertainty about parameter values' location (Lambert, 2018). Another noticeable difference is that with Bayesian statistics the null hypothesis can be accepted for practical purposes or rejected while with frequentist statistics the null hypothesis will either be rejected or fail to be rejected (Ellison, 2004).

Table 1. Comparison of Bayesian and Frequentist Statistical Methods.

Comparison	Bayesian Statistics	Frequentist Statistics
Question being asked	P (H_1 | Data)	P (Data | H_1)
Definition of Probability	An individual's belief level of an event's likelihood	The relative frequency of events run over a lengthy period of time
Use Prior Knowledge	Yes	No
Model Parameter Treatment	As random variables	As estimates of fixed values
Uncertainty about Interval	Parameter values' location	Interval itself
Null Hypothesis	Reject / Accept	Reject / Fail to reject

Besides being able to accept null hypotheses for practical purposes, there are several other benefits to using Bayesian statistics. Bayesian statistics can handle all kinds and sizes of data ranging from small data sets with ordinal dependent variables to huge data sets with lots of variables to data sets mostly consisting of zeros. Unlike confidence intervals, credible intervals also have distribution information (Kruschke & Liddell, 2018b). Bayesian statistics gives exact results for any sample size of at least one, but they are more precise with larger sample sizes (Kruschke, 2014). It is also completely valid to run Bayesian statistics on the data in the middle of a clinical trial (Kruschke & Liddell, 2018b).

Bayesian Statistical Methods

Background

The Bayesian equivalent of an ANOVA was applied to the mussel data described above. The Bayes way of thinking is more common than you think. It is all about

updating your beliefs based on new information. Martin Feldstein, a Harvard economics professor, explained Bayesian statistics using an analogy involving an umbrella (McGrayne, 2011). He said Bayesian statistics is like a person deciding whether or not to carry an umbrella even when there is a small probability of rain. Carrying around an umbrella when it does not rain is a nuisance, but getting drenched when it pours and you do not have an umbrella is not fun (Feldstein et al., 2004; Greenspan, 2004).

The mathematical basis of Bayesian statistics is the Bayes theorem: $\mathrm{P}(H_1|Data) = \frac{P(Data|H_1)P(H_1)}{P(Data)}$ (Kurt, 2019). It can also be rewritten as $P(H_1|Data) = \frac{P(Data|H_1)P(H_1)}{P(Data)} \propto P(Data|H_1)P(H_1)$ where the probability of the hypothesis given the data (the posterior probability distribution) is proportional to the probability of the data given the hypothesis (the likelihood) times the probability of the hypothesis (the prior probability distribution) (Kurt, 2019). Using the modern naming conventions, the Bayes theorem should be called the Bayes-Price-Laplace theorem (Lambert, 2018). Thomas Bayes, an English Presbyterian minister, came up with the theorem (a predecessor to the version mentioned earlier) but never published it (McGrayne, 2011). After Thomas Bayes died in 1761, a family friend and fellow Presbyterian minister, Richard Price developed Bayes's work into a paper (Bayes & Price, 1763) published a couple of years later (McGrayne, 2011). Pierre Simon Laplace[2], an astronomer-mathematician, developed the version of Bayes theorem mentioned at the start of the paragraph (McGrayne, 2011).

[2] Laplace, known as the Newton of France, developed the Central Limit Theorem and named the meter, millimeter, and centimeter (McGrayne, 2011).

Priors

With Bayesian statistics you can incorporate your prior knowledge and beliefs into statistical models. There are many kinds of priors and multiple ways to define them (Kruschke, 2014). With large data sets, the priors do not impact the posterior distribution much, but with smaller data sets they are more impactful (Lemoine, 2019). The two main categories of priors are non-informative and informative priors. Non-informative priors include flat and diffuse priors where nothing is known about the population parameters (van de Schoot et al., 2021). Common non-informative priors include using the uniform distribution where each option has an equal probability of happening and Normal(mean (μ) = 0, variance (σ^2) = 1000) (Lemoine, 2019). Informative priors are built from prior knowledge about the parameters being estimated (van de Schoot et al., 2021). Priors can be developed in several ways including among others: they can come from previous studies, they could be representative of the data and made in consultation with an expert in the topic of the data, or sub-sample could be taken of the data and set aside to develop priors with (Kruschke, 2014).

If no prior information about the parameters being estimated has been published, there is alternative to non-informative priors. There are multiple reasons to not use non-informative priors: uniform priors can sometimes be informative, non-informative priors give results identical to those from frequentist statistics, and non-informative priors have elevated type I error rates like frequentist methods (Lemoine, 2019). Kruschke (2014) recommends using weakly informative priors instead of non-informative priors. Weakly informative priors are useful when we know a little bit about the parameter, like for example, it cannot be negative (Muth et al., 2018). Weakly informative priors built on the

expected parameter range provide regularization to avoid overfitting and stabilize computation while not having a strong impact on the posterior distribution (Muth et al., 2018). Due to not having much prior knowledge about the parameters, the Bayesian statistical analysis was run mostly using weakly informative priors.

Bayes Factors

Bayes factors can be used as part of a statistical approach, but literature suggested that this was not the best approach. Bayes factors[3] ($\frac{P(Data|H_1)}{P(Data|H_2)}$), are a ratio of two hypotheses' likelihoods and can be used to compare models (Kurt, 2019). Bayes factors are highly sensitive to the priors used (Kass & Raftery, 1995). Instead of being able to get the exact value for the Bayes factor, in practice it is usually estimated using the Savage-Dickey density ratio (see Wagenmakers et al., (2010) for more information). Gelman et al. (2013)[4] does not recommend using Bayes factors for model comparison. Bayes factors can work well with discrete models, but not so well with continuous models (Gelman et al., 2013). Bayes factors for continuous models are sensitive to the variance of priors (Gelman et al., 2013). Instead of Bayes factors, Gelman et al. (2013) and Kruschke (2014)[5] recommend using Bayesian hierarchical regression instead when the goal is the Bayesian equivalent of an ANOVA.

[3] Harold Jeffreys developed the Bayes factor and helped revive Bayesian statistics with his book *Theory of Probability* (McGrayne, 2011).
[4] Gelman et al. (2013) is *Bayesian Data Analysis*, a popular Bayesian statistics textbook.
[5] Kruschke (2014) is *Doing Bayesian Data Analysis*, a little less popular Bayesian statistics textbook.

Statistical Models

Unlike a traditional ANOVA, Bayesian hierarchical regression models do not split the variance into parts, which is why they are called the Bayesian equivalent of an ANOVA instead of Bayesian ANOVA (Kruschke, 2014).

The Bayesian statistical analysis developed for this thesis consisted of three Bayesian linear regression models, one Bayesian logistic mixed effects model, and one Bayesian ordinal model. The Bayesian ordinal model was run using the brm function in the brms R package (Bürkner, 2017; Bürkner & Vuorre, 2019). The Bayesian linear regression models were run using the stan_glm function and Bayesian logistic mixed effects model was run using the stan_glmer function, both from the rstanarm R package (Goodrich et al., 2020; Muth et al., 2018). The rstanarm R package is more user-friendly but the brms packages has been developed for more types of models. See Appendix A for example R codes.

Bayesian linear regression models were run on the glochidia that are viable at 12 hours, 24 hours, and those still in the mussel marsupiums on the 10^{th} day of recovery. The glochidia viability models had weakly informative priors built using recommendations from Lemoine (2019).

Free glochidia viability at 12 hours priors:

Intercept ~ Normal (92.167, 11.884)

Coefficients ~ Normal (0, 2.5)

Sigma ~ Half-Cauchy (0, 2.5)

Free glochidia viability at 24 hours priors:

Intercept ~ Normal (88.847, 17.680)

Coefficients ~ Normal (0, 2.5)

Sigma ~ Half-Cauchy (0, 2.5)

Glochidia still in the mussel marsupiums that are viable on the 10th day of recovery priors:

Intercept ~ Normal (0.925, 0.043)

Coefficients ~ Normal (0, 2.5)

Sigma ~ Half-Cauchy (0, 2.5)

While there was potential for all three stages of mantle lure display to be observed, not all stages were observed. No instance of stage three mantle lure display was recorded, so the variable was treated as binary for the purpose of the Bayesian statistical analysis (Figure 9).

