Victoria Engberg Lowe, Vlada Caraman &
Natasha Becker Bertelsen

Methods assignment 2: Meta-analysis

Q1: The simulation

1.1 The goals of the simulation

In this assignment, we investigated a meta-analysis of pitch in schizophrenia. In order to better understand the problem we are studying before fitting a model to the empirical data, we simulated similar data, in this case a dataset of 100 studies.

1.2 The simulation process

First of all, we defined our parameters, which are the mean effect size of the studies (0.4), the mean variation by study (0.4) and the measurement error (0.8). After this we created a dataframe that contains studies with the mean number of participants (20) and the average deviation (10). Also, we specified that every study included in this dataset should have at least 10 participants, otherwise it is irrelevant for our potential meta-analysis. After we ran the simulation for all studies, we also simulated a publication bias. This way we were able to contrast between all studies that we simulated and only the studies that were published.

The simulated publication bias favors results that are both significant and positive, i.e., statistically significant results showing that schizophrenic patients have a higher pitch compared to healthy controls. The bias was simulated such that any study which lived up to these criteria had a 90% probability of getting published, and studies which did not live up to the criteria had a 10% probability of getting published. Of the 100 studies we had simulated, with this publication bias only 49 studies were published.

1.3 Build a proper Bayesian model to analyze the simulated data.

To analyze the simulated data, we follow a Bayesian workflow and define the following model formula:

```
Pitch_f <- bf(TrueStudyEffect|se(ObservedSigma) ~ 1 + (1|Study))
```

In our model, we specify that the outcome of interest is the TrueStudyEffect and the ObservedSigma. We specify both parameters since we are interested in predicting the distribution. We inform the model that the outcome variables are predicted by intercept. Lastly, we include random intercepts for studies and therefore each study has its own intercept. We specify the following priors for our model:

```
Pitch_p <- c(
  prior(normal(0,0.3), class = Intercept),
  prior(normal(0,0.3), class = sd)
)
```

We specify a prior for the intercept and a prior for the standard deviation and choose a mean of 0 and a standard deviation of 0.03 for both parameters. The reason for choosing a mean value of 0 is to ensure that we do not influence the model too much in one direction but allow the model to find a potential difference between the studies itself. In order to get outcome values that are within an appropriate range, we choose a standard deviation of 0.03. Ideally, we would like outcome values within the range of -1 to 1 based on knowledge about the topic from already existing studies in the field. We run a prior-predictive check to inspect the choice of prior values in our model and see how they fit to the simulated data for all studies and for published studies respectively.

