Final project Statistical Methods in Data Science II

Pietro Fronte

February 2019

Contents

1 Introduction

This work was carried out with the aim of

1. explain the overall features of the statistical model such as the role of the parameters and the inferential goals of the analysis

2. check the ability of a fully Bayesian analysis to recover model parameters with data simulated from the model

2 Beta-Blocker Study

2.1 Introduction to Beta-Blocker dataset

The data comes from Yusuf et al. [1] and are presented as 22 contingency table related to the same number of clinical trials. Each trail, numerically speaking, is represented by a contingency table as the one reported in 1. The main idea is to compare mortality rate between treated and control group to better state the effectiveness of beta-blocker in this specific disease.

	Death	# of patients
Control	c_d	c_t
Treated	t_d	t_t

Table 1: 2x2 contingency table

Beta blockers are a class of drug most often used to treat hypertension. A number of studies have looked at the efficacy of beta blockers in preventing death after a myocardial infarction (heart attack). The Beta-Blocker study is an example of clinical trials study with case-treatment groups. Here follow the data found in public literature [2].

Clinical Trial	Control	Treatment
1	3/38	3/39
2	7/114	14/16
3	5/ 69	11/93
4	102/1533	127/520
5	28/355	27/365
6	4/59	6/52
7	98/945	152/939
8	60/632	48/471
9	25/278	37/282
10	138/1916	188/1921
11	64/873	52/583
12	45/263	47/266
13	9/291	16/293
14	57/858	45/883
15	25/154	31/147
16	33/207	38/213
17	28/251	12/122
18	8/151	6/154
19	6/174	3/134
20	32/209	40/218
21	27/391	43/364
22	22/680	39/674

Table 2: Beta-Blockers data

In this project we will discuss 3 different models related to these data. The first two model are kind of similar, since we focus essentially on the inference of the random effect between treatment and control group assigning to this parameter two different distribution. The third model, instead, is focused on the inference of the log odds ratio between the treatment and the control group (meta-analysis); the model used is different than the previous ones.

2.2 Model 1

$$r_i^c \sim Binomial(p_i^c, n_i^c) \tag{1}$$
$$r_i^t \sim Binomial(p_i^t, n_i^t) \tag{2}$$
$$logit(p_i^c) = \mu \tag{3}$$
$$logit(p_i^t) = \mu + \delta_i \tag{4}$$
$$\mu \sim Normal(0.0, 10^{-5}) \tag{5}$$
$$\delta_i \sim Normal(\nu, \tau) \tag{6}$$
$$\delta_{new} \sim Normal(\nu, \tau) \tag{7}$$
$$\nu \sim Normal(0.0, 10^{-6}) \tag{8}$$
$$\tau \sim Gamma(0.001, 0.001) \tag{9}$$
$$\sigma = \frac{1}{\sqrt{\tau}} \tag{10}$$
$$\tag{11}$$

The number of deaths, respectively in control and treatment group, is controlled by a Binomial distribution, each of them with the appropriate mortality rate p_i^c, p_i^t. On the other hand, the mortality rate, translated into the corresponding log odds, it is supposed to assume generic values expressed in the model by a Non-infromative prior Normal distribution μ with mean 0 and very big variance [1](in the model above reported the parameter for the Normal distribution are mean and precision). Notice that having applied the log odds to the probabilities we have created a mapping of the values as follow:

$$0 \leq p_i^c \leq 1 \quad -\infty \leq logit(p_i^c) \leq +\infty \tag{12}$$
$$0 \leq p_i^t \leq 1 \quad -\infty \leq logit(p_i^t) \leq +\infty \tag{13}$$

As can be seen from the model written above the logit of probabilities are expressed by the same distribution μ, however we are assuming that there is a random effect δ that affect the treatment group. It is, indeed, δ that we want to infer and understand how its values are distributed, as well as the predictive distribution of the effect δ_{new} in a new hypothetical trial and the variance σ. Both random effects are assumed to be distributed as Normal with mean ν and precision τ. For ν and τ then a non-informative prior distribution is chosen (Normal with mean 0 and really small precision; Gamma distribution with 0.001 for shape and scale parameter).

First of all, to have a visual idea on how the data appear, we plot into an histogram (Figure 1) the log odds of the probabilities coming from each contingency table as reported in 4.

Clinical Trial	Control	Treatment
1	-2.456736	-2.484907
2	-2.726919	-1.985915
3	-2.549445	-2.008824
4	-2.6411	-2.395028
5	-2.457756	-2.527209
6	-2.621039	-2.036882
7	-2.156733	-1.644348
8	-2.254794	-2.176171
9	-2.314514	-1.890340
10	-2.555991	-2.221167
11	-2.536916	-2.323518
12	-1.577833	-1.538924
13	-3.444682	-2.851429
14	-2.642810	-2.924356
15	-1.640937	-1.319603
16	-1.662548	-1.527200
17	-2.074967	-2.215574
18	-2.883403	-3.205453
19	-3.332205	-3.776585
20	-1.710414	-1.492904
21	-2.601317	-2.010241
22	-3.398162	-2.790063

Table 3: Log odds of probabilities of each clinical trial

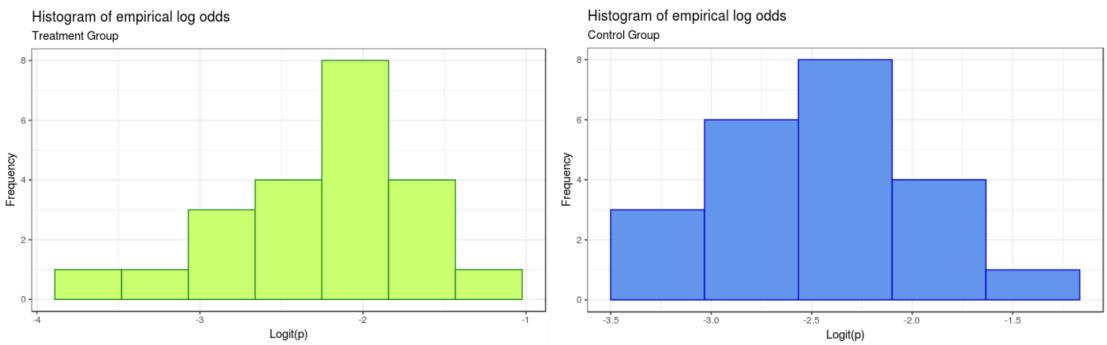

Figure 1: Histograms of empirical log odds in treatment and control groups

We plot first the prior distribution on ν and τ, Figure 2 3. In order to plot the Normal prior distribution of δ, Figure 4, we need a point estimate of expected value (ν) and variance ($\sigma^2 = \frac{1}{\tau}$); we consider then the expected value of each parameter according to their distribution.

$$\delta \sim Normal(\nu, \tau) \tag{14}$$
$$\nu \sim Normal(0.0, 10^{-6}) \tag{15}$$
$$\tau \sim Gamma(0.001, 0.001) \tag{16}$$
$$\tag{17}$$

$$\nu \sim Normal(0.0, 10^{-6}) \quad \mathbb{E}(\nu) = 0 \tag{18}$$
$$\tau \sim Gamma(0.001, 0.001) \quad \mathbb{E}(\tau) = \frac{0.001}{0.001} = 1 = \frac{1}{\tau} = \sigma^2 \tag{19}$$
$$\tag{20}$$