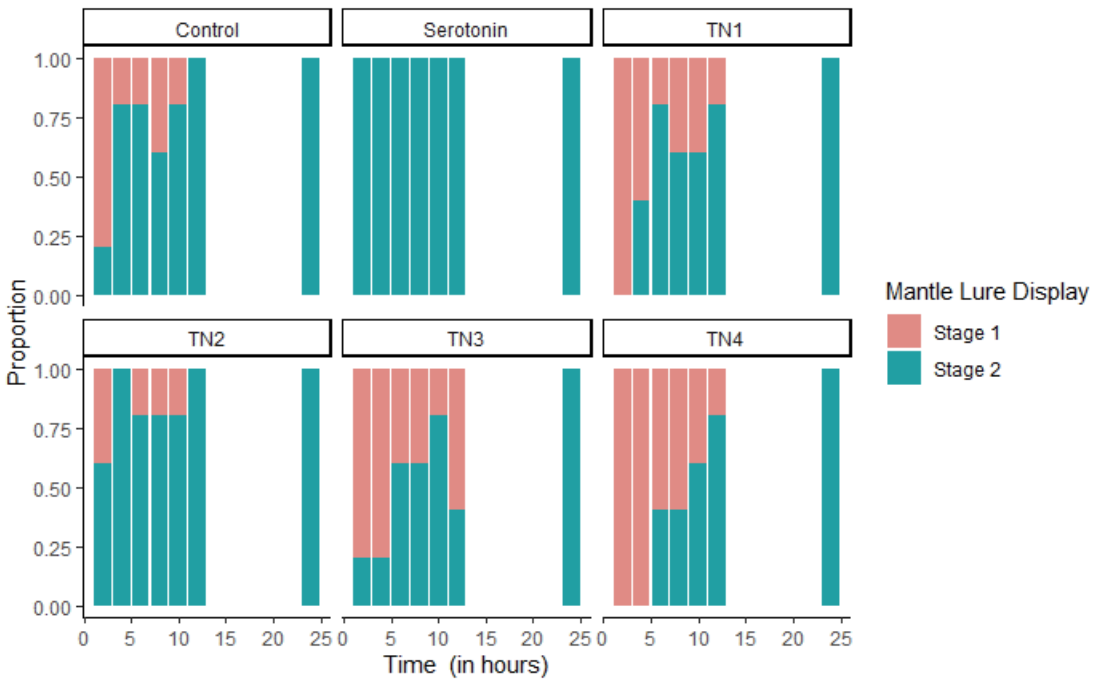

Figure 9. Proportion of Mantle Lure at Different Stages Recorded for the 24 Hours of TFM: N Exposure.

Several models were considered to look at the effect of the four treatments and the two controls on the level of mantle lure display. The final model used treatment as a fixed factor and time (in hours) and mussel ID as random factors (random intercepts but not slopes). A logistic mixed effects model was used to account for the binary nature of the mantle lure variable and the repeated measures for each mussel.

The default priors for the model type were used since no prior information about the level of mussel lure display was known. The covariance matrix accounts for the between mussel variation in the intercept and slope (Muth et al., 2018). The decov prior was developed as a robust model regularizing prior that works well most of the time (Muth et al., 2018).

Level of mantle lure display priors:

Intercept ~ Normal (0,2.5) (after the predictors are centered)

Coefficients ~ Normal (0, 6.69)

Covariance ~ decov (regularization=1, concentration=1, shape=1, scale=1)

All three levels of foot protrusion were present, so foot protrusion was considered an ordinal variable with ordered categories for the Bayesian statistical analysis (Figure 10).

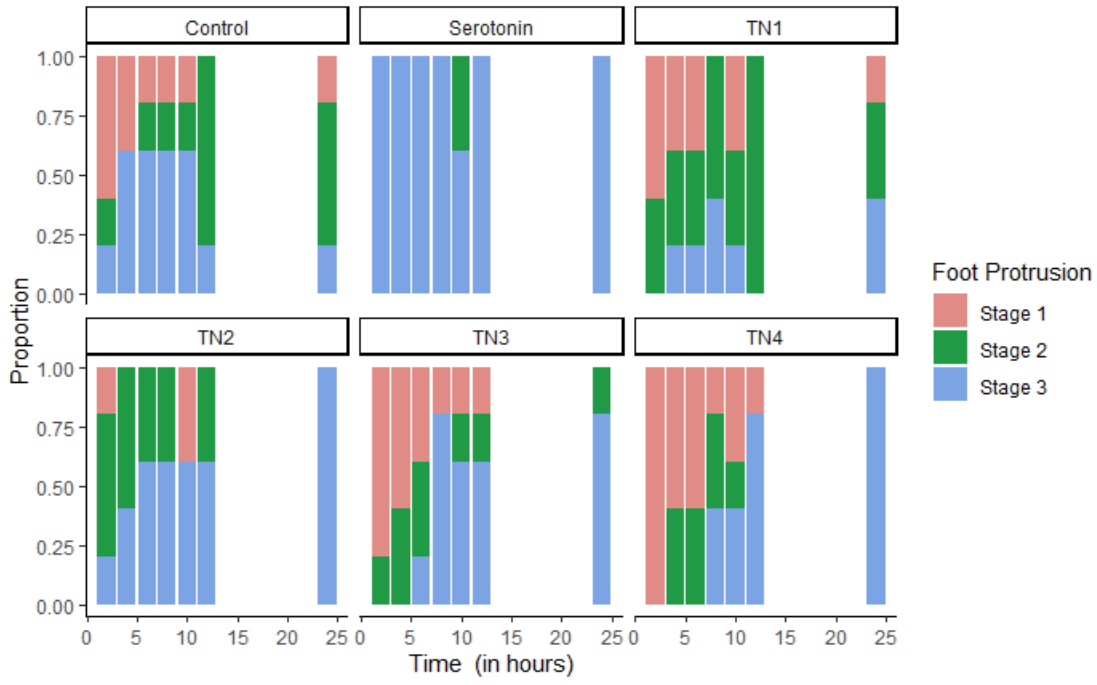

Figure 10. Proportion of Foot Protrusion at Different Stages Recorded for the 24 Hours of TFM: N Exposure.

A cumulative mixed effects model was run to account for the ordinal nature of the foot protrusion data and the repeated measures for each mussel (Bürkner & Vuorre 2019). Like the mantle lure model, treatment was treated as a fixed effect and time (in hours) and the mussel id were treated as random effects. The default brms priors were used for this model.

Level of foot protrusion priors:

Intercept ~ t (μ=0, σ=2.5, degrees of freedom=3)

Standard deviation ~ t (μ=0, σ=2.5, degrees of freedom=3)

Markov Chain Monte Carlo

For the most basic of examples, the posterior distribution can be calculated using paper and a pencil, but with the Bayesian statistical models mentioned above, computing power and techniques like Markov Chain Monte Carlo (MCMC) are needed to study the

posterior distributions. MCMC combines the process of Markov chains and Monte Carlo[6] integration and uses simulation to estimate the posterior distribution of interest (van de Schoot et al., 2021). The process of Markov chains allows parameter values to be generated from the posterior distribution and Monte Carlo integration estimates the posterior distribution by using computer simulation to estimate integrals (van de Schoot et al., 2021). A Markov chain is determined by its transition kernel (how the Markov chain parameter values are updated) and beginning parameter values (van de Schoot et al., 2021). Transition kernels are determined by proposal distributions like the Gibbs sampler, Metropolis-Hastings algorithm, and the No-U-Turn Sampler (NUTS) (van de Schoot et al., 2021). The Metropolis-Hastings algorithm allows for the sampling of constrained parameter spaces from a probability distribution (Lambert, 2018). NUTS is an extension of Hamiltonian Monte Carlo, a Metropolis-Hastings algorithm useful for small sample sizes (van de Schoot et al., 2021). NUTS is built into Stan[7], a probabilistic programming language designed for Bayesian inference (Lambert, 2018). Stan is faster and more general than BUGS and JAGS, two other Bayesian inference software (Lambert, 2018). The R functions used to run the statistical models in this thesis were developed to run using Stan without the user having to know Stan code (Bürkner, 2017; Goodrich et al., 2020).

