Portfolio
Assignment 2

Sabrina Zaki Hansen (au693815)

GitHub Repository: https://github.com/sabszh/Assignment3

Assignment Description

In this following report, a Bayesian meta-analysis on acoustic traits in patients with schizophrenia has been conducted and reporte. The report will start with a simulation and model building, to make clear inferences on data given by the meta-analysis performed by Parola et al (2020).

Q1 - Data Simulation

Aim of the following section is to report on a pipeline and findings of a simulation of study effects on acoustic markers of schizophrenia, to gain insight of the field.

One data-set of 100 studies was simulated. Following plot shows the distribution of effect of the mean with publication bias included. What this means, is that we in the simulation accounted and extracted for publication bias, and p-hacking.

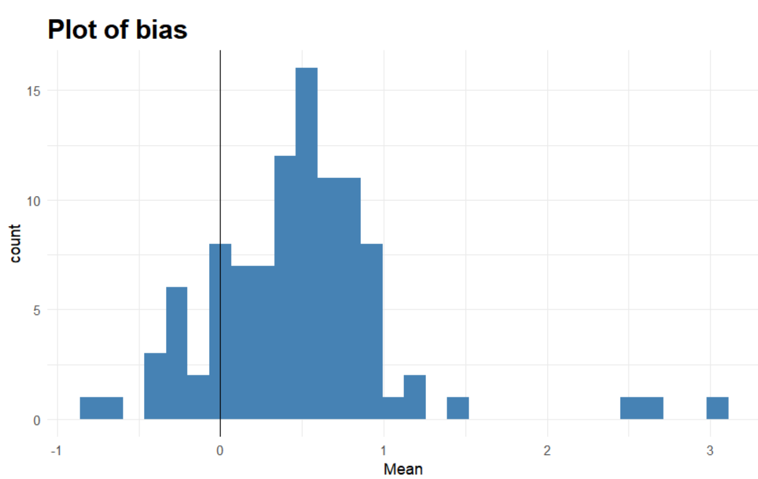

Figure 1

The criteria threshold set in this simulation for a study being published is that if the absolute value of the difference between the effect of the mean and two times the standard error is greater than zero, then study has 90% chance of being published. However, if the absolute value of the difference between the effect of the mean and two times the standard error is less than or equal to zero, then the study has a 10% chance of being published. The following plot shows how the effect of the study is spread:

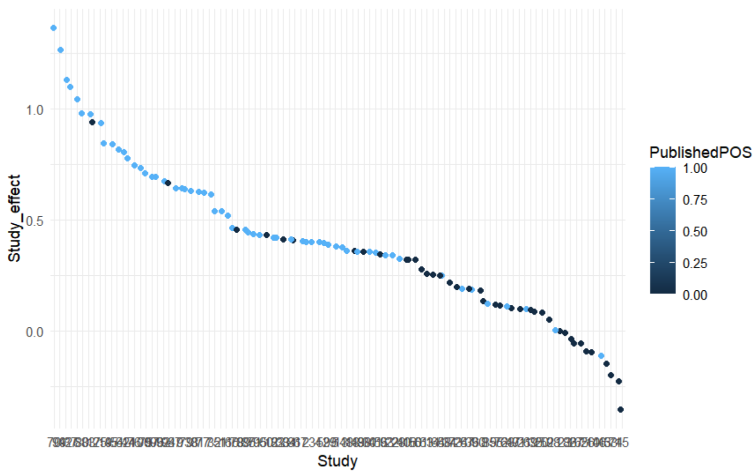

Figure 2

Through visual inspection it is detected that the simulation catches the desired outcome effect of most published studies being with higher effect, although some are not published and some with lower are-this matches the empirical understanding of the publication bias. The theoretical understanding of it, is that publication bias is a type of bias that can occur when studies with positive results are more likely to be published than studies with negative results.

Next step in the Bayesian workflow, is to build a formula that will be used to explain the data. Given the literature on the matter, we end with the following formula:

$$Study = Study\ effect|se(Standard\ error) \sim 1 + (1|Study) \tag{1}$$

What the formula (1) tells us, is that the predictor of the model is the effect size as a distribution, predicted by the standard error, and each study allowing for random intercepts. We are interested in predicting a distribution for this specific task, as we want to understand the variability of the data, whereas a point estimate would only represent a single value.