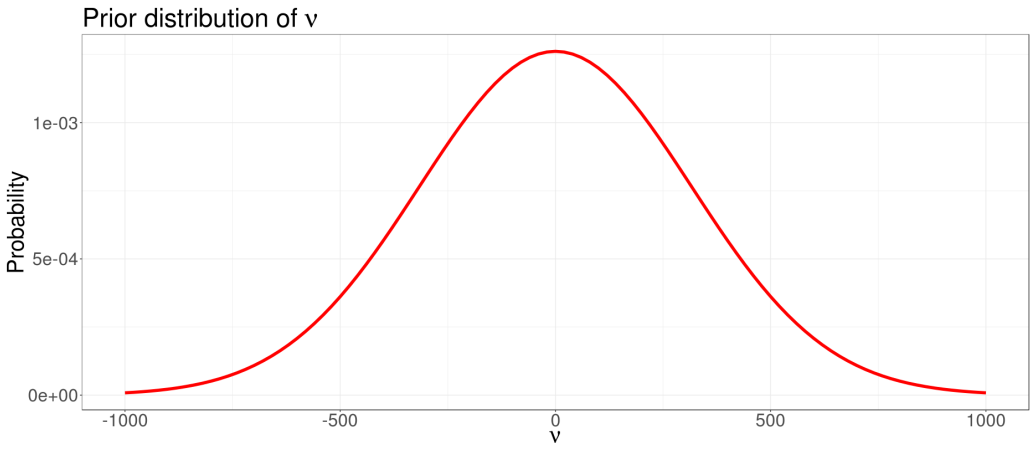

Figure 2: Prior distribution of ν

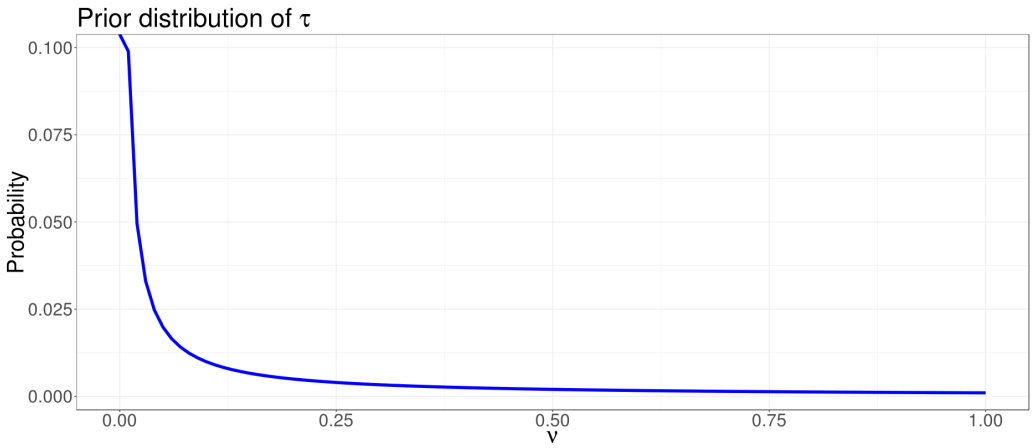

Figure 3: Prior distribution of τ

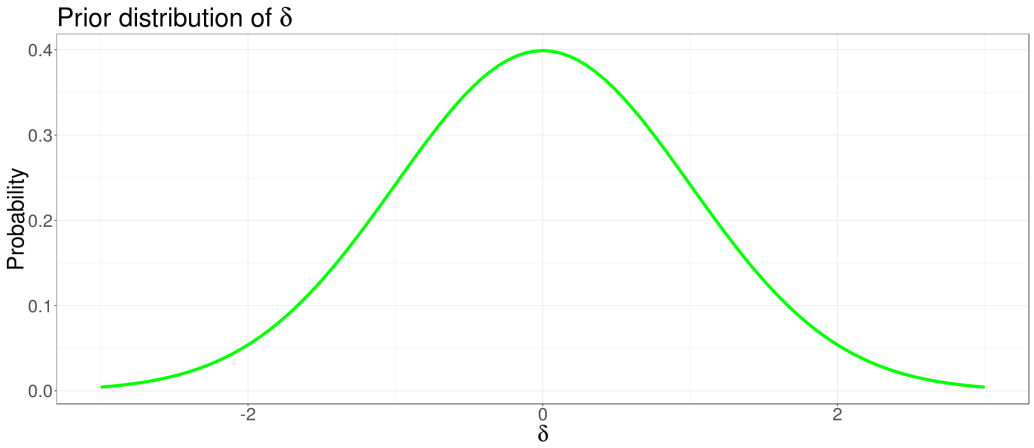

Figure 4: Prior distribution of δ

2.3 Model 2

The second model is pretty close to the first one. The only difference between them is that the random effect δ is suppose to still have a bell-shaped distribution but with heavier tails; for that reason we suppose δ to follow a non-standardized t-Student distribution with parameter ν (mean) τ (precision) and k (degree of freedom).

$$x \sim t - Student(\nu, \tau, k) = \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})}\sqrt{\frac{\tau}{k\pi}}\Big(1 + \frac{\tau(x-\nu)^2}{k}\Big)^{-\frac{k+1}{2}} \tag{21}$$

Keeping the same distributions for the mean μ (Normal(0,1)) and the precision τ (Gamma(0.001,0.001)) we can have a look to the prior distribution of δ and compare it with the previous one 5. As said

before with the t-Student distribution we want to consider δ's distribution to have heavier tails; in 5 is not that clear so zooming the tails (also in log-log scale to emphasize the gap) was the solution to check this property 6.

$$r_i^c \sim Binomial(p_i^c, n_i^c) \tag{22}$$
$$r_i^t \sim Binomial(p_i^t, n_i^t) \tag{23}$$
$$logit(p_i^c) = \mu \tag{24}$$
$$logit(p_i^t) = \mu + \delta_i \tag{25}$$
$$\mu \sim Normal(0.0, 10^{-5}) \tag{26}$$
$$\delta \sim t - Student(\nu, \tau, 3) \tag{27}$$
$$\delta_{new} \sim t - Student(\nu, \tau, 3) \tag{28}$$
$$\nu \sim Normal(0.0, 10^{-6}) \tag{29}$$
$$\tau \sim Gamma(0.001, 0.001) \tag{30}$$
$$\sigma = \frac{1}{\sqrt{\tau}} \tag{31}$$
$$\tag{32}$$

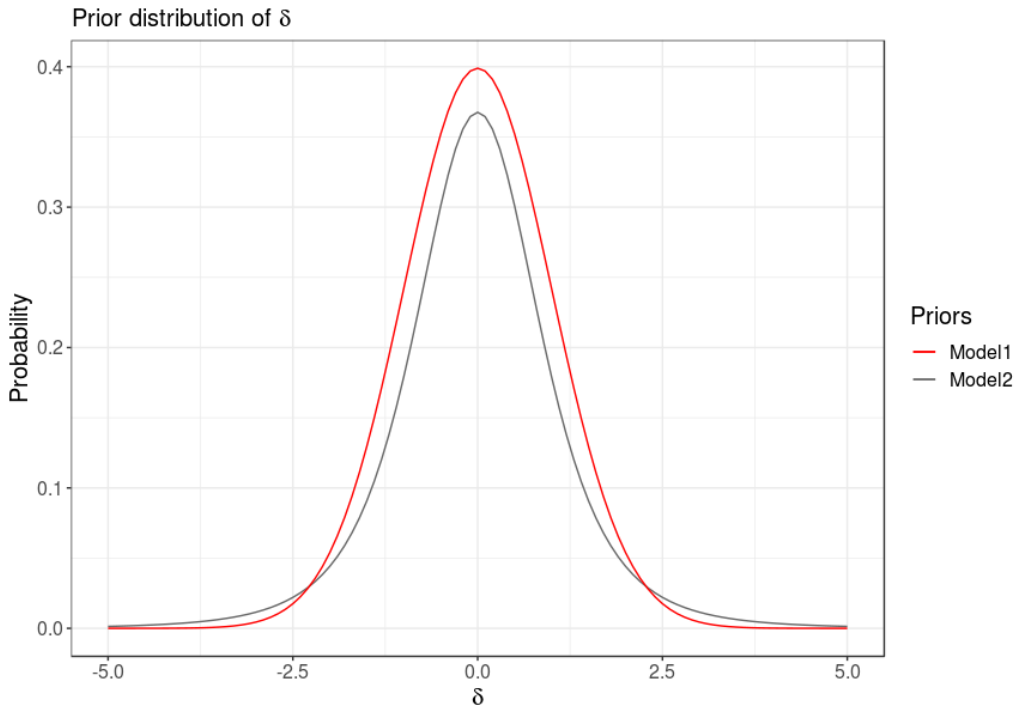

Figure 5: δ priors in model 1 and model 2

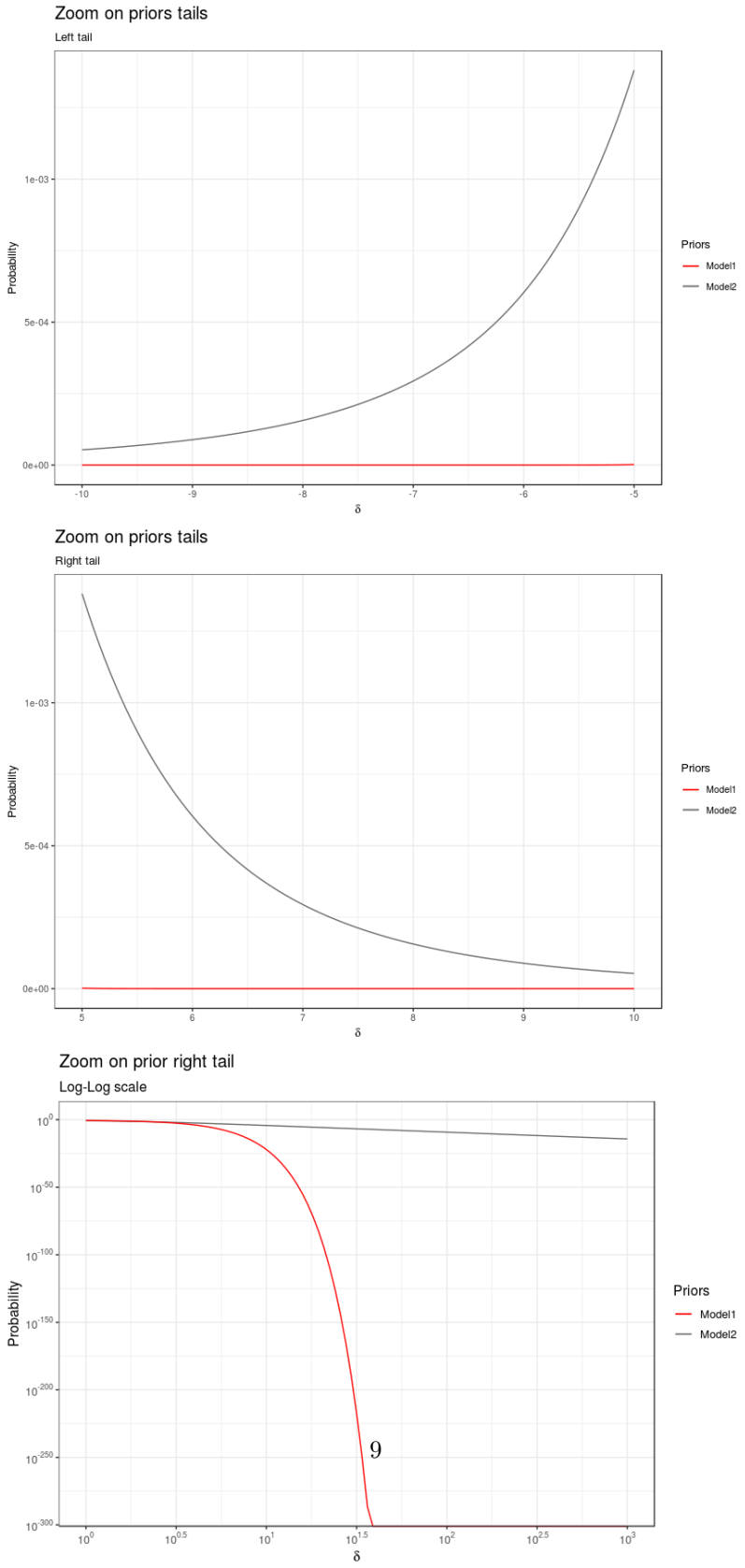

Figure 6: Tails zoom of prior and posterior of δ

2.4 Model 3

The third model, as anticipated before, belongs to the so-called meta-analysis branch of statistics field [3]. Meta-analysis is the statistical procedure for combining data from multiple studies. When the treatment effect (or effect size) is consistent from one study to the next, meta-analysis can be used to identify this common effect. The data are combined in an unique, relatively abstract, value that is the log odds ratio defined as