---

[6] Monte Carlo integration is named after the Monte Carlo casino in Monaco where Stanislaw Ulam's (one of the developers) uncle would gamble away money (McGrayne, 2011).

[7] Stan was developed by a group of people including Andrew Gelman, a prolific Bayesian statistician who also helped develop $\hat{R}$ (Lambert, 2018).

Chains and Convergence

If a Markov chain is run for long enough, the parameter values will converge to a stable state where the values stay within the stable distribution (van de Schoot et al., 2021). Convergence to the desired distribution is important because the goal is to get accurate estimates of parameter values. If several Markov chains (4-8 chains for simple models) are run instead of one, the whole range of possible parameter values is more likely to have been explored and convergence to the same stable distribution signifies the convergence to the desired distribution (Lambert, 2018).

There are several ways to assess model fit and whether convergence has been reached including trace plots, $\hat{R}$, effective sample size (ESS), Monte Carlo standard error (MCSE), and posterior predictive checking (PPC) plots (Muth et al., 2018). Trace plots display the generated parameter values for all iterations and show how the range of possible parameter values is being explored (van de Schoot et al., 2021). The Gelman-Rubin statistic, $\hat{R}$, is a ratio of within-chain and between-chain variability, where the model is suggested to have converged when the $\hat{R}$ values for all the parameters are less than 1.1 (Lambert, 2018; van de Schoot et al., 2021). Effective sample size is the effective length of the MCMC chain, yields an indication of algorithm efficiency, and is recommended to be greater than 1000 (Muth et al., 2018; van de Schoot et al., 2021). Monte Carlo standard error is approximately the posterior standard deviation divided by square root of ESS (Muth et al., 2018). PPC plots are used to see if the model fits the data well and are built by simulating data from the posterior predictive distribution and comparing it to observed data (Muth et al., 2018). All metrics were used to measure

convergence and model fit for the statistical models[8] mentioned above but the trace plots and $\hat{R}$ values (all about 1.00) were not included for brevity.

Credible Intervals

While both credible intervals (CrI) and confidence intervals are uncertainty intervals, they are calculated and interpreted differently. Confidence intervals are generally interpreted as with (1-α)% (usually 95%) confidence, the true value of some parameter is between number a and number b (Wackerly, 2007). Credible intervals describe uncertainty in parameter value location and are calculated from the posterior distribution (Lambert, 2018). Credible intervals are interpreted as there is a X% (generally 89% or 95%) probability that the value of some parameter is between number c and number d (Lambert, 2018). Highest density intervals (HDI), a type of credible intervals, contain the most credible values for a parameter (Kruschke & Liddell, 2018a). The HDI is usually a 95% HDI due to how common 95% confidence intervals are (Kruschke & Liddell, 2018a).

Comparison of Treatment Levels Using the ROPE-HDI Approach

The ROPE is the region of practical equivalence surrounding the null value (ROPE'd value) where any value in the region is essentially the null value (Kruschke & Liddell, 2018a). The default size of the ROPE is $\pm 0.1SD(Dependent\_Variable)$ for linear models because Cohen (1988) defined a small effect size as 0.2 and a half of that was used for the ROPE (Kruschke & Liddell, 2018a). The default ROPE size for logistic

[8] By default, the first 50% of the values simulated were discarded as part of the burn-in period.

models is $\pm 0.1 * \frac{\pi}{3}$, using $\frac{\pi}{3}$ from Cohen (1988) to convert log odd ratios to standardized differences (Kruschke, 2014). The ROPE-HDI approach has three outcomes. If the ROPE completely excludes the 95% HDI, then the ROPE'd value is rejected (Kruschke 2014, Figure 11a). If the ROPE completely includes the 95% HDI, then the ROPE'd value is accepted "for practical purposes" (otherwise it is left undecided) (Kruschke 2014, Figure 11b). The ROPE'd value is rejected instead of a hypothesis, but essentially it means that there is an effect (Kruschke & Liddell, 2018a). So, if two treatment levels are compared using the ROPE-HDI approach and the ROPE'd value is rejected; it is in essence saying that there is a difference between the two treatment levels (see Appendix A for example R code).

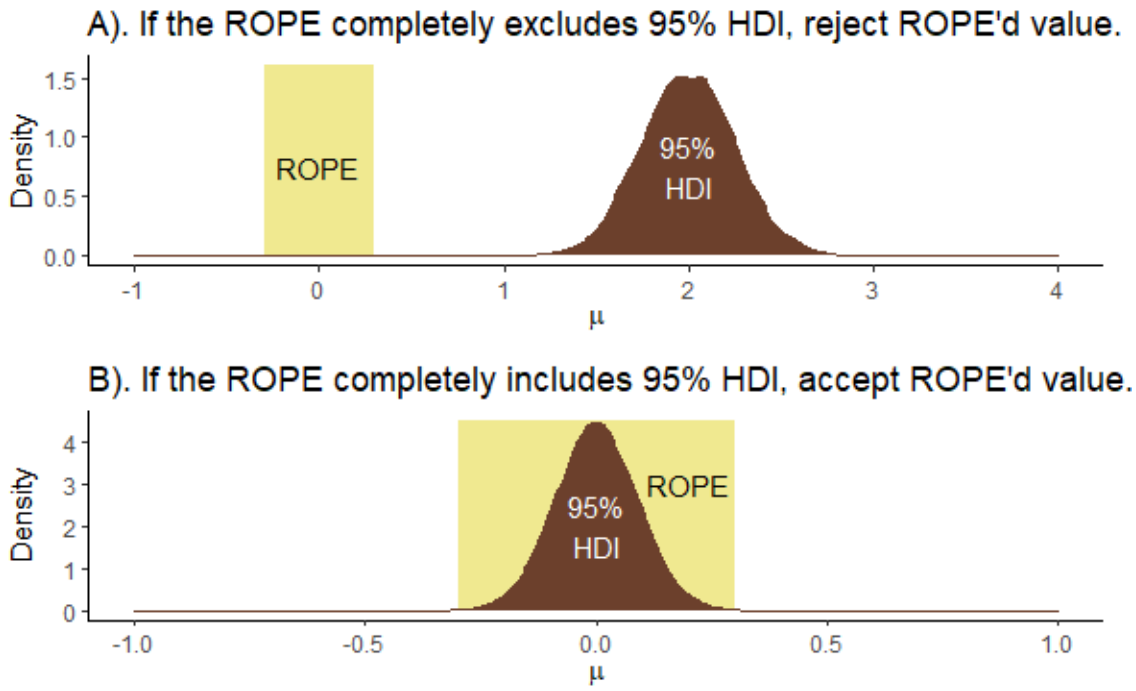

Figure 11. Pictorial Representation of the ROPE-HDI Approach.

RESULTS

Glochidia Toxicity Tests

I. Glochidia Viability after 12 Hours of Exposure

According to the results of the Kruskal-Wallis test, there was not sufficient evidence that at least one treatment has a different mean percent glochidia viability from the other treatments after 12 hours of TFM: N exposure ($\chi^2 = 5.4215$, degrees of freedom(df)=5, p=0.366).

The Bayesian linear regression model, run with four chains and 10,000 iterations, passed all the metrics: it converged well with all the $\widehat{R}$ values equal to one and the effective sample sizes were all greater than 1000 (Table 2). The data from the posterior predictive distribution and the observed data do have similar distributions (Figure 12).

