

Coursera__Bayesian

Michael Belias

June 3, 2017

For the peer-graded assignment we will use a multi-level model, such as a meta-analysis. We gathered 10 studies in which patients were treated with intravenous immunoglobulin to prevent infections held in hospital.

Understand the problem

We are trying to answer the question ‘does the administration of intravenous immunoglobulin (IVIG) prevent infections compared to placebo?’.

Plan and properly collect relevant data

Justify why the data chosen provide insight to answering your question Describe how the data were collected
In order to answer the abovementioned question we conducted a systematic review studies that estimated the proportion of patients that had a sepsis event with and without IVIG treatment.

Describe any challenges relating to data acquisition/preparation (such as missing values, errors, etc.)

A meta-analysis is a powerful statistical tool because it has more power than a single trial and helps us check for other study-level characteristics that may influence our results. The challenge lies in the fact that these results are driven from a multi-level model, meaning that participants may be within study correlated and not completely independent.

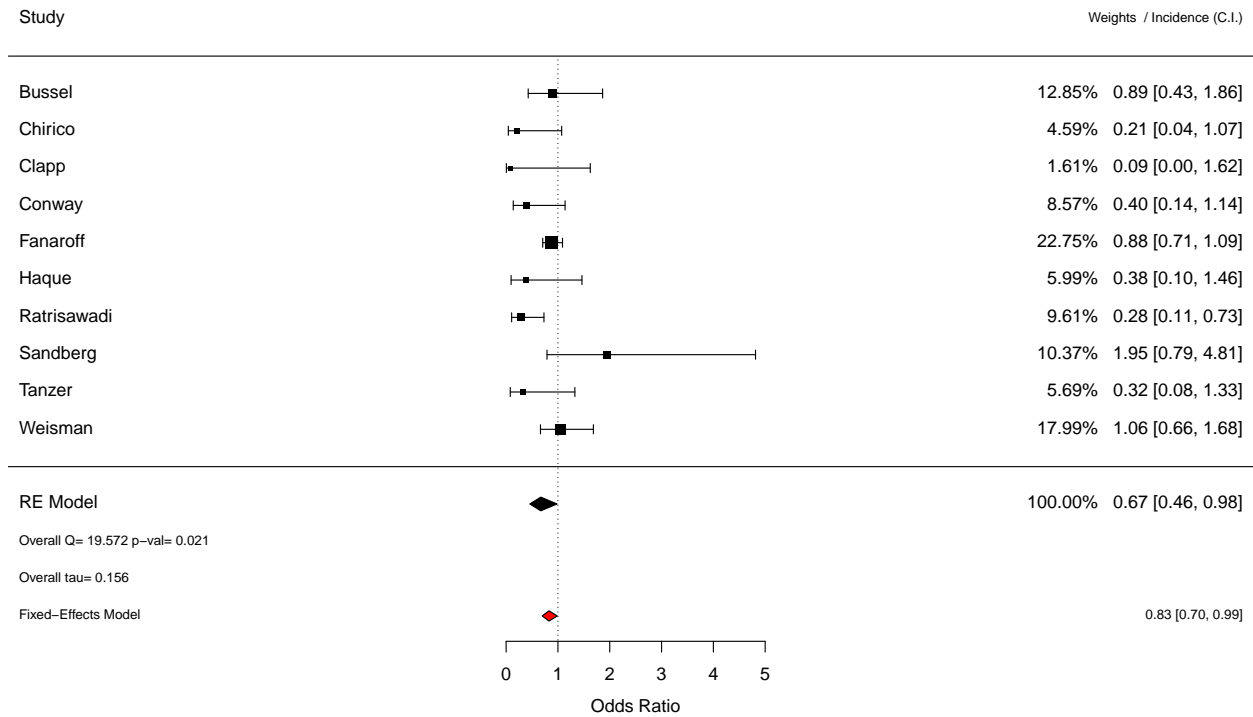
Explore data

Graphically explore the data using plots that could potentially reveal insight relating to your question

The data consist of 5 columns indicating the ‘Country’ that the study was conducted, two columns represent the number of patients allocated in Treatment or Control and two the sepsis events that occurred in each arm.