$$\Delta_i = log\left(\frac{r_i^t}{n_i^t - r_i^t}\right) - log\left(\frac{r_i^c}{n_i^c - r_i^c}\right) \tag{33}$$

$$= log\left(\frac{\frac{r_i^t}{n_i^t - r_i^t}}{\frac{r_i^c}{n_i^c - r_i^c}}\right) \tag{34}$$

$$= log\left(\frac{odds_{treatment}}{odds_{control}}\right) \tag{35}$$

where r_c and r_t are respectively the number of events registered in control and treatment group; n_c and n_t are instead the population size of each trial i. We can also retrieve an approximation of the standard error for each log odds ratio value

$$\sigma(\Delta_i) = log\left(\frac{1}{r_i^c} + \frac{1}{n_i^c - r_i^c} + \frac{1}{r_i^t} + \frac{1}{n_i^t - r_i^t}\right) \tag{36}$$

$$\Delta_i \sim N(\delta_i, \sigma^2(\Delta_i)) \tag{37}$$

$$\delta_i \sim N(d, prec) \tag{38}$$

$$d \sim N(0, 10e^-6) \tag{39}$$

$$\tau \sim Unif(0, 10) \tag{40}$$

$$prec = 1/\tau^2 \tag{41}$$

The odds ratio usually range between (0,∞); applying the log to those value we standardize them around 0. For that reason we assume them to follow a normal distribution whose mean has a Normal prior distribution with hypeparameters 0 for the mean and a very large variance (to give a non-informative prior). We assume the variance of d to be distributed as Uniform in the interval (0,10). The variance of the log-odds are known and evaluated with (36).

Evaluating the odds we actually want to check if the odds of the event "death" are greater either in the Control or Treatment case. The odds ratio range from 0 to ∞ with twofold meaning: all values in the interval (0,1), 1 excluded, mean that the odds is greater in Control group than the Treatment; in case of 1 it means that the event has the same probability to show up in both groups; values from 1 to $+\infty$ mean that the odds of group treatment are greater than the control group. The log-odds show the same information with values both positive and negative centered in 0.

Clinical Trial	OR	$\sigma(OR)$
1	0.02817088	0.72301587
2	-0.74100320	0.23343543
3	-0.54062120	0.31872921
4	-0.24612808	0.01909462
5	0.06945337	0.07876801
6	-0.58415690	0.45658762
7	-0.51238549	0.01923431
8	-0.07862326	0.04161232
9	-0.42417337	0.07506123
10	-0.33482340	0.01370499
11	-0.21339753	0.03797510
12	-0.03890844	0.05265218
13	-0.59325371	0.18076732
14	0.28154593	0.04220784
15	-0.32133359	0.08863069
16	-0.13534792	0.06808023
17	0.14060645	0.13262283
18	0.32204972	0.30541643
19	0.44438052	0.51358597
20	-0.21750973	0.06751770
21	-0.59107599	0.06615537
22	-0.60809913	0.07419013

Table 4: log OR and log OR variance

3 Monte Carlo Markov Chain with JAGS

Using Jags for Rstudio and the model available in the public literature [2] I run 3 different chains of 20000 iterations with 2000 burn-in iterations; the iteration saved are then only 18000 on a single chain. Before listing the models implemented and the results obtained it is needed an introduction about the metric we are going to check in order to choose the best models among the proposals.

3.1 Number effective sample size and Gelman-Rubin $\hat{R}$

The number of effective sample size is defined in [4] as the number of independent samples with the same with the same estimation power as the original N auto-correlated samples.

$$n_{EFF} = \frac{n}{1 + 2\sum_k^{\infty} \rho(k)} \tag{42}$$

If $\rho(k)$ is positive and decreases too slowly the sum at the denominator diverges and $N_{EFF} = 0$; if $\sum_k^{\infty} \rho(k)$ is negative then the N_{EFF} may result grater than the original N. Usually we expect N_{EFF} to be smaller than the original chain length N. In case of more than one chain generated, let's say m chains, we expect $N_{EFF} < mN$. In case all samples are independent $n_{EFF} = N$ and consequently $\rho(k) = 0\ \forall$k. However, the MCMC is generating correlated samples, due to the Markov chain nature of the sampler - that is, since every sample generated is dependent on its predecessor a correlation between samples is introduced.

For what concern the $\hat{R}$ is just a diagnostic measure of convergence when below 1.1 [5]. At convergence $\hat{R} = 1$.

3.2 Deviance

The deviance is a goodness of fit for a statistical model when the model fitting is achieved by maximum likelihood. The total deviance for a model may be defined as follow:

$$D(y, \hat{\mu}) = 2\big(log(p(y|\hat{\theta}_s)) - log(p(y|\hat{\theta}_0))\big)$$

where $\hat{\theta}_0$ denotes the fitted values of the parameters in the model M_0 while $\hat{\theta}_s$ denotes the fitted parameters for the saturated model. With saturated model we refer to a model with a parameter for each observation in order to achieve a perfect fitting of the data [6]. Intuitively, can be though as the sum of squares of residuals in OLS (ordinary least squares).

3.3 Effective number of parameters of the model

The larger p_D is, the easier it is for the model to fit the data. The effective number of parameters of the model implemented in JAGS package is $p_D = \frac{var(D)}{2}$ as defined in Gelman et al.[7] and [8].

3.4 DIC - Deviance Information Criterion

The deviance information criterion (DIC) is a hierarchical modeling generalization of the Akaike information criterion (AIC). It is particularly useful in Bayesian model selection problems where the posterior distributions of the models have been obtained by Markov chain Monte Carlo (MCMC) simulation. In JAGS the DIC is evaluated as $DIC = \bar{D} + p_D$, where $\bar{D}$ is the expected value of the

deviance and p_D is the effective number of parameters as defined in section 3.3. When comparing different models will be chosen the one having a lower DIC.

4 Results

4.1 Model 1

In the first model we are considering the random effect δ, as well as the random effect δ_{new} for an unseen clinical trial,distributed as a Normal distribution with a mean ν and a precision τ.

```
for( i in 1 : Num ) {
    rc[i] ~ dbin(pc[i], nc[i])
    rt[i] ~ dbin(pt[i], nt[i])
    logit(pc[i]) <- mu[i]
    logit(pt[i]) <- mu[i] + delta[i]
    mu[i] ~ dnorm(0.0,1.0E-5)
    delta[i] ~ dnorm(nu, tau)
}
nu ~ dnorm(0.0,1.0E-6)
tau ~ dgamma(0.001,0.001)
delta.new ~ dnorm(nu, tau)
sigma <- 1 / sqrt(tau)
```

4.1.1 Results, plots and comments about Model 1

```
Inference for Bugs model at "m1", fit using jags,
 3 chains, each with 20000 iterations (first 2000 discarded)
 n.sims = 54000 iterations saved
          mu.vect sd.vect    2.5%     25%     50%     75%    97.5%  Rhat n.eff
delta.new  -0.250   0.148  -0.552  -0.328  -0.253  -0.174    0.065 1.002  3500
nu         -0.250   0.062  -0.367  -0.292  -0.251  -0.210   -0.126 1.003   930
tau       230.529 386.938  12.735  40.831  91.099 238.819 1355.760 1.002  2400
deviance  262.061 120.505 244.882 254.035 259.220 264.719  276.744 1.001 24000

For each parameter, n.eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule, pD = var(deviance)/2)
pD = 7261.0 and DIC = 7523.1
DIC is an estimate of expected predictive error (lower deviance is better).
```

As can be seen from the model results printed in R we have some interesting results to comment. Talking about point estimates of the parameter saved we can see that both δ_{new} and ν are stuck to the value -0.250 with a tight variance for both. The precision τ seems to have a point estimate of around 230, however with a very large variance. Same conclusion we can have about the deviance..high value of deviance (so first symptom that our model is not that great) with high

standard deviation too. For what concern $\hat{R}$ the value seems pretty close to 1 so we can state that the chain reached the convergence. ν, τ and δ_{new} show, with this specific setup (20000 iterations with 2000 burn-in iterations and the seed set to 2), a really low number of effective sample size n_{EFF} considering the 20000 iterations. This means that the value registered for those parameters are highly auto-correlated and only few of them can be considered independent, having the same estimation power of the whole chain. 24000 is the effective number of sample size for deviance.
As said before the model shows an high deviance, high value of effective number of parameters p_D and DIC; for that reason we can state that this model is not a good representation of the true random effect δ. Traceplot and posterior density are shown below.