Table 2. Glochidia Viability after 12 Hours of Exposure Model Summary

Treatment	Mean	MCSE	Standard Deviation	95% CrI Low	95% CrI High	ESS
(Intercept)	92.40	0.01	2.35	87.83	97.16	25973
TN1	-1.05	0.01	2.29	-5.53	3.46	27884
TN2	0.25	0.01	2.30	-4.30	4.75	27076
TN3	0.57	0.01	2.30	-3.98	5.11	28477
TN4	-1.42	0.01	2.26	-5.88	3.07	27793
TN5	0.33	0.01	2.30	-4.14	4.81	28422
Sigma	11.92	0.01	1.63	9.22	15.63	23195

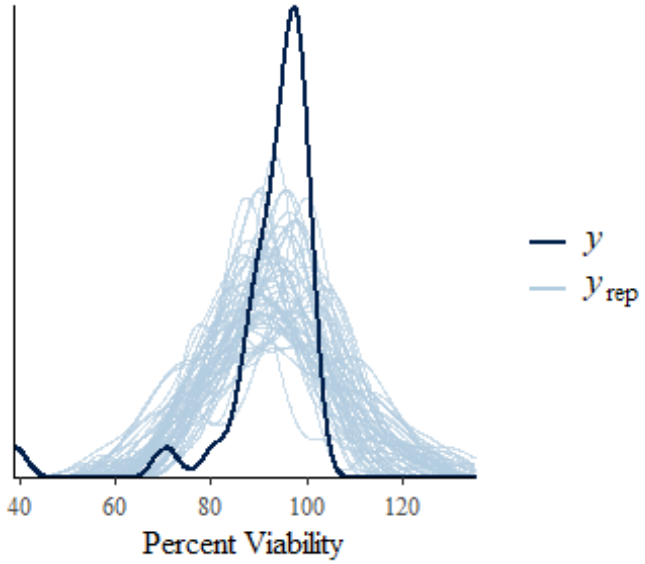

Figure 12. Comparison of Observed Data and Data from the Posterior Predictive Distribution

Using the ROPE-HDI approach with a ROPE range of -1.188 to 1.188, the contrasts are all undecided as a small percent of the 95% credible interval is within the ROPE (Table 3).

Table 3. Contrasts for the Percentage of Viable Glochidia after 12 Hours of TFM: Niclosamide Exposure.

Level 1	Level 2	Difference	95% CrI Low	95% CrI High	ROPE Percentage	ROPE Equivalence
Control	TN1	1.05	-3.46	5.52	0.03	Undecided
Control	TN2	-0.26	-4.70	4.34	0.03	Undecided
Control	TN3	-0.58	-5.15	3.91	0.03	Undecided
Control	TN4	1.44	-2.93	6.00	0.03	Undecided
Control	TN5	-0.32	-4.74	4.20	0.04	Undecided
TN1	TN2	-1.31	-7.69	4.89	0.02	Undecided
TN1	TN3	-1.63	-7.86	4.63	0.02	Undecided
TN1	TN4	0.39	-5.69	6.68	0.03	Undecided
TN1	TN5	-1.38	-7.55	4.93	0.02	Undecided
TN2	TN3	-0.32	-6.54	5.96	0.02	Undecided
TN2	TN4	1.68	-4.53	7.79	0.02	Undecided
TN2	TN5	-0.06	-6.30	6.20	0.02	Undecided
TN3	TN4	2.00	-4.11	8.50	0.02	Undecided
TN3	TN5	0.23	-6.10	6.34	0.02	Undecided
TN4	TN5	-1.77	-7.60	4.68	0.02	Undecided

II. Glochidia Viability after 24 Hours of Exposure

There was not sufficient evidence that at least one treatment has a different mean percent glochidia viability from the other treatments 24 hours after the exposure started (Kruskal-Wallis test, $\chi^2 = 1.7226$, df =5, p=0.886).

The Bayesian linear regression model, run with four chains and 10,000 iterations, passed all the metrics: it converged well with all the $\widehat{R}$ values equal to one and the effective sample sizes were all greater than 1000 (Table 4). The data from the posterior predictive distribution and the observed data do have somewhat similar distributions (Figure 13).

Table 4. Percentage of Viable Glochidia 24 Hours after the Start of TFM: Niclosamide Exposure Model Summary

Treatment	Mean	MCSE	Standard Deviation	95% CrI Low	95% CrI High	ESS
(Intercept)	88.88	0.02	3.31	82.36	95.46	26112
TN1	0.21	0.01	2.41	-4.54	4.99	29816
TN2	0.25	0.01	2.38	-4.39	4.88	29998
TN3	0.33	0.01	2.39	-4.33	5.05	28242
TN4	-1.14	0.01	2.43	-5.86	3.63	28898
TN5	0.15	0.01	2.42	-4.63	4.97	28324
Sigma	17.80	0.02	2.41	13.87	23.25	22840

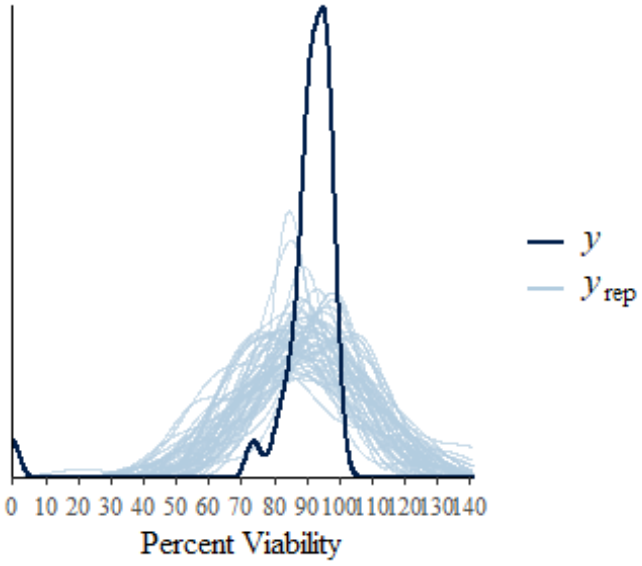

Figure 13. Comparison of Observed Data and Data from the Posterior Predictive Distribution.

Using the ROPE-HDI approach with a ROPE range of -1.768 to 1.768, the contrasts are all undecided as a small percent of the posterior distribution is within the ROPE (Table 5).

Table 5. Contrasts for the Percentage of Viable Glochidia 24 Hours after the Start of TFM: Niclosamide Exposure.

Level 1	Level 2	Difference	95% CrI Low	95% CrI High	ROPE Percentage	ROPE Equivalence
Control	TN1	-0.20	-5.04	4.48	0.04	Undecided
Control	TN2	-0.27	-4.87	4.40	0.03	Undecided
Control	TN3	-0.32	-5.01	4.35	0.04	Undecided
Control	TN4	1.14	-3.63	5.85	0.03	Undecided
Control	TN5	-0.13	-5.08	4.48	0.03	Undecided
TN1	TN2	-0.04	-6.77	6.58	0.02	Undecided
TN1	TN3	-0.10	-6.62	6.49	0.03	Undecided
TN1	TN4	1.35	-5.45	7.86	0.03	Undecided
TN1	TN5	0.03	-6.34	6.82	0.02	Undecided
TN2	TN3	-0.08	-6.79	6.38	0.02	Undecided
TN2	TN4	1.40	-5.25	7.88	0.02	Undecided
TN2	TN5	0.09	-6.56	6.65	0.03	Undecided
TN3	TN4	1.47	-5.04	8.17	0.02	Undecided
TN3	TN5	0.19	-6.30	6.77	0.03	Undecided
TN4	TN5	-1.29	-7.87	5.40	0.02	Undecided

Adult Toxicity Tests

I. Glochidia Viability in Marsupium on the 10th Day of Recovery

There was not sufficient evidence that at least one treatment has a different average percent of glochidia still in the mussel marsupiums that are viable from the other treatments on the 10th day of recovery (ANOVA, Table 6, F=0.849, df=5, p=0.534).

Table 6. ANOVA for Percentage of Glochidia Still in the Mussel Marsupiums that are Viable on the 10th Day of Recovery.

	Degrees of Freedom	Sum of Squares	Mean Square Error	F Statistic	P-Value
Treatment	5	0.008	0.002	0.849	0.534
Residuals	17	0.033	0.002		

The Bayesian linear regression model, run with four chains and 10,000 iterations, passed all the metrics: it converged well with all the $\widehat{R}$ values equal to one and the effective sample sizes were all greater than 1000 (Table 7). The data from the posterior predictive distribution and the observed data do have closely similar distributions (Figure 14).