	Country	Treat.Events	Treat.Total	Control.Events	Control.Total
Bussel	USA	20	61	23	65
Chirico	Italy	2	43	8	43
Clapp	USA	0	56	5	59
Conway	UK	8	34	14	32
Fanaroff	USA	186	1204	209	1212
Haque	Saudi Arabia	4	100	5	50
Ratrisawadi	Thailand	10	68	13	34
Sandberg	Sweden	19	40	13	41
Tanzer	Turkey	3	40	8	40
Weisman	USA	40	372	39	381

integer(0)



Forest plots are a useful tool for investigating the between studies heterogeneity. In our case there is an increased level indicated by the Q and its corresponding p-value. Also is we fit a random and a fixed effects model we see great difference, probably due to the Sandberg trial.

Postulate a model

Justify why the model you chose is appropriate for the type of data you have

We wish to fit both types of meta-analysis models using odds ratio as effect size. So the distributions of the variables are:

- $y_{sT} \sim \text{Bin}(n_{sT}, p_{sT})$ the per study s number of events on treatment arm
 - n_{sT} number of patients per study on treatment group
 - p_{sT} number of events per study on treatment group
- $y_{sC} \sim \text{Bin}(n_{sC}, p_{sC})$ the per study s number of events on control arm
 - n_{sC} number of patients per study on control group
 - p_{sC} number of events per study on control group

We will try to estimate the difference in the effects between the two arms. Odds are calculated in log-scale using the logit transformation so we will have a base effect in the control group plus a extra effect on the treatment.

Describe how the model is well suited to answer your question

we will assess the intervention effect by two methods a fixed and a random-effects approach. This way we will have the pooled estimate of the intervention effect.

Write the full hierarchical specification of the model

For the Fixed-effects model we are setting the following: * $\text{logit}(p_{sC}) = a_s$ * $\text{logit}(p_{sT}) = a_s + \delta$, where a_s is the study-specific baselines

For the random-effects model we are setting the following: * $\text{logit}(p_{sC}) = a_s$ * $\text{logit}(p_{sT}) = a_s + \mu_s$, where a_s is the study-specific baselines + $\mu_s \sim N(\delta, \tau^2)$, instead of a common effect used before + δ is the overall log odds ratio treatment effect + τ^2 the precision

Justify your choice of prior distributions

Odds ratios are approximatly normally distributed using the logit transformation, so this is the appropriate simulation techniqe. In the fixed effect model we have a common effect so there a common parameter to be estimated, whereas in the random-effect a common assumption is that the trial estimates are driven from a normal distributio with a mean μ and σ^2 SD (JAGS uses precision $\tau = 1/\sigma^2$).

Fit the model

First, we will try the fixed - effects approach described above.

```

Compiling model graph
  Resolving undeclared variables
  Allocating nodes
Graph information:
  Observed stochastic nodes: 20
  Unobserved stochastic nodes: 12
  Total graph size: 122

```

Initializing model

```

Iterations = 2001:12000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

```

1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
delta[2]	-0.2066	0.08892	0.0008892	0.001889
delta.or	0.8166	0.07259	0.0007259	0.001544

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
delta[2]	-0.3834	-0.2660	-0.2067	-0.1448	-0.03341
delta.or	0.6816	0.7664	0.8133	0.8652	0.96714

Afterwards we will try the random - effects approach described above.

```

Compiling model graph
  Resolving undeclared variables
  Allocating nodes
Graph information:
  Observed stochastic nodes: 20
  Unobserved stochastic nodes: 22
  Total graph size: 141

```

Initializing model

```

Iterations = 2001:52000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 50000