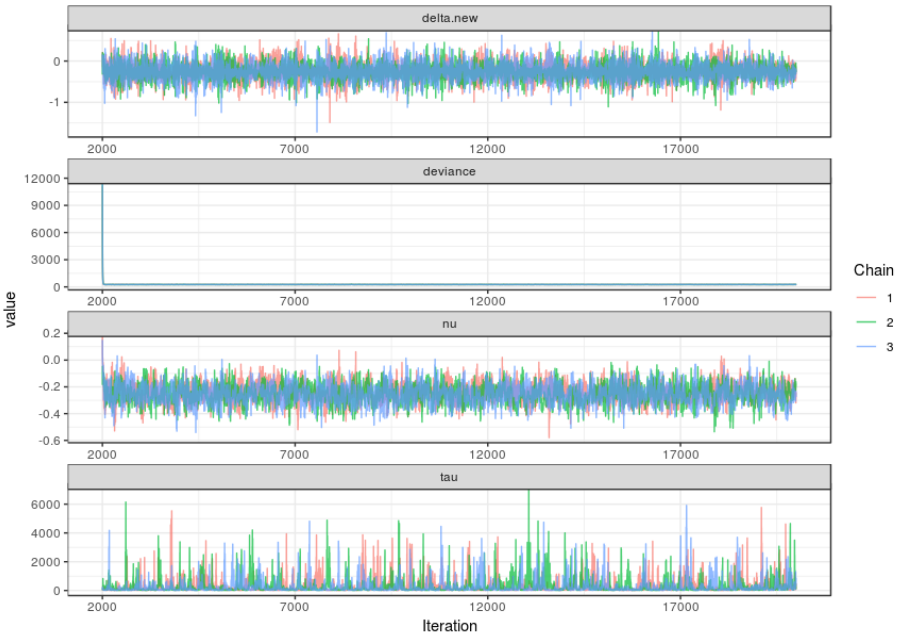

Figure 7: Traceplot of parameters

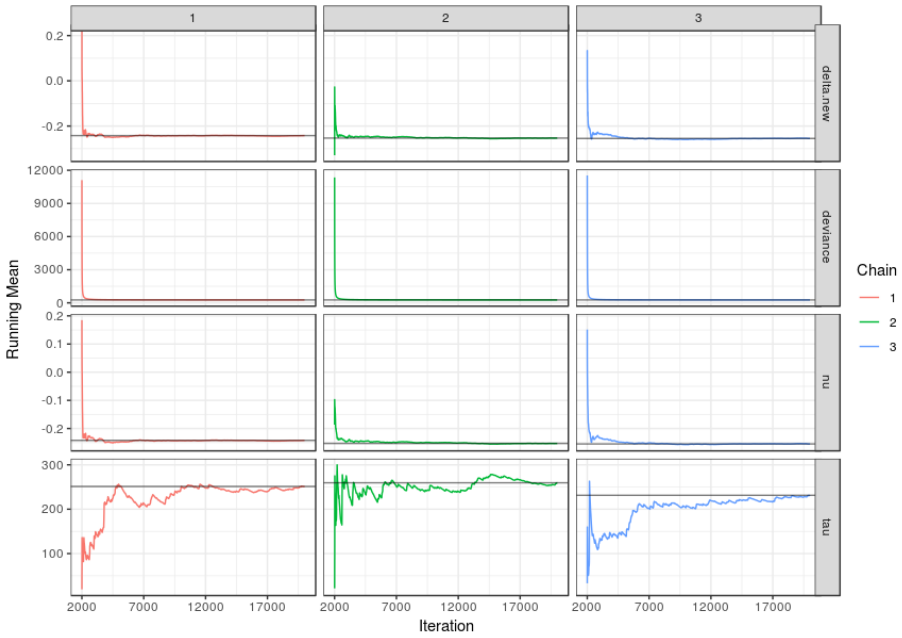

Figure 8: Traceplot of running means

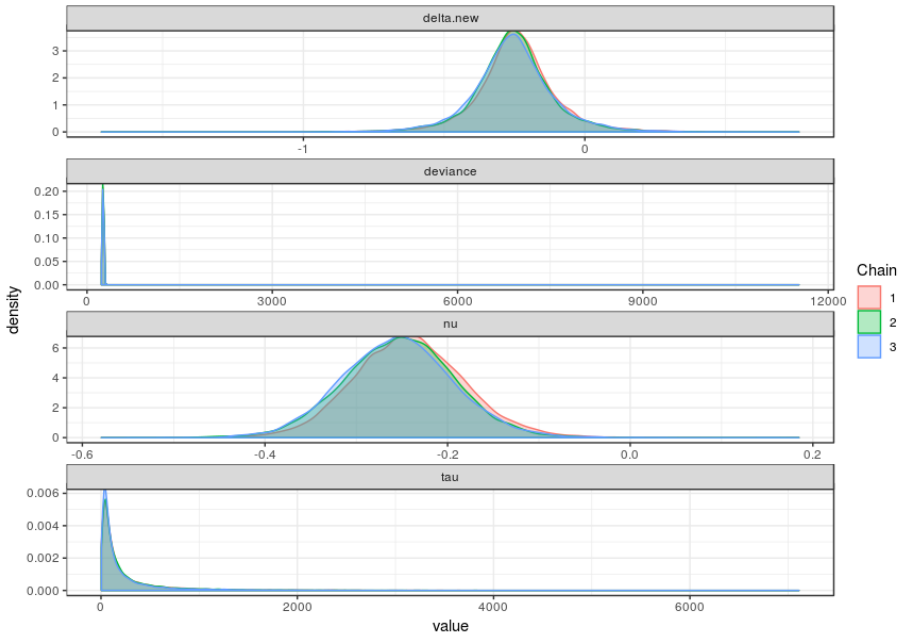

Figure 9: Posterior densities of parameters based on data

The plots confirm what previously found analytically. In Figure 9 we can also visually understand why deviance and τ show a so big variance. Deviance traceplot, however, result unclear in Figure 7 so zooming the deviance chains between iterations [1;50] and iterations [100;1000] is it possible to understand that the variance start with a huge values that quickly decrease to then stabilize around 260. This is actually the reason of the large standard deviation.

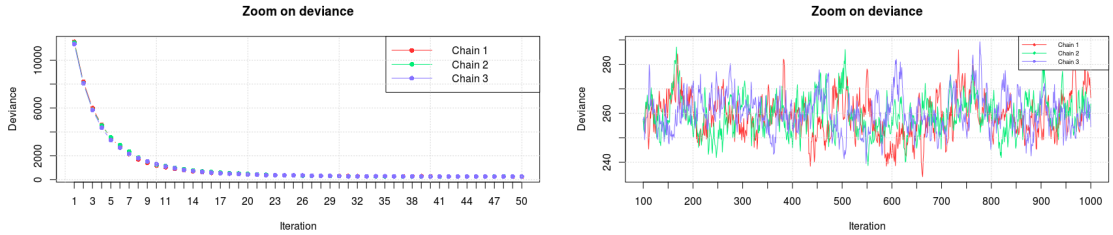

Figure 10: Zoom on deviance Chains

Going back to the auto-correlation topic we can have a look on the auto-correlation plots to have a double check on how our data are composed.

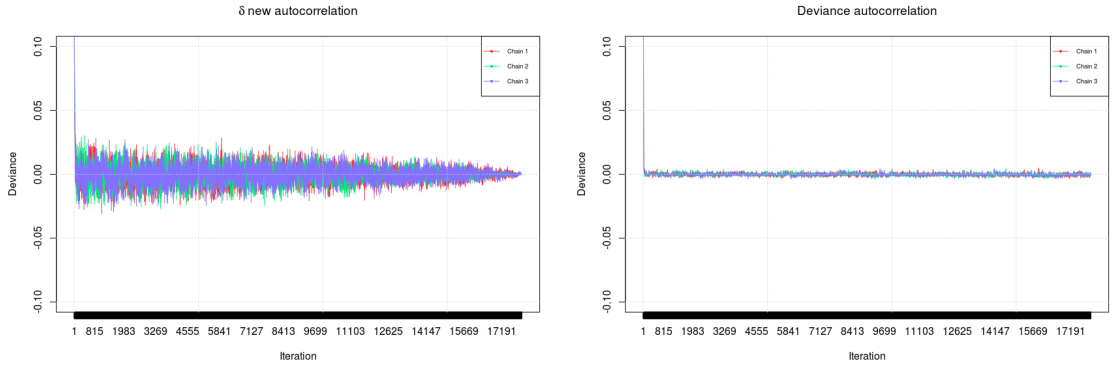

Figure 11: δ_{new} and deviance auto-correlation plot

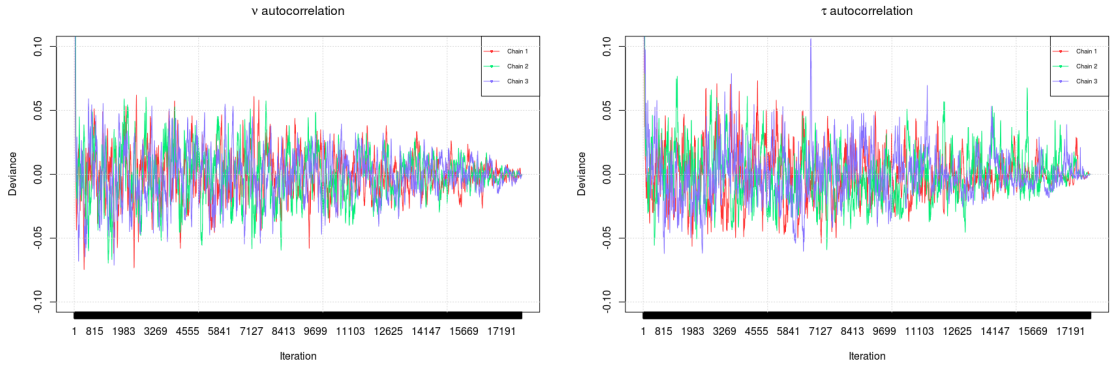

Figure 12: ν and τ auto-correlation plot

We can see how Deviance auto-correlation dies out really quickly, just after few iterations. The correlation is so weak that, as reported in 42, we end up having a large enough value of effective sample size N_{EFF}. On the other hand we can see how big and unstable are the values of the auto-correlation of the other parameters. This prove, in fact, the small value of n_{EFF}, number of effective sample size, obtained for δ_{new}, ν, τ. In Figure 13 we can see how the posterior distribution of δ looks like and have a comparison with the prior.