Table 7. Percentage of Glochidia Still in the Mussel Marsupiums that are Viable on the 10th Day of Recovery Model Summary

Treatment	Mean	MCSE	Standard Deviation	95% CrI Low	95% CrI High	ESS
(Intercept)	0.91	0	0.02	0.87	0.96	7969
Serotonin	-0.05	0	0.05	-0.16	0.05	13738
TN1	0.00	0	0.03	-0.05	0.06	9792
TN2	0.03	0	0.03	-0.03	0.09	9846
TN3	0.03	0	0.03	-0.04	0.09	10352
TN4	0.01	0	0.04	-0.06	0.08	10839
Sigma	0.05	0	0.01	0.03	0.07	10779

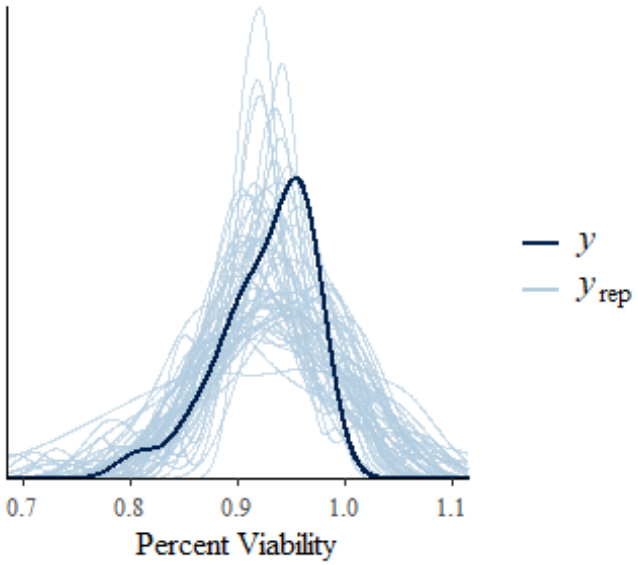

Figure 14. Comparison of Observed Data and Data from the Posterior Predictive Distribution

Using the ROPE-HDI approach with a ROPE range of -0.0043 to 0.0043, there are some undecided contrasts, but there are also several contrasts where the null value of 0 was accepted "for practical purposes" meaning there is no difference (Table 8). The control, TN1, TN2, TN3, and TN4 were all found to not be different from each other (Table 8).

Table 8. Contrasts for Percentage of Glochidia Still in the Mussel Marsupiums that are Viable on the 10th Day of Recovery.

Level 1	Level 2	Difference	95% CrI Low	95% CrI High	ROPE Percentage	ROPE Equivalence
Control	Serotonin	0.05	-0.05	0.16	0.83	Undecided
Control	TN1	0.00	-0.07	0.05	1.00	Accepted
Control	TN2	-0.03	-0.09	0.03	1.00	Accepted
Control	TN3	-0.03	-0.09	0.04	1.00	Accepted
Control	TN4	-0.01	-0.08	0.06	1.00	Accepted
Serotonin	TN1	-0.06	-0.16	0.04	0.81	Undecided
Serotonin	TN2	-0.09	-0.19	0.02	0.61	Undecided
Serotonin	TN3	-0.08	-0.18	0.03	0.66	Undecided
Serotonin	TN4	-0.06	-0.17	0.05	0.76	Undecided
TN1	TN2	-0.03	-0.09	0.03	1.00	Accepted
TN1	TN3	-0.02	-0.08	0.04	1.00	Accepted
TN1	TN4	0.00	-0.07	0.06	1.00	Accepted
TN2	TN3	0.01	-0.06	0.07	1.00	Accepted
TN2	TN4	0.02	-0.05	0.09	1.00	Accepted
TN3	TN4	0.02	-0.06	0.09	1.00	Accepted

II. Level of Mantle Lure Display (1, 2, or 3)

The Bayesian logistic mixed effects model, run with six chains and 12,000 iterations, passed all the metrics: it converged well with all the $\widehat{R}$ values equal to one and the effective sample sizes were all greater than 1000 (Table 9). The estimates could have been more precise, but data from the posterior predictive distribution and the observed data do have similar distributions (Figure 15).

Table 9. Level of Mantle Lure Display Model Summary.

Treatment	Mean	MCSE	Standard Deviation	95% CrI Low	95% CrI High	ESS
(Intercept)	1.73	0.01	1.09	-0.38	3.95	11944
Serotonin	7.12	0.02	3.21	2.25	14.70	21467
TN1	-1.21	0.01	1.06	-3.45	0.81	14301
TN2	1.09	0.01	1.15	-1.15	3.40	17387
TN3	-1.63	0.01	1.09	-3.92	0.40	14349
TN4	-2.16	0.01	1.08	-4.46	-0.16	14340

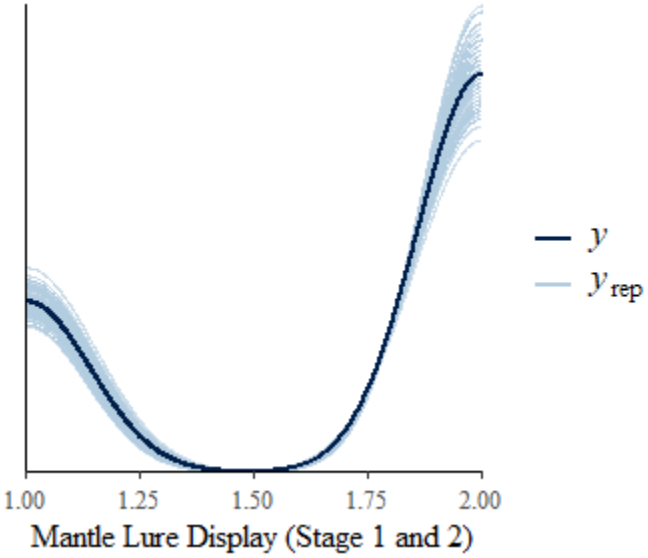

Figure 15. Comparison of Observed Data and Data from the Posterior Predictive Distribution

The ROPE-HDI approach with a ROPE range of -0.181 to 0.181 was applied to the output from the final model to contrast the different treatment levels (Table 10). Using the ROPE-HDI approach, differences were found between Serotonin and all other treatment levels, TN2 and TN3, and TN2 and TN4 (Table 10).

Table 10. Contrasts for Level of Mantle Lure Display.

Level 1	Level 2	Log Odds Difference	95% CrI Low	95% CrI High	ROPE Percentage	ROPE Equivalence
Control	Serotonin	-6.61	-13.92	-1.85	0.00	Rejected
Control	TN1	1.17	-0.86	3.39	0.04	Undecided
Control	TN2	-1.07	-3.36	1.18	0.05	Undecided
Control	TN3	1.59	-0.45	3.85	0.02	Undecided
Control	TN4	2.12	0.10	4.37	0.00	Undecided
Serotonin	TN1	7.81	2.75	15.00	0.00	Rejected
Serotonin	TN2	5.50	0.55	12.91	0.00	Rejected
Serotonin	TN3	8.24	3.26	15.45	0.00	Rejected
Serotonin	TN4	8.75	3.88	16.13	0.00	Rejected
TN1	TN2	-2.25	-4.65	-0.03	0.00	Undecided
TN1	TN3	0.41	-1.63	2.48	0.08	Undecided
TN1	TN4	0.94	-1.11	2.97	0.05	Undecided
TN2	TN3	2.65	0.42	5.14	0.00	Rejected
TN2	TN4	3.18	1.03	5.67	0.00	Rejected
TN3	TN4	0.53	-1.59	2.55	0.08	Undecided

III. Level of Foot Protrusion (1, 2, or 3)

The Bayesian cumulative mixed effects model was run with six chains and 12,000 iterations. The model converged well with all the $\widehat{R}$ values equal to one and the effective sample sizes were all greater than 1000 (

Table *11*). The estimates could have been more precise but data from the posterior predictive distribution and the observed data do follow the same general distribution (Figure 16 and Figure 17).

Table 11. Level of Foot Protrusion Model Summary.