For our simulation we want to set a weakly informed prior, as we want to allow the data to have more influence on our posterior distribution.

```
# Setting weakly informed prior
Study_p <- c(
  prior(normal(0,0.3),class=Intercept),
  prior(normal(0,0.2),class=sd))
```

The mean for the intercept is 0, as a way of indicating that the prior is not very informative about the likely value of the intercept. The standard deviation are set as for what we usually would expect as effect sizes in psychological studies, and iterative process of adjusting.

After setting the priors, 'Prior Predictive Checks' (PP-check) were performed.

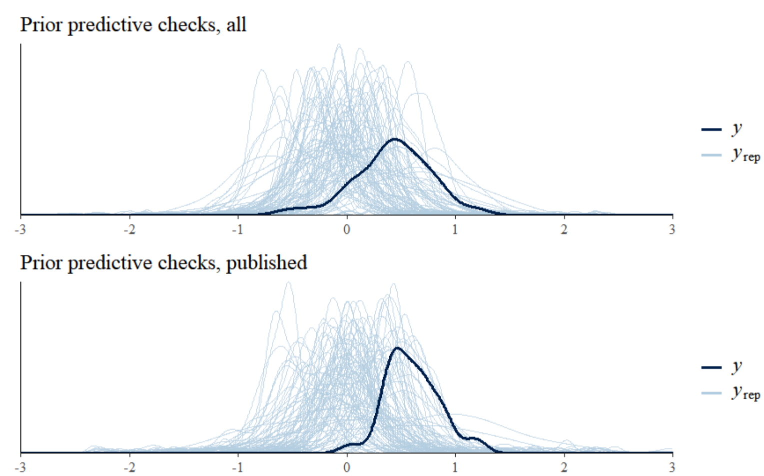

Figure 3

The plots in figure 3 is for the data generated in the simulation: the first is for all of the studies (published and not) whereas the last one is for only the published studies. The pp-check seems to generate data that is similar to our simulated data, with some slight variations - which makes sense given the priors are weakly informed. The set prior are accepted and thus move on to look at how the posterior generated.

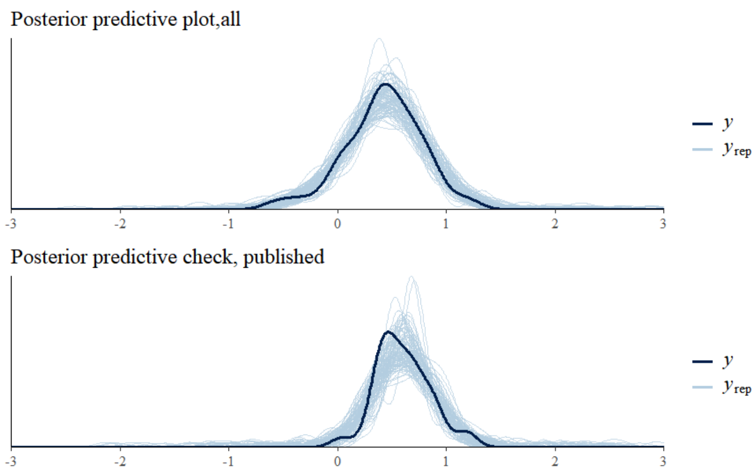

Figure 4

The plot tell us that our model is a good fit, as we can see by looking at the posterior predictive distribution of the response variable, our observed values of y, are within the range and variability of the predicted values.

Before we accept this model of explanation of our simulated data, we move on to investigate how to the posterior update when observing the data. By drawing from the fit, we get the following plot for the whole simulation and the published:

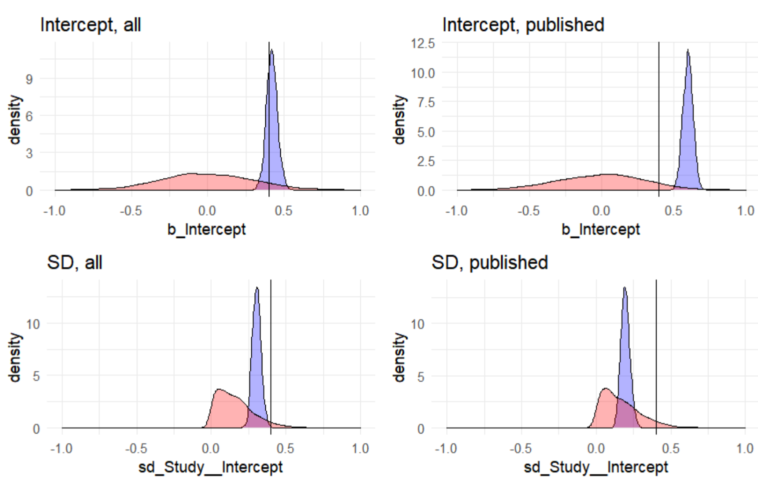

Figure 5

The vertical black lines indicate the theoretical intercept and standard deviation set in our model. Through inspection it is clear that the described model is a good fit, as the weakly informed allowing for our data to work hard to influence on the posterior distribution. With the different plot pointing in the desired direction, the model is accepted as a good fit and will be used when introducing collected data in the next question.