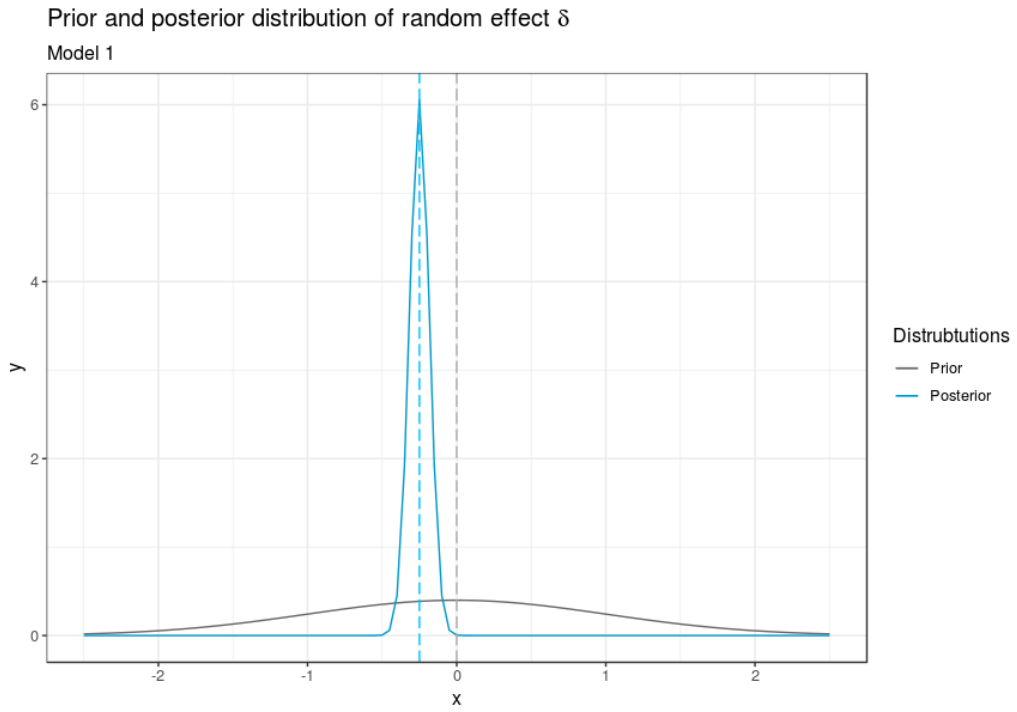

Figure 13: Prior and posterior distribution of δ with respect to model 1

4.2 Model 2

```
for( i in 1 : Num ) {
  rc[i] ~ dbin(pc[i], nc[i])
  rt[i] ~ dbin(pt[i], nt[i])
  logit(pc[i]) <- mu[i]
  logit(pt[i]) <- mu[i] + delta[i]
  mu[i] ~ dnorm(0.0,1.0E-5)
  delta[i] ~ dt(nu, tau, 3)
}
nu ~ dnorm(0.0,1.0E-6)
tau ~ dgamma(0.001,0.001)
delta.new ~ dt(nu, tau, 3)
```

4.2.1 Results, plots and comments about Model 2

```
Inference for Bugs model at "m2", fit using jags,
 3 chains, each with 20000 iterations (first 2000 discarded)
 n.sims = 54000 iterations saved
           mu.vect sd.vect    2.5%     25%     50%     75%    97.5%  Rhat n.eff
delta.new  -0.253   0.185  -0.598  -0.327  -0.252  -0.178    0.096 1.001 10000
nu         -0.252   0.063  -0.378  -0.294  -0.253  -0.211   -0.127 1.002  2200
tau       323.654 441.798  21.861  76.557 167.807 384.791 1566.208 1.002  1500
deviance  259.469   7.926 245.146 253.888 259.038 264.578  276.182 1.002  2900

For each parameter, n.eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule, pD = var(deviance)/2)
pD = 31.4 and DIC = 290.9
DIC is an estimate of expected predictive error (lower deviance is better).
```

As it is possible to check in the output reported above, in terms of point estimate of the parameters we face two different cases: on one hand we have estimate really close to those obtained with model 1 (ν, deviance, and δ_{new}) while on the other hand we have a point estimate for τ different from the previous one (around 323.6 in model 2 against 230.5 in model 1). This is actually a good factor for our bayesian analysis because with this model it seems we got a higher precision. We should look then to the other metrics to determine if this model is actually better or not than the previous one. In terms of standard deviation we have a small estimate for ν and δ_{new} meaning that -0.25 is a good expected value for those parameters. The variance of the deviance is now way smaller than the previous one and this is also reflected in the evaluation of p_D, the effective number of parameters of the model. Greater, instead, the variance of the precision τ. We can easily say that all of the parameters have converged ($\hat{R}$ is pretty close to 1). If we check the value of DIC, basing our conclusion only on this metric, we can say that this second model is absolutely better than the first one. This is due to the low value of variance registered for the deviance with respect to the previous model, even though the variance per se is not that small.