Treatment	Estimate	Standard Deviation	95% CrI Low	95% CrI High	Bulk ESS	Tail ESS
Intercept (1->2)	-1.35	0.90	-3.12	0.45	10938	14416
Intercept (2->3)	0.75	0.89	-0.98	2.56	11087	14960
Serotonin	4.62	1.34	2.17	7.45	16668	20864
TN1	-0.78	1.00	-2.80	1.23	12409	17170
TN2	1.25	1.01	-0.75	3.29	12860	17712
TN3	-0.22	1.01	-2.22	1.79	12003	16583
TN4	-0.68	1.02	-2.72	1.33	12157	17989
Discrimination	1.00	0.00	1.00	1.00	36000	36000
Mussel ID SD	1.33	0.33	0.77	2.09	12458	20078
Time (in hours) SD	1.43	0.54	0.70	2.76	11164	18213

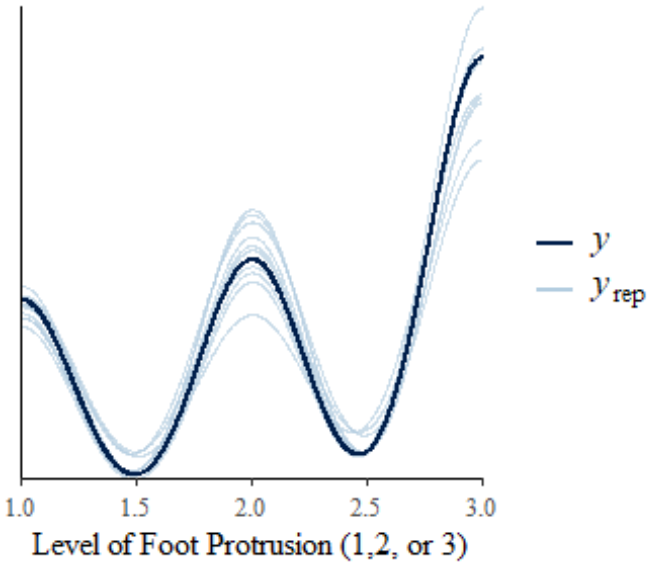

Figure 16. Comparison of Observed Data and Data from the Posterior Predictive Distribution

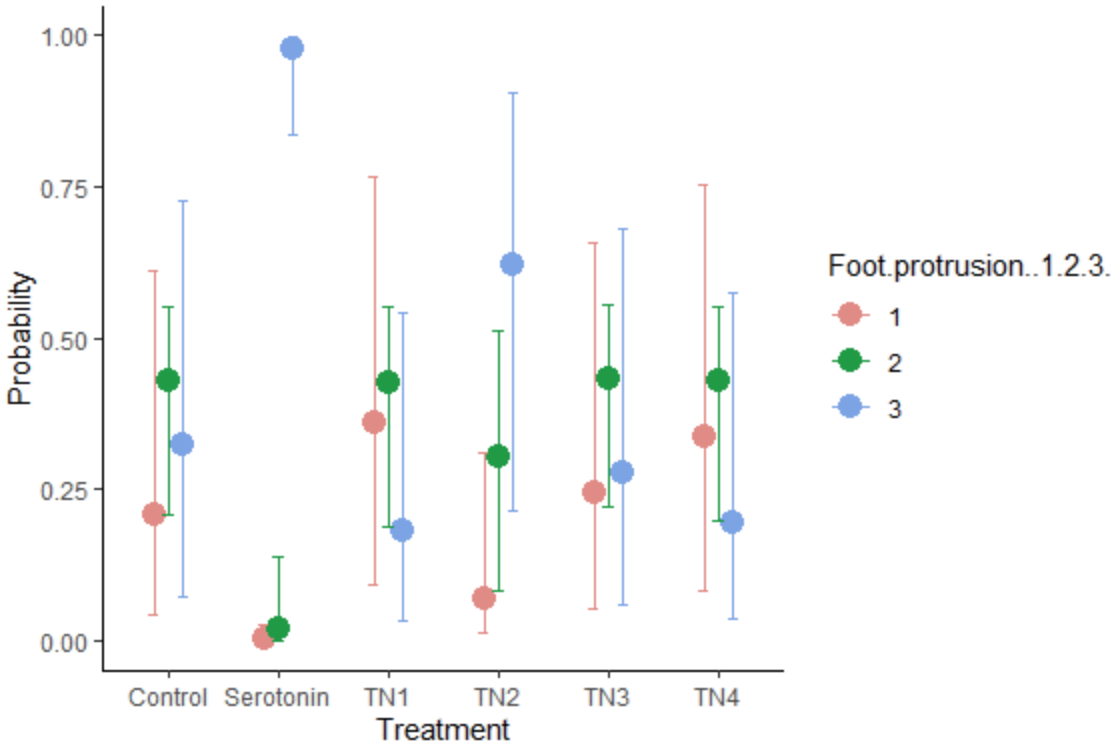

Figure 17. Conditional Effects Plot for the Foot Protrusion Model.

The ROPE-HDI approach with a ROPE range of -0.1 to 0.1 was used on the foot protrusion model to compare treatment levels (

Table *12*). Using the ROPE-HDI approach, differences were found between Serotonin and all other treatment levels (Table 12).

Table 12. Contrasts for Level of Foot Protrusion.

Level 1	Level 2	Difference	95% CrI Low	95% CrI High	ROPE Percentage	ROPE Equivalence
Control	Serotonin	-4.55	-7.32	-2.07	0.00	Rejected
Control	TN1	0.77	-1.21	2.81	0.06	Undecided
Control	TN2	-1.23	-3.23	0.80	0.04	Undecided
Control	TN3	0.21	-1.72	2.27	0.08	Undecided
Control	TN4	0.67	-1.27	2.77	0.07	Undecided
Serotonin	TN1	5.31	2.81	8.11	0.00	Rejected
Serotonin	TN2	3.30	0.83	6.05	0.00	Rejected
Serotonin	TN3	4.77	2.21	7.49	0.00	Rejected
Serotonin	TN4	5.22	2.72	8.04	0.00	Rejected
TN1	TN2	-2.00	-4.02	0.00	0.01	Undecided
TN1	TN3	-0.56	-2.52	1.43	0.07	Undecided
TN1	TN4	-0.10	-2.11	1.87	0.09	Undecided
TN2	TN3	1.44	-0.59	3.45	0.03	Undecided
TN2	TN4	1.91	-0.11	3.95	0.01	Undecided
TN3	TN4	0.47	-1.53	2.50	0.08	Undecided

DISCUSSION

Biological Interpretations of Data

Although no decisions could be made about the effect of treatment on the percentage of free glochidia that are viable after 12 or 24 hours of TFM: N exposure, the same cannot be said for the percentage of glochidia still in the mussel marsupiums that are viable on the 10^{th} day of recovery. The original frequentist analysis did not find sufficient evidence that at least one treatment was different from the others (Table 6), while the Bayesian analysis was actually able to say that several treatment levels are no different from each other (Table 8). Differences were found between treatment levels, mainly between Serotonin and all other treatment levels, in the level of mantle lure display and level of foot protrusion models (Table 10, Table 12).

Statistical Implications of Results

As illustrated in the previous sections, Bayesian methods can be used to analyze environmental toxicology data and the results are similar to results obtained using frequentist statistical methods. It was done using modern, user-friendly R packages which use the same formula syntax that the frequentist linear regression R function lm() and the frequentist ANOVA R function aov() use. Although the underlying mechanism used by the Bayesian approach is different, the way to get the results using the R functions stan_glm, stan_glmer, and brm are similar to the frequentist version. No new statistical programming languages needed to be learned to implement Bayesian statistical methods. This would make it easy for statisticians or data analysts to implement Bayesian statistical methods. The final code for the statistical analysis did not take that long to run either, clocking in at about 15 minutes for whole analysis including graphs and tables.

The Bayesian equivalent of an ANOVA was found to be most useful with data that does not follow the normal distribution. If the goal is comparing treatment levels with an ordinal dependent variable, the recommended frequentist options are, a Kruskal-Wallis test or a Friedman test if there are repeated measures (Tabachnick & Fidell, 2007). Bayesian ordinal models along with the ROPE- HDI approach incorporate the ordinal data structure along with any randomized groups or repeated measures into the statistical model while still being able to compare treatment levels.

Limitations

There were multiple limitations to the Bayesian statistical analysis. The default method of calculating ROPE ranges were used but the results would probably be more consequential if the regions of practical equivalence (ROPE) were defined in consultation

with a domain knowledge expert before collecting the data or at least before looking at the data. This was not possible for the Bayesian statistical analysis, as the data analysis stage had already begun before learning about purposefully defined ROPE ranges. The results would have also been more relevant if the priors had been deliberately designed using prior knowledge before the data was collected. This was not feasible as no prior knowledge was known about the parameters of interest. The priors used in the Bayesian statistical analysis were also limited to the options of prior distributions developed for the specific R functions used; had Stan been known, a wider variety of prior distribution options would have been available to use in the statistical models (like beta priors for proportions).