```

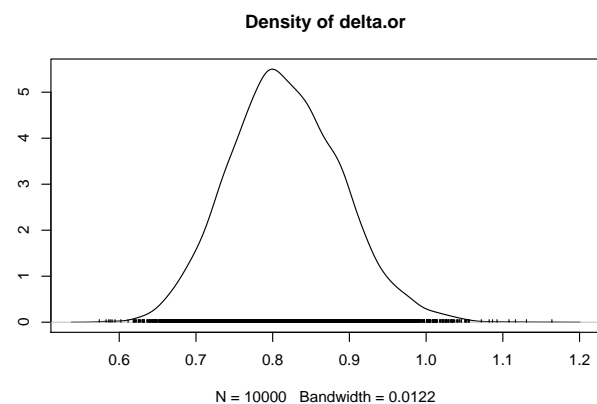
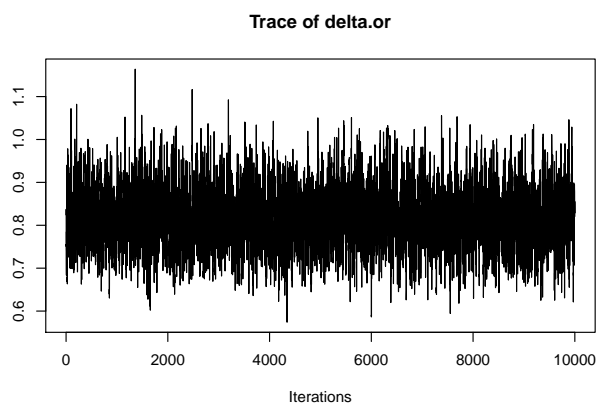
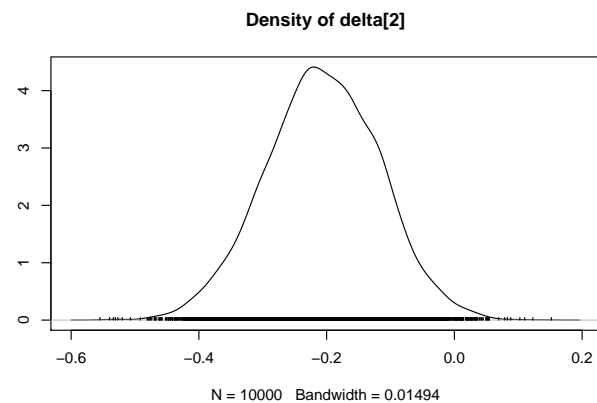
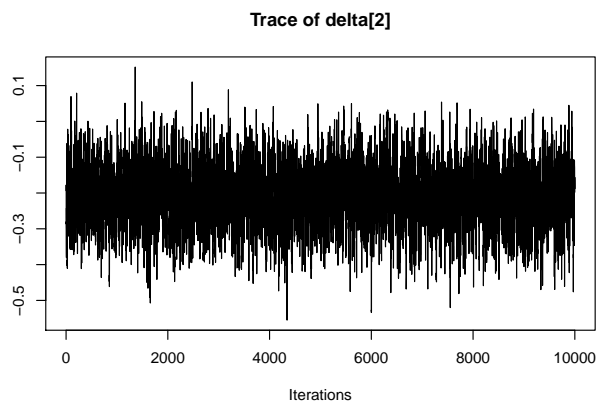
1. Empirical mean and standard deviation for each variable,
plus standard error of the mean:

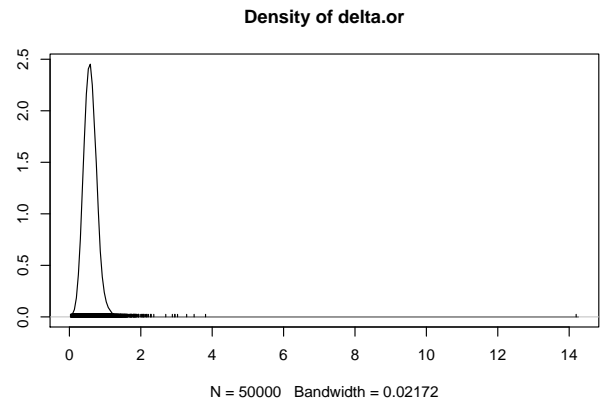
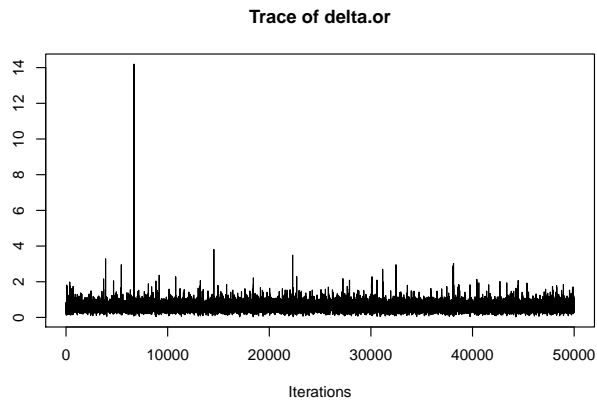
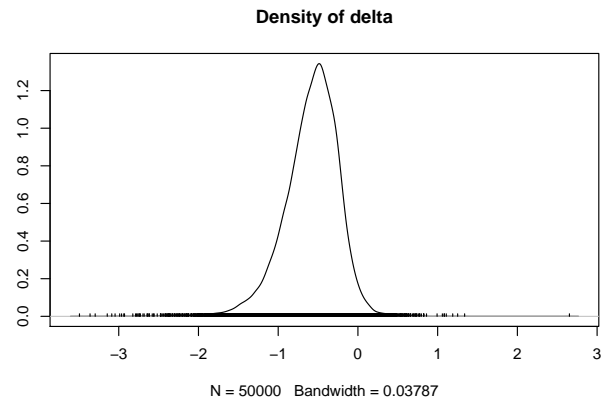
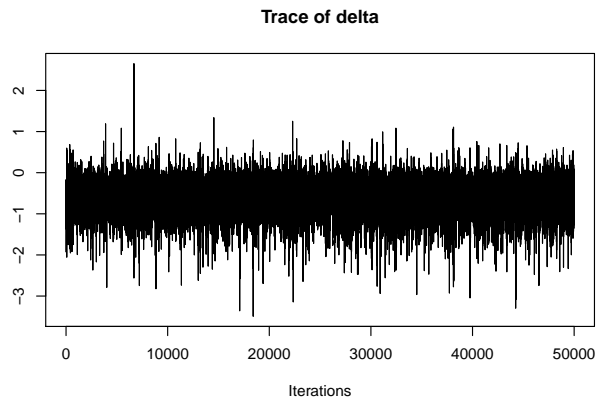
	Mean	SD	Naive SE	Time-series SE
delta	-0.5825	0.3432	0.0015350	0.003872
delta.or	0.5902	0.2025	0.0009055	0.001959

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
delta	-1.3529	-0.7715	-0.5471	-0.3547	-0.01183
delta.or	0.2585	0.4623	0.5786	0.7014	0.98824

Check the model





The fixed effects 95% posterior CI for the log odds ratio, delta, is $(-0.38, -0.033)$, random effects 95% posterior CI for the log odds ratio, delta, is $(-1.35, -0.01)$, suggesting that IVIG is reducing the chances of in-hospital infection. The difference in the results lies due to the lack of shrinkage in the fixed effects model e.g. the Sandberg (2000) study is not shrunk down towards the mean.

Decide if your model is adequate. Postulate and fit at least one alternative model and assess which is best for answering your question. If neither is adequate, report that and move on (I have already fitted a second model)

The Random effects model is performing better by shrinking the estimates to the mean.