final

Robin

April 28, 2019

# Introduction

The data consist of randomized trials before 1980 of corticosteroid therapy in premature labor and its effects on neonatal death. This experiment has 7 trails where each trial consists of a treatment group and control group. Within each group the total count of patients was recorded along with the corresponding mortality for each group(events). The goal of my experiment is to first consult an authoritative source for an appropriate Bayesian meta-analysis method to model this data. Defend the use of this method. Simulate type 1 errors based on the chosen method. Finally, apply chosen method to the data and summarize findings.

# Methods

This study had lots of information generated for each trial so using meta-analysis is an advantage because the method allows the researcher to pool information to determine the effectiveness of the therapy on the population (*d)*. When considering this type of model, I though about the therapy effects varying across the 7 trials. If we assume that that each trial was carried out in different medical centers within different locations. The researchers might want to deploy their study on females of different child bearing ages within different environmental conditions, either rural or urban. Where some trails might have a higher segment of one group over the other group. For example, Block trial might be conducted in a location where most of the females might be in their early 20-30 group while in another location like Auckland, females might be in their 35 -45 age group.

Based on my assumptions the data for each trial is independent of the order in which they were carried out over the population of interest and can be considered exchangeable. This exchangeability points to the effects of the therapy from each trial coming from a common distribution. As a result of this reasoning I preferred a Bayesian meta- analysis random effects model. Let be the number of trials.

***y0i ~ bin(ci, n0i)***

***y1i ~ bin(ti, n1i)***

***logit(ci) = μi***

***logit(ti) = μi + δi***

***μi ~ norm(0, .0001), δi ~ norm(d, τ), d ~ norm(0, .0001), τ = σ−2, σ ~ uni(0, 2), or = exp(d)***

In the model’s data layer there are two likelihoods, one likelihood for the control group() and the other for the treatment group().Within each group we have the total number of patients and the corresponding events, so binominal likelihoods were used. Where the total number of patients for the control group is and for the treatment group . and are the probabilities in the two likelihoods. In the process layer we use logit function and so the probabilities of success are modelled on a logit scale. are trial specific baselines representing the log odds of the outcome in the control group. While are the trial specific log odds ratio in the control group compared with the treatment group. Where , the prior for therapy effects and prior for between trials standard deviations are vague or uninformative because I wanted to keep the focus on the data.

# Computation

The experiment was conducted in R and JAGS. I wrote a R function for my chosen model in JAGS that accept 4 vector parameters. These 4 parameters are the treatment event, treatment total, and control event, and control total. This data was packaged into a list along with the common length of all 4 parameters and passed into JAGS. JAGS was set up to run 2 chains with 20000 iterations for each chain where 10000 were cast away as burn-ins. Resulting in a total of 20000 posterior samples for all parameters with 5 thinning. With seed set to 1, the experiment for type 1 error was carried out in 3 separate loops using 1000, 2000, and 3000 iterations.

Within these loops a random generator function for generating the events using rbinom was used for both the controlled and treatment group. Then the results were passed into the Bayesian function that defines my chosen meta-analysis. Convergence and sample size were randomly checked for a few of the iterations in the loop All samples checked showed the Gelman test for parameter was less than 1.1 for both the point estimates and upper CI. Also, sample size was large. As a result of the Gelman test and the large sample size, there is good convergence.

# Simulation

|  |  |  |  |
| --- | --- | --- | --- |
| **Type 1 for pooled difference between Treatment (*d***) | | | |
| Number of Iterations | 1000 | 2000 | 3000 |
| Type 1 Error | 0.01 | 0.01 | 0.01 |

For the type 1 error, a check was implemented within my simulation for 0 within the 95% CI for each simulated dataset. If 0 was within the interval, these CI were not counted. Only CI where 0 was not in the credible set was counted and the proportion is shown in the table above based on the number of iterations. Overall, my type error is .01. I was not expecting the type 1 error to be .01 at the 95% but .05.