Future Directions

Multiple future directions for this research include a power simulation to compare the power of Bayesian statistical methods to frequentist statistical methods and more kinds of models. It can be done but it gets complicated as Bayesian methods do not base decisions on controlling Type I and II error rates (Kruschke & Liddell, 2018b). Kruschke & Liddell (2018b) even suggests using frequentist methods if the goal is error control. Kruschke & Liddell (2018b) later describes the concept of Bayesian power analysis and even mentions a R package developed to run a Bayesian power analysis with two treatment levels, but nothing has been developed for more than two treatment levels.

Additional models could be developed using Bayesian method for all the unused variables. One of the unused variables is the number of conglutinates (clumps of eggs) released (Newton et al., 2020). In the original frequentist statistical analysis, no statistical tests were run on the variable since it contains mostly zeros. However, there are statistical

models called zero-inflated models that were developed for count data with lots of zeros (Congdon, 2005). The Bayesian version of a zero-inflated model (possible with the brms package, (Bürkner, 2017)) could be used with the ROPE-HDI approach to actually learn more about whether treatment level has an effect on the number of conglutinates released, something not possible with frequentist statistical methods.

Finally, the brms R package can also be used to run the Bayesian equivalent of a multivariate ANOVA (MANOVA) to compare between the TFM data and the TFM: N data. The TFM data and the TFM: N data were not collected at the same time, but the experiments were setup and conducted the same way except for the change in chemical composition.

Conclusion

This thesis started with the aim of finding a Bayesian way to analyze environmental toxicology data. Three Bayesian models (Bayesian linear regression, Bayesian logistic with mixed effects, and Bayesian ordinal model) were found to be appropriate for the five variables selected from the USGS native freshwater mussel data. These Bayesian models can be implemented using modern, user-friendly R packages that use the same formula syntax as the frequentist linear regression and ANOVA R functions. The Bayesian equivalent of an ANOVA can be applied to a variety of variable types and sample sizes. Bayesian statistics allows for null hypotheses to be accepted or rejected unlike frequentist statistics which only allows for null hypotheses to fail to be rejected or rejected. Given the complexity of the USGS freshwater mussel data described above, the small sample sizes (because of the endangered status of mussels), and the combination of continuous and categorical data, frequentist statistical analyses could be

inadequate to accurately describe the data. Bayesian statistical methods offer an alternative and should be considered as an option when deciding what statistical method(s) to use to analyze environmental toxicology data.

REFERENCES

Barnhart, M. C., Haag, W. R., & Roston, W. N. (2008). Adaptations to host infection and larval parasitism in Unionoida. *Journal of the North American Benthological Society*, *27*(2), 370–394. https://doi.org/10.1899/07-093.1

Bayes, T., & Price, R. (1763). LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London*, *53*, 370–418. https://doi.org/10.1098/rstl.1763.0053

Birceanu, O., McClelland, G. B., Wang, Y. S., Brown, J. C. L., & Wilkie, M. P. (2011). The lampricide 3-trifluoromethyl-4-nitrophenol (TFM) uncouples mitochondrial oxidative phosphorylation in both sea lamprey (*Petromyzon marinus)* and TFM-tolerant rainbow trout (*Oncorhynchus mykiss*). *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, *153*(3), 342–349. https://doi.org/10.1016/j.cbpc.2010.12.005

Boogaard, M. A., Newton, T. J., Hubert, T. D., Kaye, C. A., & Barnhart, M. C. (2015). Evaluation of the short term 12 hour toxicity of 3-trifluoromethyl-4-nitrophenol (TFM) to multiple life stages of *Venustaconcha ellipsiformis* and *Epioblasma triquetra* and its host fish (*Percina caprodes*). *Environmental Toxicology and Chemistry*, *34*(7), 1634–1641. https://doi.org/10.1002/etc.2959

Bringolf, R. B., Heltsley, R. M., Newton, T. J., Eads, C. B., Fraley, S. J., Shea, D., & Cope, W. G. (2010). Environmental occurrence and reproductive effects of the pharmaceutical fluoxetine in native freshwater mussels. *Environmental Toxicology and Chemistry*, *29*(6), 1311–1318. https://doi.org/10.1002/etc.157

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101. https://doi.org/10.1177/2515245918823199

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second edition.). Hillsdale, N.J. : L. Erlbaum Associates.

Congdon, P. (2005). *Bayesian Models for Categorical Data*. Wiley.

Cunha, E. M., & Machado, J. (2001). Parturition in *Anodonta cygnea* induced by selective serotonin reuptake inhibitors (SSRIs). *Canadian Journal of Zoology*, *79*(1), 95–100. https://doi.org/10.1139/z00-187

Ellison, A. M. (2004). Bayesian inference in ecology. *Ecology Letters*, *7*(6), 509–520. https://doi.org/10.1111/j.1461-0248.2004.00603.x

Erickson, R. A., & Rattner, B. A. (2020). Moving Beyond $p < 0.05$ in Ecotoxicology: A Guide for Practitioners. *Environmental Toxicology and Chemistry*, *39*(9), 1657–1669. https://doi.org/10.1002/etc.4800

Feldstein, M., King, M., & Yellen, J. L. (2004). Panel Discussion. *American Economic Review*, *94*(2), 41–48. https://doi.org/10.1257/0002828041302307

FMCS - Freshwater Mussels. (n.d.). Retrieved March 28, 2021, from https://molluskconservation.org/MUSSELS/Reproduction.html

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis*. CRC Press.

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2020). *rstanarm: Bayesian applied regression modeling via Stan.* (R package version 2.21.1) [Computer software]. https://mc-stan.org/rstanarm

Great Lakes Fishery Commission—Sea Lamprey Lifecycle. (n.d.). Retrieved March 28, 2021, from http://www.glfc.org/sea-lamprey-lifecycle.php

Greenspan, A. (2004). Risk and Uncertainty in Monetary Policy. *American Economic Review*, *94*(2), 33–40. https://doi.org/10.1257/0002828041301551

Haag, W. R., & Williams, J. D. (2014). Biodiversity on the brink: An assessment of conservation strategies for North American freshwater mussels. *Hydrobiologia*, *735*(1), 45–60. https://doi.org/10.1007/s10750-013-1524-7

Hackshaw, A. (2008). Small studies: Strengths and limitations. *European Respiratory Journal*, *32*(5), 1141–1143. https://doi.org/10.1183/09031936.00136408

Hajduk, G. K. (2019, September 10). *Introduction to linear mixed models*. https://ourcodingclub.github.io/tutorials/mixed-models/

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kruschke, J. (2014). *Doing Bayesian Data Analysis*. Saint Louis: Elsevier Science & Technology.

Kruschke, J. K., & Liddell, T. M. (2018a). Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review*, *25*(1), 155–177. https://doi.org/10.3758/s13423-017-1272-1

Kruschke, J. K., & Liddell, T. M. (2018b). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4

Kurt, W. (2019). *Bayesian Statistics the Fun Way: Understanding Statistics and Probability with Star Wars, LEGO, and Rubber Ducks*. No Starch Press.

Lambert, B. (2018). *A Student's Guide to Bayesian Statistics*. SAGE.

Lemoine, N. P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7), 912–928. https://doi.org/10.1111/oik.05985

Leonard, J. A., Cope, W. G., Barnhart, M. C., & Bringolf, R. B. (2014). Metabolomic, behavioral, and reproductive effects of the synthetic estrogen 17 α-ethinylestradiol on the unionid mussel *Lampsilis fasciola*. *Aquatic Toxicology*, *150*, 103–116. https://doi.org/10.1016/j.aquatox.2014.03.004

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of Effect Existence and Significance in the Bayesian Framework. *Frontiers in Psychology*, *10*, 2767. https://doi.org/10.3389/fpsyg.2019.02767

Makowski, D., Lüdecke, D., & Ben-Shachar, M. S. (2020). *Estimation of Model-Based Predictions, Contrasts and Means.* https://github.com/easystats/modelbased

McDonald, D. G., & Kolar, C. S. (2007). Research to Guide the Use of Lampricides for Controlling Sea Lamprey. *Journal of Great Lakes Research*, *33*(sp2), 20–34. https://doi.org/10.3394/0380-1330(2007)33[20:RTGTUO]2.0.CO;2

McGrayne, S. B. (2011). *The Theory That Would Not Die*. New Haven: Yale University Press.