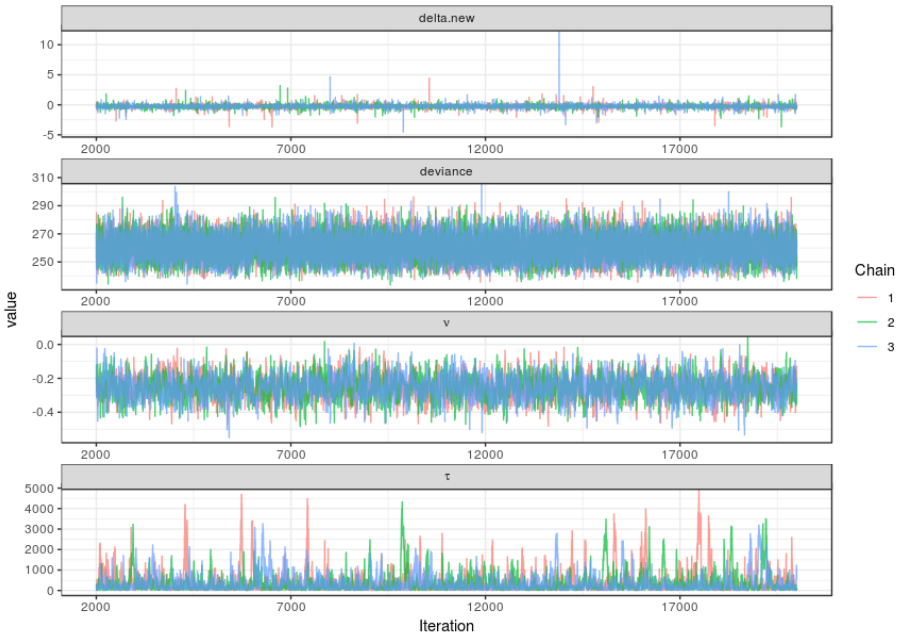

Figure 14: Traceplot of parameters

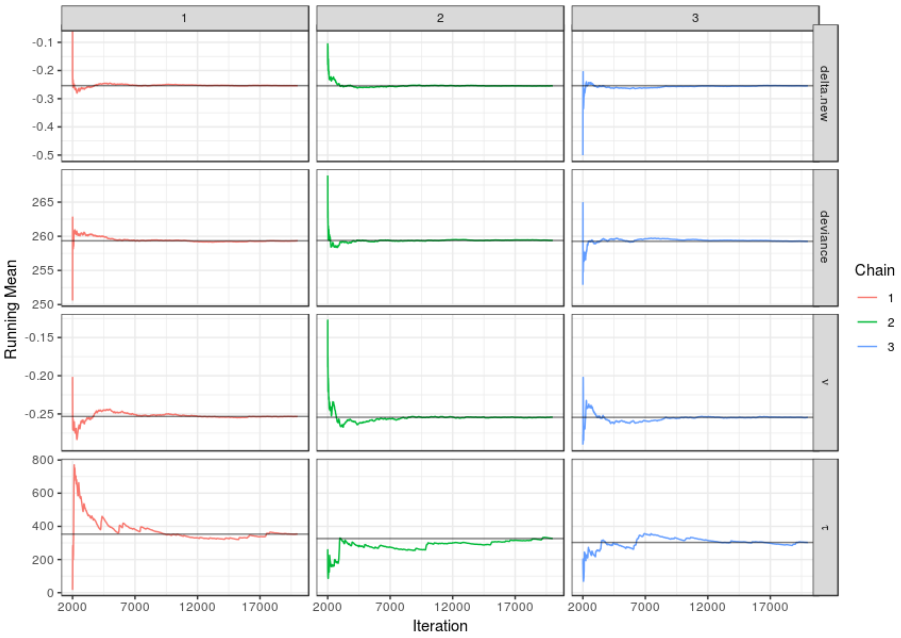

Figure 15: Traceplot of running means

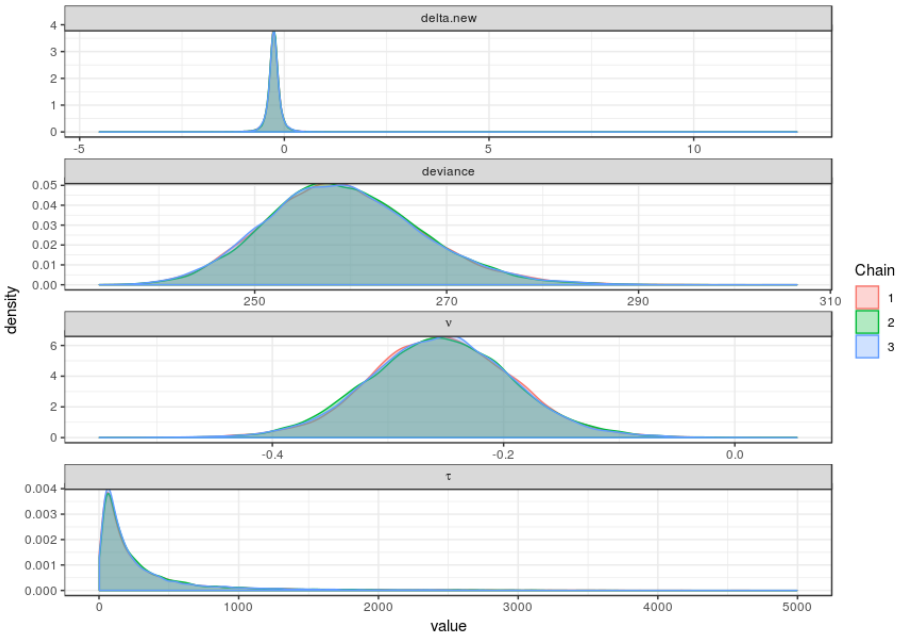

Figure 16: Posterior densities of parameters based on data

Talking about n_{EFF} the effective sample size we can see all the parameters show a strong positive autocorrelation as can be seen in Figures 17 18; this cause a small value of n_{EFF}.

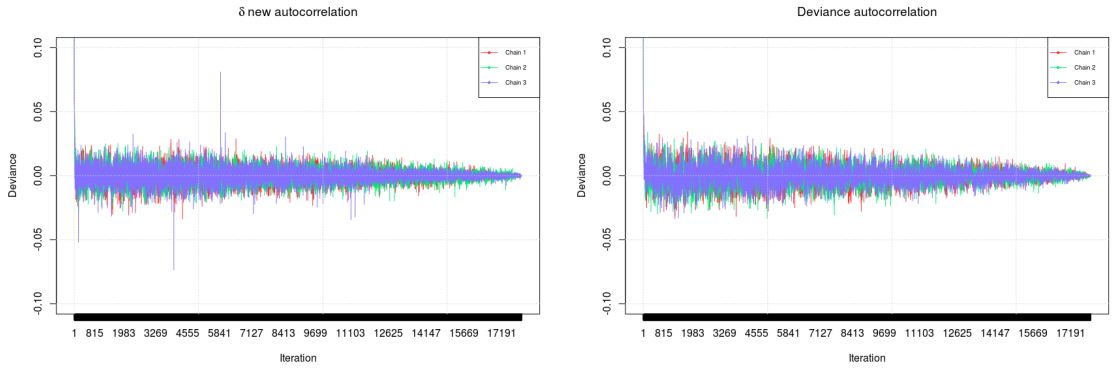

Figure 17: δ_{new} and deviance auto-correlation plot

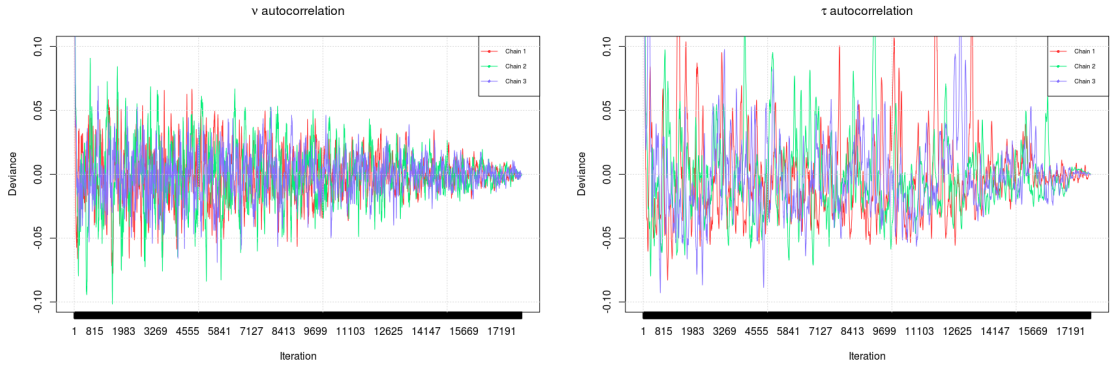

Figure 18: ν and τ auto-correlation plot

As can be seen in Figure 12 the parameter τ shows, unlike the other 3 parameters, high auto-cocorrelation values. Usually we expect the auto-correlation values to decrease as the number of iterations increase; in this specific case, even after 15000 iterations those value are still relatively high. This explain the low number of expected sample size in the JAGS output of model 2.

We can finally check how the posterior densities for the random effect δ look like and it is possible to compare them with the empirical density, even though the latter is based on only 22 values.

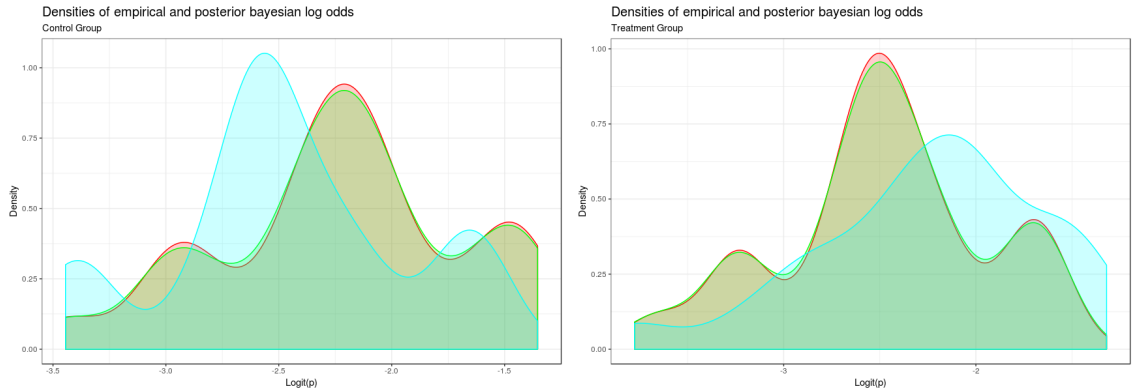

Figure 19: Posterior densities of δ based respectively on Model 1 (green) and Model 2 (red) against the empirical density (cyan)

4.3 Model 3

```
model {

# Likelihood

for (j in 1:Nstud) {
  P[j] <- 1/V[j]    # precision from calculated variance
  Y[j] ~ dnorm(delta[j],P[j])
  delta[j] ~ dnorm(d,prec)
}

# Priors
d ~ dnorm(0,1.0E-6)
prec <- 1/tau.sq
tau.sq <- tau*tau   # tau.sq = between-study variance
tau ~ dunif(0,10)
OR<-exp(d)          # exponentiate to get back to OR scale
}
```

4.3.1 Results, plots and comments about Model 3

```
Inference for Bugs model at "m3.txt", fit using jags,
 3 chains, each with 20000 iterations (first 2000 discarded)
 n.sims = 54000 iterations saved
         mu.vect sd.vect   2.5%    25%    50%    75%  97.5%  Rhat n.eff
OR         0.784   0.051  0.690  0.750  0.781  0.815  0.889 1.003  1100
d         -0.246   0.064 -0.371 -0.288 -0.247 -0.204 -0.117 1.003  1100
tau        0.132   0.080  0.012  0.070  0.124  0.182  0.306 1.016   240
deviance   7.695   4.497 -1.161  4.450  7.984 11.115 15.670 1.003   870

For each parameter, n.eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule, pD = var(deviance)/2)
pD = 10.1 and DIC = 17.8
DIC is an estimate of expected predictive error (lower deviance is better).
```

Y_i is the so called within-study variation while δ_i is the between-study variation. We got an estimate on d, the expected value of the between-study variation and an estimate on τ,variance of the distribution of δ_i. The point estimate of d is negative as well as the confidence interval with confidence level $\alpha = 5\%$ admits only negative values. Same for the OR value, the exponentiated version of d; both point estimate and confidence interval are under the 0. Both values carries the same information that is, the odds is greater in control group; in other words people who received the treatment are, statistically speaking, less likely to die. The deviance also is really small, meaning that the model fit well the data. It is not possible to compare this last model with the previous one because they are based on different data, but we can say that the DIC value is way small than the

previous cases.