Q2 - Current Evidence

Given that there now has been build a theoretical foundation, an empirical approach will now be introduced. Here we are interested in understanding the current evidence for distinctive vocal patterns in schizophrenia. To do this, data from a meta analytical study will be analyzed. The following is a summary of highlights of the data, containing studies on patients with schizophrenia (sz) and healthy controls (hc).

Studies	Mean n sz	Mean n hc	n pitch	Sd pitch mean sz	Sd pitch sz	n pitch	Sd pitch mean hc	Sd pitch sd hc
57	30.88	27.05	20	11.25	6.11	15	17.36	7.91

Figure 6

To get started, two vectors: the effect sizes and sampling variances where calculated by using the values specified in figure 6 for the measure of the standardized mean difference. The effect size, or Cohen's d including pooled SD uses the following formula:

$$d = \frac{\bar{x_1} - \bar{x_1}}{\sqrt{\frac{s_1^3 + s_2^3}{2}}} \tag{2}$$

And sampling variance:

$$SV = \frac{\sum_{i=1}^{n}(x_1 - \mu)^2}{n - 1} \tag{3}$$

By using the two specified variables, a Random-Effects Model was included to produce a funnel plot, to describe the relationship between the effect size and standard error.

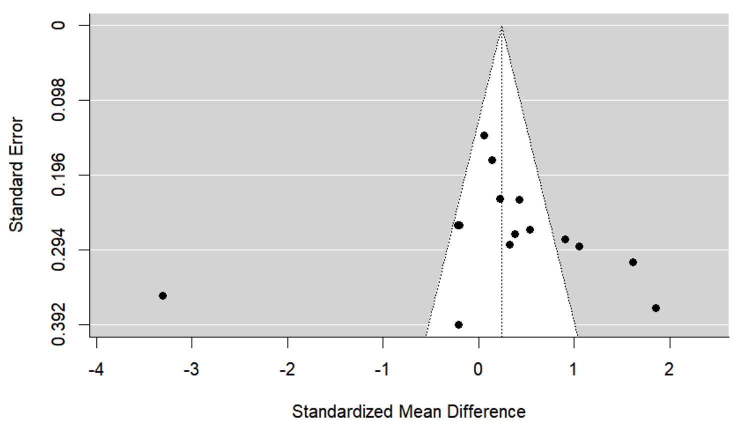

Figure 7

The funnel plot is used to visually check for existence for publication bias. In the plot we see the effect size (ES) from each study against the precision of the estimate. If there is a publication bias, the plot will be asymmetrical-which seems to be the case here. However, this is not enough to conclude anything for sure about the publication bias. The Bayesian workflow will then be used to investigate prior and posterior of the data.

By adjusting the prior, with the same model as for in the simulation, the prior predictive check looks as the following:

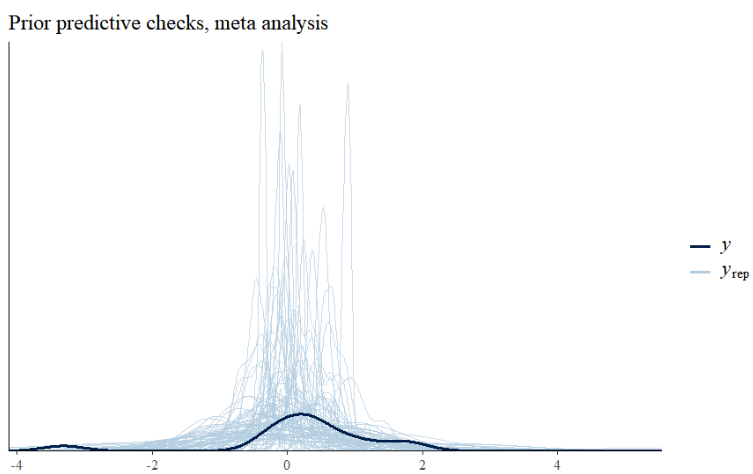

Figure 8

Looking at the posterior predictive check plot:

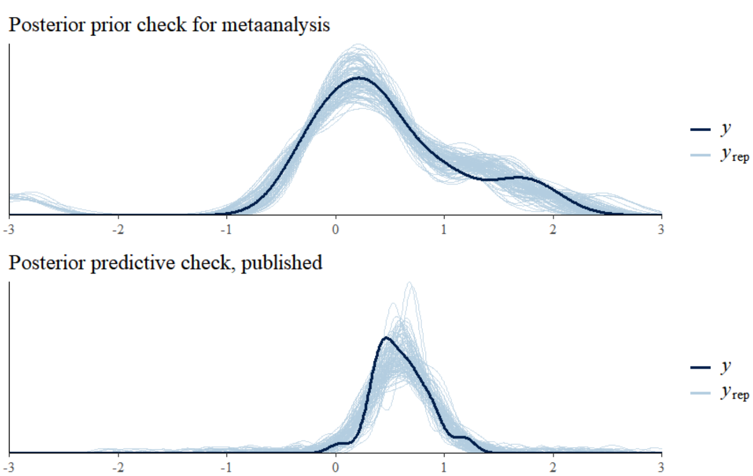

Figure 9

In the plot (figure 9) the first plot shows the PP-check for the meta analysis, and the last is for the simulated data that is deemed published. The reason they are side-by-side is that it gives an opportunity to compare the distribution as the simulation is showing the tendencies in a positive publication bias. They seem to be a bit similar, in the way they vary - but as this is only a visual inspection further investigation would have to contribute to a conclusion about the publication bias. We look at how the distribution update, when including the simulated data

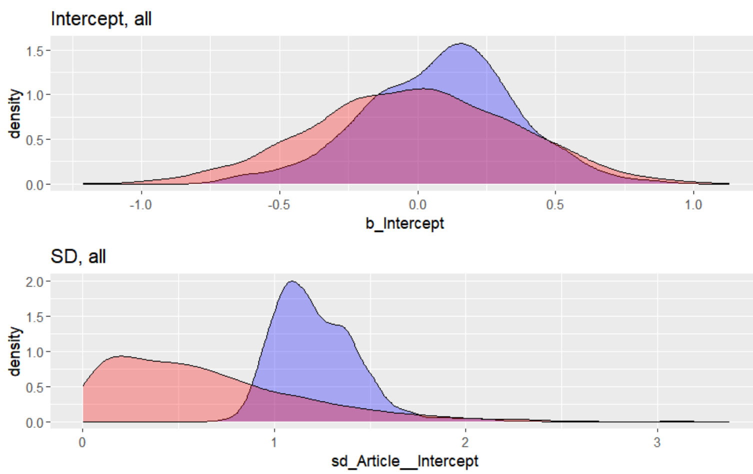

Figure 10

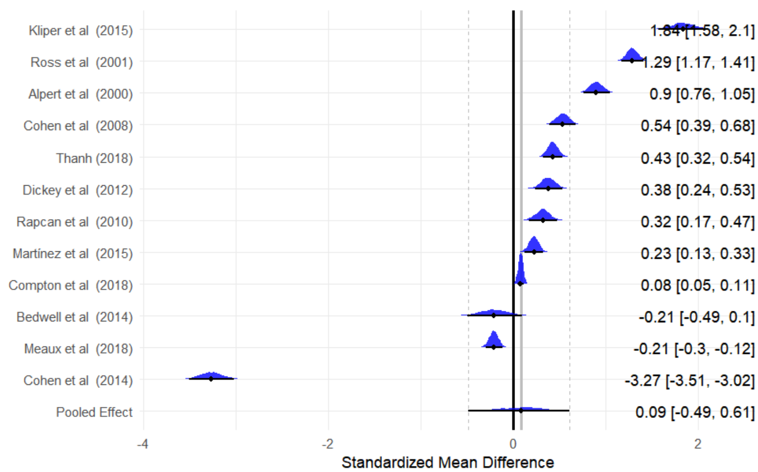

Figure 11

The forest plot above, gives us a graphical representation of the results of the meta analysis. The plot shows the standardized mean difference. What we see is that it is heavily right-skewed thus indicating a publication bias as foreseen.

Conclusion

Through inspection of the simulation and data from the meta analysis, it can be concluded that there is a publication bias within this sample of the field of study of pitch of patients with schizophrenia. However, the sample size is only 15 which might not be a good representation of the field as a whole. Further investigation would be needed with a larger sample size to confirm these results.

Q3 - Prediction

Part 3 has not been performed. But is planned to be executed with a LOO.