Final

Robin Baldeo
ST 540

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.

Method

Meta-analysis is useful when we must combine the results of several studies. This method allows the researcher to pool information to determine the effectiveness of the therapy on the population. However, if for some reason this information is incorrect, we can have bias or misleading conclusions.

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 $i = 1, \ldots, 7$ be the number of trials.

$$Y0_i \sim bin(c_i, N0_i)$$

$$Y1_i \sim bin(t_i, N1_i)$$

$$logit(c_i) = \mu_i$$

$$logit(t_i) = \mu_i + \delta_i$$

$$\mu_i \sim norm(0, .0001),\ \delta_i \sim norm(d, \tau),\ d \sim norm(0, .0001),\ \tau = \sigma{-}2,\ \sigma \sim uni(0, 2),\ or = exp(d)$$

In the model's data layer there are two likelihoods, one likelihood for the control group($Y0_i$) and the other for the treatment group($Y1_i$).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 $N0_i$ and for the treatment group $N1_i$. c_i and t_i are the probabilities in the two likelihoods. In the process layer we use logit function on c_i and t_i so the probabilities of success are modelled on a logit scale. μ_i are trial specific baselines representing the log odds of the outcome in the control group.

While δ_i are the trial specific log odds ratio in $N0_i$ compared with the $N1_i$.. Where μ_i, d the prior for therapy effects and σ prior for between trials standard deviations are vague or uninformative because the focus should be 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 d 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.

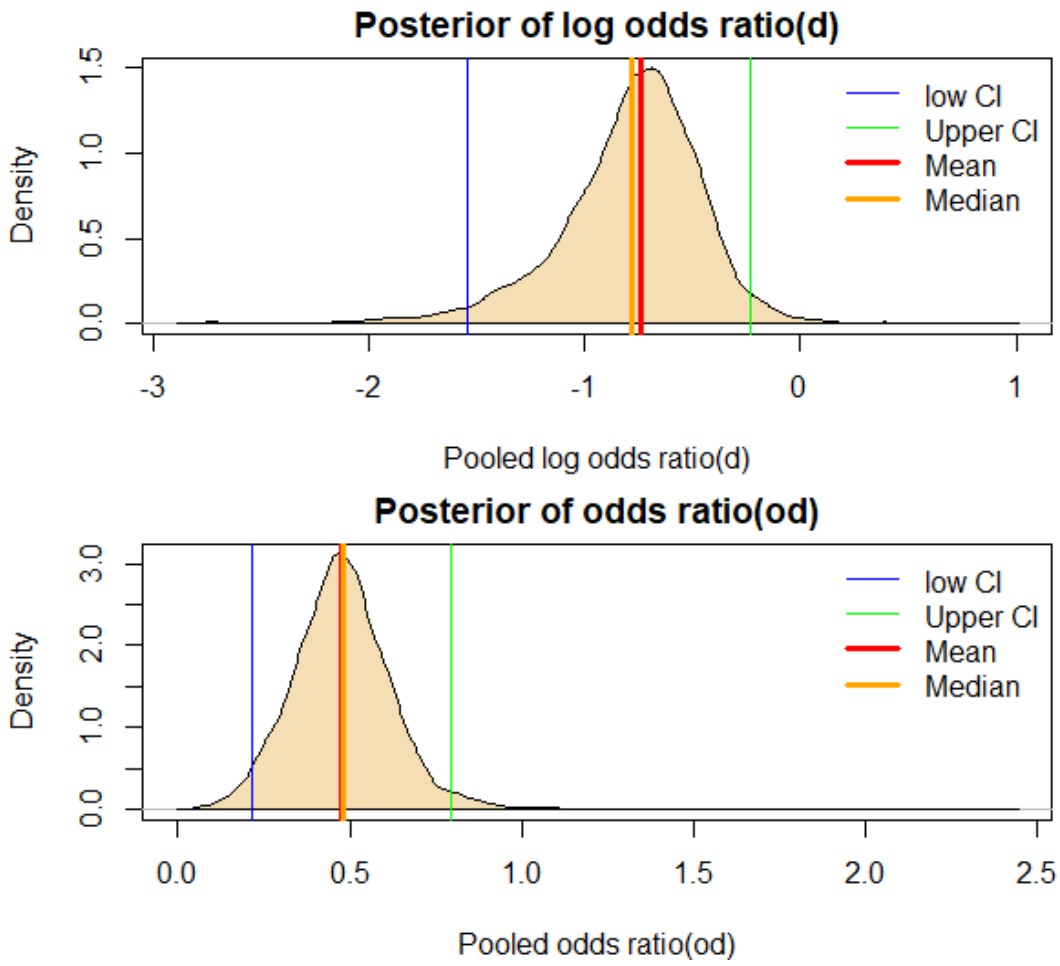
Simulation

Type 1 for pooled differences between Treatment (*d*)			
# iterations	1000	2000	3000
Type 1	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 interval was counted and the proportion is shown in the table above based on the number of iterations. Overall, my type 1 error is .01. I was expecting the type 1 error to be .05 at the 95% instead of .05.

Application to Cochrane data

My method produced effective sample size of 3203 for d and 3710 for or. Gelman test for this method is less than 1.1 for the point estimate and upper CI and because of the high sample and Gelman results there is good convergence. The pooled log odds ratio d median is -.74 with a 95% CI of -1.58 and -.22. The pooled odds ratio or has a median of .48 with a 95% CI of .21 and .80. In the plot below, the median and the means are near identical for each of the 2 parameters.



Since 0 is not within my CI for both parameters, these parameters are statistically significant indicating there is a difference between $N1_i.$ and the $N0_i.$. The negative value for d indicates that the odds of mortality under the treatment is much lower compared to the control. OR median of .48 indicates that treatment reduces neonatal mortality by 52% when compared with the control. As a result, based on the data, corticosteroid therapy is effective in premature labor, reducing neonatal mortality by 52%.

Appendix

Reference

Dias, Sofia. Network Meta-Analysis for Decision-Making (Statistics in Practice) (p. 19-31). John Wiley & Sons, 2018

JAGS Code

```
#function to run Bayesian method
 model<-function(et, nt, ec, nc){
  n<-length(et)
  data <- list(n=n, y1= et, n1 = nt, y0= ec, n0= nc)
  model_string <- textConnection("model{

 # Likelihood
   for(i in 1:n){
      y0[i]~dbin(c[i], n0[i])
      y1[i]~dbin(t[i], n1[i])

 #process
      logit(c[i])<- mu[i]
      logit(t[i])<- mu[i] + delta[i]

      mu[i]~dnorm(0, .0001)
      delta[i]~dnorm(d,taub)
   }

   # Priors
   d ~ dnorm(0,0.0001)
   taub <- 1/(sigma*sigma)
   sigma~ dunif(0,2)
   or <- exp(d)

}")

  model <- jags.model(model_string,data = data, n.chains=2,quiet=TRUE)
  update(model, 10000, progress.bar="none")
  params <- c( "d", "sigma", "or")
  samples <- coda.samples(model,variable.names=params,n.iter=20000,thin=5, progress.bar="none")
  return(samples)
}

#simulation of data
#function to generate random numbers
random<- function(nt, nc){
  n<- length(nt);  t<-rep(NA, n);  c<-rep(NA, n)
  for( i in 1:n ){
    t[i]= rbinom(1,nt[i], .1);    c[i]= rbinom(1,nc[i], .1)
  }
  return(cbind(t,c))}
```