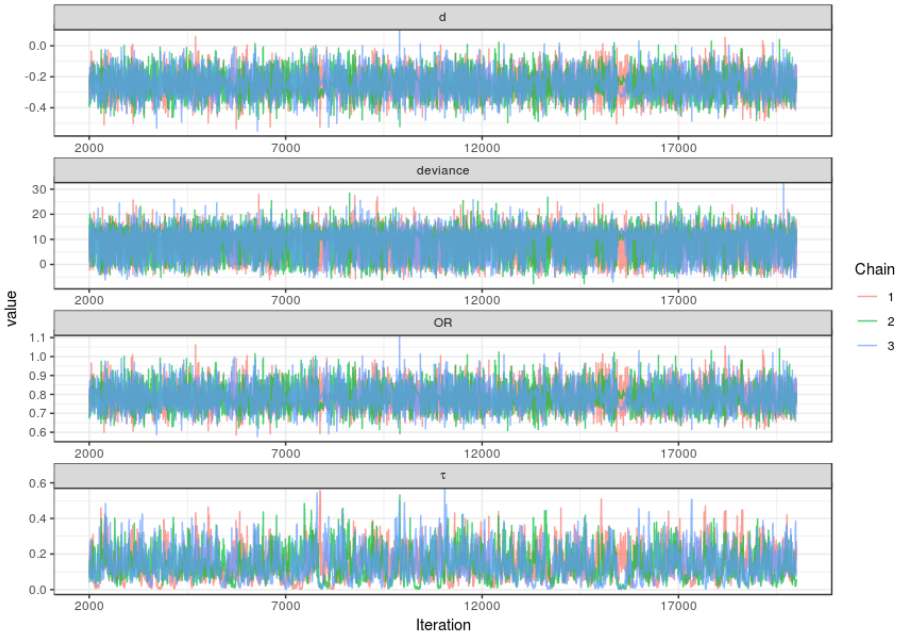

Figure 20: Traceplot of parameters of model 3

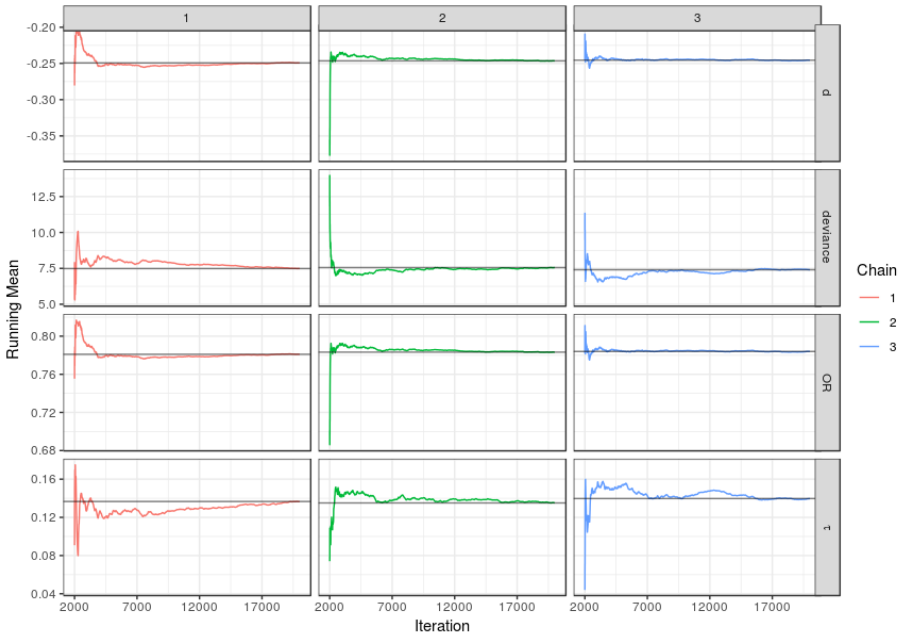

Figure 21: Running mean of parameters of model 3

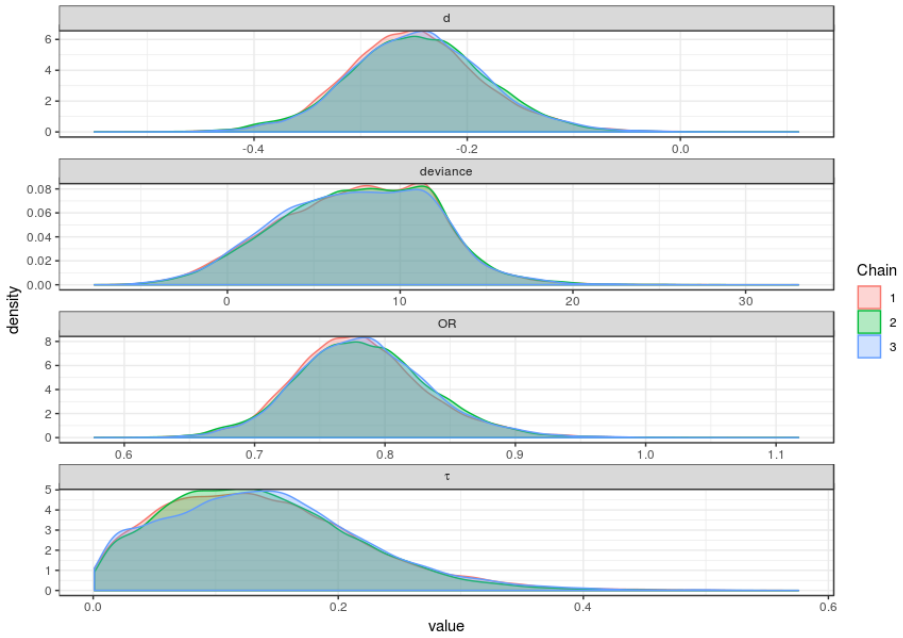

Figure 22: Posterior density of parameters of model 3

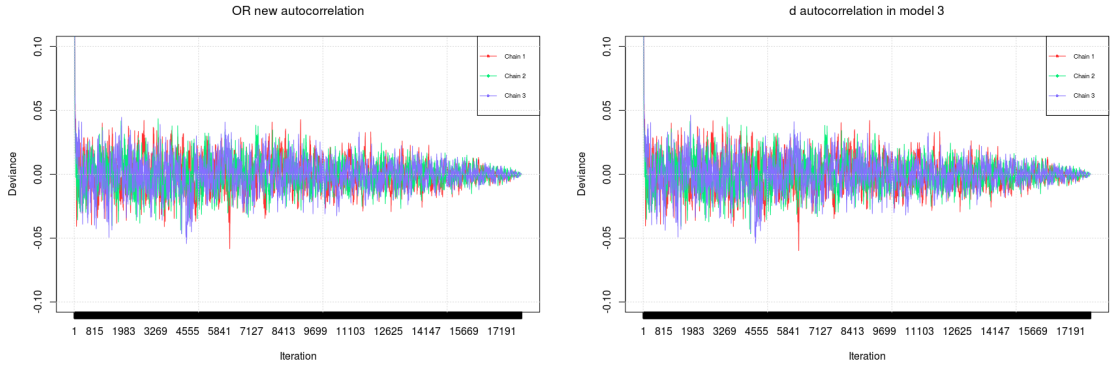

Figure 23: OR and between studies log odds autocorrelation plot

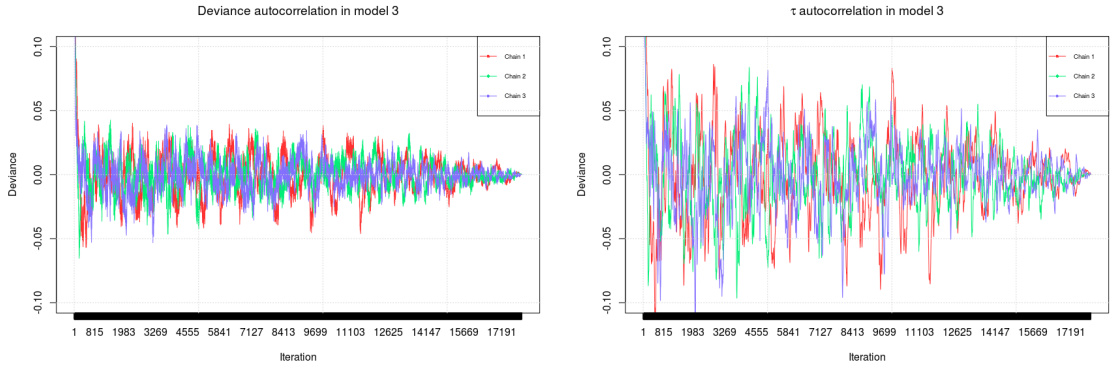

Figure 24: Deviance and τ autocorrelation of model 3

In Figures 24,23 autocorrelation values are not so small, this then affect the effective number of sample size way smaller than the original chain length. Notice that in 23 the autocorrelation is exactly the same, as well as N_{EFF} since OR is just the exponentiated version of d.

In order to check if the odds ratio obtained are statistically significant we can perform an hypothesis testing. Here we perform two kind of hypothesis testing for each contingency table related to each clinical trial: the Fisher exact test and the χ^2 test. The values obtained, reported in table 5, are all way smaller than 0.05; this makes all the odds ratio statistically significant.