Muth, C., Oravecz, Z., & Gabry, J. (2018). User-friendly Bayesian regression modeling: A tutorial with rstanarm and shinystan. *The Quantitative Methods for Psychology*, *14*(2), 99–119. https://doi.org/10.20982/tqmp.14.2.p099

Newton, T. J., Boogaard, M. A., Gray, B. R., Hubert, T. D., & Schloesser, N. A. (2017). Lethal and sub-lethal responses of native freshwater mussels exposed to granular

Bayluscide®, a sea lamprey larvicide. *Journal of Great Lakes Research*, *43*(2), 370–378. https://doi.org/10.1016/j.jglr.2016.12.010

Newton, T., Schloesser, N., Boogaard, M., & Toribio, S. (2020). *Potential Reproductive Effects of Lampricides on Native Mussels*.

Park, S.-J., Shin, J.-H., Kang, H., Hwang, J.-J., & Cho, D.-H. (2011). Niclosamide induces mitochondria fragmentation and promotes both apoptotic and autophagic cell death. *BMB Reports*, *44*(8), 517–522. https://doi.org/10.5483/BMBRep.2011.44.8.517

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Species List- Wisconsin Mussel Monitoring Program. (n.d.). Wisconsin Mussel Species. Retrieved October 12, 2020, from http://wiatri.net/inventory/mussels/About/speciesList.cfm?theOrder=C

Tabachnick, B. G., & Fidell, L. S. (2007). *Experimental Designs Using ANOVA*. Thomson/Brooks/Cole.

Tukey, J. W. (1949). Comparing Individual Means in the Analysis of Variance. *Biometrics*, *5*(2), 99–114. JSTOR. https://doi.org/10.2307/3001913

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., & Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, *1*(1), 1. https://doi.org/10.1038/s43586-020-00001-2

Vaughn, C. (2020, August 25). *The Functional Ecology of Freshwater Mussels*. https://fws.rev.vbrick.com/#/videos/74130f8c-6cfe-4c6e-9607-7a3cafbcf815

Wackerly, D. D. (2007). *Mathematical statistics with applications* (Seventh edition.). Thomson Brooks/Cole.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, *70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a World Beyond "p< 0.05." *The American Statistician*, *73*(sup1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

Wilkie, M. P., Hubert, T. D., Boogaard, M. A., & Birceanu, O. (2019). Control of invasive sea lampreys using the piscicides TFM and niclosamide: Toxicology, successes & future prospects. *Aquatic Toxicology*, *211*, 235–252. https://doi.org/10.1016/j.aquatox.2018.12.012

Wisconsin's rare mussels and clams -Wisconsin DNR. (2016, May). https://dnr.wi.gov/topic/endangeredresources/Animals.asp?mode=list&Grp=19#legend

APPENDIX A

EXAMPLE CODE

To Setup Stan:

Rtools is needed to run the Bayesian analysis as it contains a C++ compiler needed for Stan. Rtools versions correspond to R versions, so Rtools 35 works with R 3.5.x and R 3.6.x and Rtools 40 works with R 4.0.x (install instructions can be found at https://cran.r-project.org/bin/windows/Rtools/). To setup Rtools run the following commands in R:

```
install.packages("pkgbuild")

rt_path = pkgbuild::rtools_path()
rt_bin = paste0(substr(rt_path,1,nchar(rt_path)-4),"/mingw_$(WIN)/bin/")
writeLines(paste0('PATH="',rt_path,';${PATH}"'), con = "~/.Renviron")
writeLines(paste0('Sys.setenv(BINPREF = "',rt_bin,'")'), con = "~/.Rprofile")
install.packages("jsonlite",type="source")
```

After setting up Rtools, RStan can be installed and setup. The following RStan setup commands came from https://github.com/stan-dev/rstan/wiki/RStan-Getting-Started.

```
remove.packages("rstan")
if (file.exists(".RData")) file.remove(".RData")

Sys.setenv(MAKEFLAGS = "-j4") # four cores used

install.packages("rstan", type = "source")

library(rstan)
example(stan_model, package = "rstan", run.dontrun = TRUE)

options(mc.cores = parallel::detectCores())
rstan_options(auto_write = TRUE)
```

To Run a Bayesian Regression Model using the rstanarm Package:

After Rtools and Rstan have been set up, we can now move on to building a Bayesian regression model. The rstanarm R package is a user-friendly way to run Bayesian regression models and connects to Stan without the user having to know how to code in Stan (Goodrich et al., 2020; Muth et al., 2018).The following is example code for a Bayesian linear regression model where the formula is the same as a frequentist linear regression with Y as the dependent variable and X as the independent variable.

```
library(rstanarm)
model_lm<- stan_glm(Y~X, data, family=gaussian())
```

In a model, priors can be set but there are also default weakly informative priors that can be used instead. Running the command prior_summary(model_lm) gives a summary of the priors used in the model. The model summary is given by summary(model_lm). Trace plots are created for each parameter in a model using plot(model_lm, “trace”) and posterior predictive checking plots are created using pp_check(model_lm).

```
summary(model_lm)
plot(model_lm, “trace”)
pp_check(model_lm)
```

For the Bayesian logistic mixed effects model, the repeated measure variable (like measuring over multiple time points) and subject ID variable (connects the multiple measurements for each subject) were added as random variables.

```
log_mixed_effects <- stan_glmer(Y~ X + (1| repeated_measure) + (1 | subject_id), data,
family=binomial(link= ”logit”))
```

Fixed variables are variables like X that we expect to have an impact on Y while random variables are grouping factor that we are trying to control (Hajduk, 2019). There

are multiple types of random variables. Random variables where the intercept varies by the random variable use the form (1 | random_variable) in the R code for a mixed effects or brms model (Bürkner & Vuorre, 2019). If the slopes vary by the random variable the R code is (0+ fixed_variable | random_variable) and if the slopes and intercepts vary by the random variable the R code is (1+ fixed_variable | random_variable) (Hajduk, 2019).

For a more thorough tutorial of the rstanarm R package, read Muth et al (2018). Muth et al (2018) includes both Bayesian linear regression and Bayesian hierarchical models with mixed effects.

To Run a Bayesian Regression Model using the brms Package:

I have found the rstanarm R package more user-friendly than the brms R package, but the brms R package has been developed for more types of models. Basic Bayesian linear regression models can be built using brms, but the focus here is on Bayesian ordinal models. Models built with the brm function in the brms R package, use the same Y~X format for the formula like the rstanarm R package and the lm function for frequentist linear regression (Bürkner, 2017). The three main types of ordinal models are cumulative models, sequential models, and adjacent-categories models (Bürkner & Vuorre, 2019). The below model is a cumulative model, but a sequential model can be run by changing cumulative() to sratio() and an adjacent-categories model by changing it to acat() (Bürkner & Vuorre, 2019).

```
library(brms)
brmsmodel <- brm(Y~X, data, family=cumulative())
```

The brms package uses the same commands as the rstanarm package for a summary of model priors, model summary, and posterior predictive checking plots. The brms plot function works a little differently because enter must be hit after running the

command to see the plots. By default, the rstanarm and brms plot functions display the plots for all the parameters, which can be too small to read if there are a lot of parameters. However, parameters can be specified to limit the number of plots and increase their size.

```
summary(brmsmodel)
plot(brmsmodel)
pp_check(brmsmodel)
```

How to Apply the ROPE-HDI approach:

The ROPE-HDI approach was applied using the bayestestR package (Makowski, Ben-Shachar, & Lüdecke, 2019). ROPE ranges were calculated using the rope_range function in the bayestestR package. The treatment level contrasts were calculated using the estimate_contrasts function in the modelbased R package (Makowski et al., 2020). The ROPE-HDI decision rule as it is described in Kruschke (2014) was applied by setting test = “equitest” and rope_ci= 0.95 (Makowski et al., 2020). If you do not set rope_ci, it defaults to one due to a simulation study done by the collaborators of these two packages (Makowski, Ben-Shachar, Chen, et al., 2019). The ROPE-HDI decision rule can still be applied when the rope_ci defaults or is set to one, but then it is not done the way it is mentioned in Kruschke (2014).

```
library(bayestestR)
library(modelbased)
library(logspline)
rope_range(model)
estimate_contrasts(model, test="equitest", ci=0.95, rope_ci=0.95, cores=4)
```