Trial	Fisher results	χ^2 results
1	8.435e-19	3.114e-15
2	7.900e-58	2.267e-45
3	1.093e-39	5.021e-32
4	0.000e+00	0.000e+00
5	3.519e-173	3.626e-135
6	2.496e-27	7.115e-22
7	0.000e+00	0.000e+00
8	6.033e-246	8.603e-194
9	2.840e-131	5.115e-103
10	0.000e+00	0.000e+00
11	0.000e+00	1.472e-265
12	7.271e-104	8.280e-83
13	2.451e-158	1.394e-120
14	0.000e+00	0.000e+00
15	7.822e-60	4.827e-48
16	9.592e-85	1.076e-67
17	1.743e-71	9.115e-60
18	3.366e-78	7.642e-61
19	5.312e-81	5.145e-63
20	2.130e-87	8.614e-70
21	1.720e-184	1.649e-143
22	0.000e+00	4.775e-277

Table 5: Hyptothesis testing p-values

5 Comparison with frequentist inference

Maximum Likelihood estimation of parameter p in case of binomial data, is just the ratio between the number of success (death in our specific case) over the size of the sample (patients of each group).

$$y \sim Bin(n, p) \tag{43}$$

$$\hat{p}_{ML} = \frac{y}{n} \tag{44}$$

For that reason we can have an estimation of the probability of death of each trial for control and treatment case and compare it with the estimates coming from the 2 bayesian models.

Trial	Frequentist	Model 1	Model 2
1	0.0789	0.079	0.0806
2	0.0614	0.1006	0.101
3	0.0724	0.1058	0.1058
4	0.0665	0.0835	0.0835
5	0.0788	0.0827	0.0827
6	0.0678	0.097	0.0976
7	0.1037	0.1523	0.1528
8	0.0949	0.1085	0.1083
9	0.09	0.1239	0.1236
10	0.0720	0.0960	0.0958
11	0.0733	0.0900	0.0901
12	0.1711	0.1880	0.1881
13	0.0309	0.0470	0.0476
14	0.0664	0.0615	0.0602
15	0.1623	0.2049	0.2052
16	0.1594	0.1836	0.1842
17	0.1115	0.1194	0.1197
18	0.0529	0.0489	0.0489
19	0.0344	0.0309	0.0310
20	0.1531	0.1845	0.1847
21	0.0690	0.1057	0.1057
22	0.0323	0.0511	0.0513

Table 6: Control groups

Trial	Frequentist	Model 1	Model 2
1	0.0769	0.0629	0.0641
2	0.1206	0.0775	0.0773
3	0.1182	0.0829	0.0827
4	0.0835	0.0664	0.0661
5	0.0739	0.0685	0.0685
6	0.1153	0.0760	0.0763
7	0.1618	0.1123	0.1121
8	0.1019	0.0898	0.0892
9	0.1312	0.0965	0.0960
10	0.0978	0.0736	0.0736
11	0.0891	0.0720	0.0720
12	0.1766	0.1581	0.1578
13	0.0546	0.0358	0.0362
14	0.0509	0.0546	0.0558
15	0.2108	0.1654	0.1654
16	0.1784	0.1514	0.1515
17	0.0983	0.0991	0.0992
18	0.0389	0.0396	0.0396
19	0.0223	0.0248	0.0249
20	0.1834	0.1501	0.1501
21	0.1181	0.0794	0.0792
22	0.0578	0.0378	0.0379

Table 7: Treatment group

We can also perform a linear regression using the group (control vs treatment) as predictor and the probability of death/logit of the probability and check the coefficients registered for the two types of group.

```
lm(formula = c(logit_pt, logit_pc) ~ group)

Residuals:
     Min       1Q   Median       3Q      Max
-1.53356 -0.24676  0.00827  0.31976  0.92343

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.4655     0.1216 -20.280   <2e-16 ***
groupTreat    0.2225     0.1719   1.294    0.203
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5702 on 42 degrees of freedom
Multiple R-squared:  0.03834,Adjusted R-squared:  0.01545
F-statistic: 1.675 on 1 and 42 DF,  p-value: 0.2027 as response.

lm(formula = c(p_t, p_c) ~ group)

Residuals:
      Min        1Q    Median        3Q       Max
-0.084957 -0.026451 -0.009232  0.018845  0.103539

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.086531   0.009877   8.761 4.98e-11 ***
groupTreat  0.020814   0.013968   1.490    0.144
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04633 on 42 degrees of freedom
Multiple R-squared:  0.05022,Adjusted R-squared:  0.0276
F-statistic: 2.221 on 1 and 42 DF,  p-value: 0.1437
```

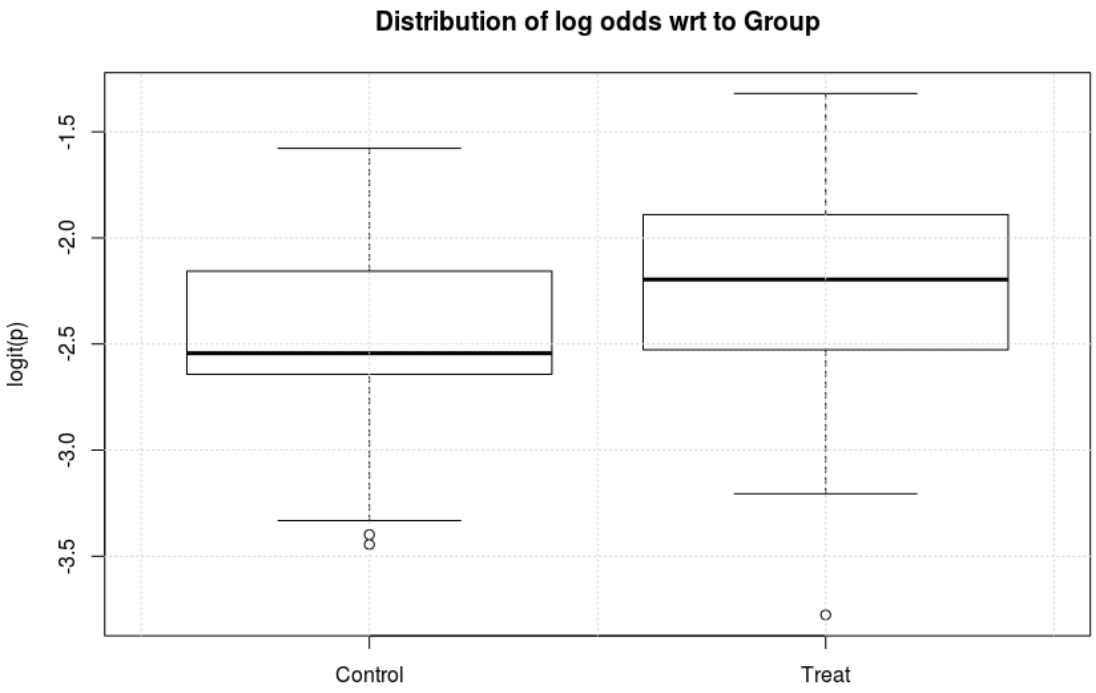

Figure 25: Boxplot for control and treatment group

Checking the coefficients of both the linear regressions it seems that the results obtained are just the opposite of those obtained in the bayesian inference. In the first linear regression the (Intercept) shows a value of -2.4655, this is related to the control group, while the coefficient of treatment group account for a 0.2225. This actually means that -2.4655 is point estimate for the logit in control case while the logit of treatment group has a point estimate that differs 0.2225 from the logit of control group. $p_{control} = logit^{-1}(-2.4655) = 0.0783$ while $p_{treatment} = logit^{-1}(-2.2432) = 0.0959$. Same happens in linear regression that use the probabilities as response: point estimate for $p_{control} = 0.0865$ and for $p_{control} = 0.107345$. This is an opposite trend with respect to the bayesian inference. It seems that the treatment group accounts a greater probability of death (around 10%) with respect to the control group(around 8%). However it must be highlighted that according to both the lm output, the coefficients for treatment group appears to not be statistically significant having a p-value greater than the 0.05.

6 Conclusions

In conclusion we can say that, from a Bayesian point of view, the difference between control and treatment group is statistically evident (see the results of the first two models showing a δ random effect of around -0.25 in logit of death probability for treatment group). Also model 3 shows evidence about the effectiveness of the beta-blocker treatment with an odds ratio of $\frac{1}{0.784} = 1.275$ in favour of control group for what concerns the *death* event. Interesting are the results of frequentist inference that, based only on the collected data, shows a greater chance to die in treatment group rather than the control one.

References

[1] S. Yusuf, R. Peto, J. Lewis, R. Collins, P. Sleight, Beta blockade during and after myocardial infarction: an overview of the randomized trials, Progress in cardiovascular diseases 27 (5) (1985) 335–371.

[2] O. Website, Openbugs.
URL `http://www.openbugs.net/Examples/Blockers.html`

[3] C. DiMaggio, Bayesian analysis for epidemiologists part iv: Meta-analysis.
URL `http://www.columbia.edu/~cjd11/charles_dimaggio/DIRE/styled-4/styled-11/code-9/`

[4] S. Manual, Definition of effective sample size.
URL `https://mc-stan.org/docs/2_18/reference-manual/effective-sample-size-section.html`

[5] A. Gelman, D. B. Rubin, et al., Inference from iterative simulation using multiple sequences, Statistical science 7 (4) (1992) 457–472.

[6] Wikipedia, Deviance(statitics).
URL `https://en.wikipedia.org/wiki/Deviance_(statistics)`

[7] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, Bayesian data analysis. texts in statistical science series (2004).

[8] Wikipedia, Deviance information criterion.
URL `https://en.wikipedia.org/wiki/Deviance_information_criterion`