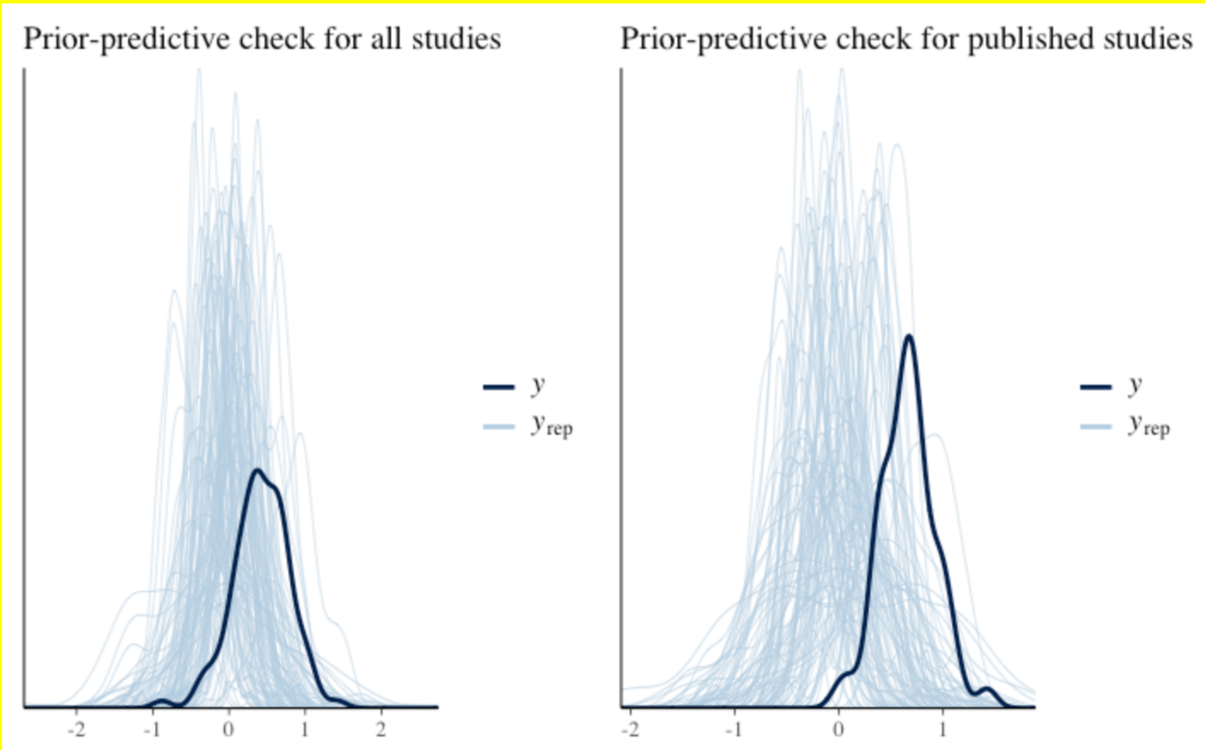

Figure 1: Prior-predictive checks

The prior-predictive checks show that for the unbiased-data, we see a range of roughly -2 to 2 and for the biased-data that range narrows down to -2 to 1. Here we also see that the distributions are centered at different values, with the biased-data generating a distribution which is positively skewed compared to the distribution created by the unbiased-data. Subsequently, we run a posterior predictive check to see how the models learn from the data.

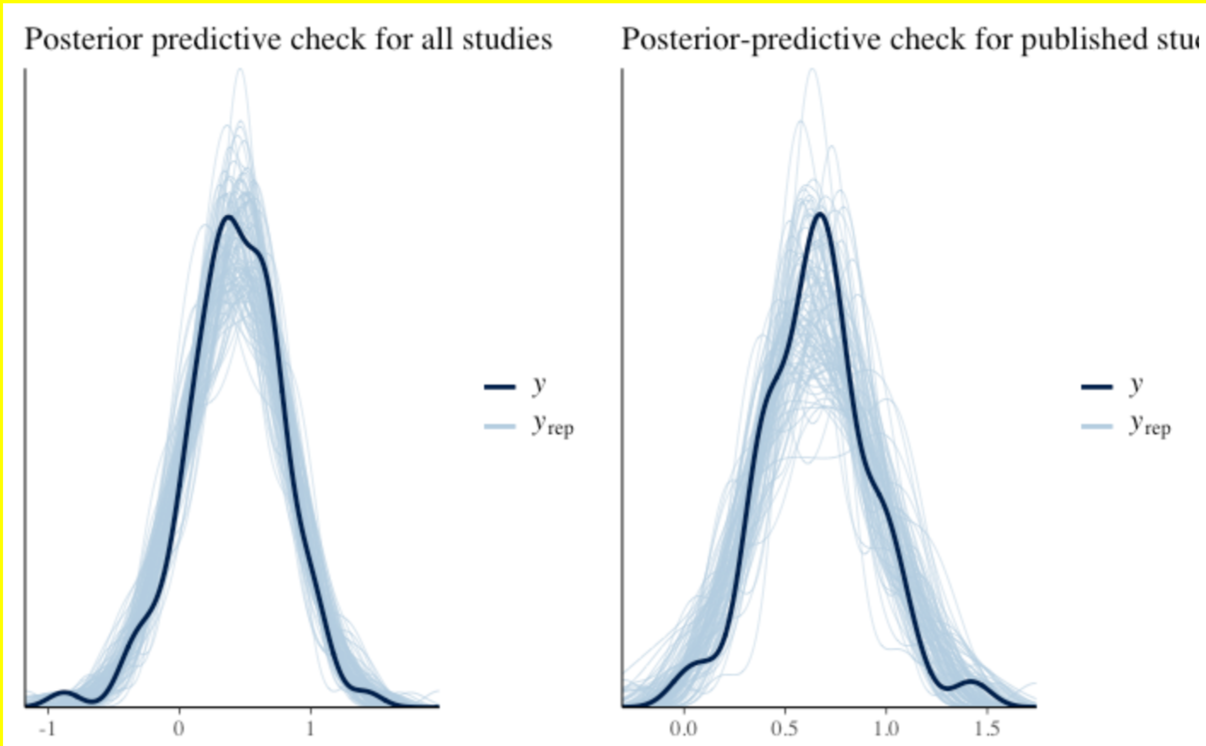

Figure 2: Posterior predictive checks

The posterior predictive checks show that the posterior distributions are now much closer to the simulated data, and both distributions are more confident (i.e., pointy) and more clearly resemble the distribution. We observe that the posterior predictive check generated by the biased-data is noisier than the one generated by the unbiased-data. This is most likely due to the biased-dataset being much smaller than the un-biased one, and thus the biased-dataset is more uncertain/noisy. Additionally, we also observe a difference in range, again with the distribution generated by the unbiased-data being broader, here from around -1 to 1.5, and the distributions generated by the biased-data ranging from 0 to 1.5.

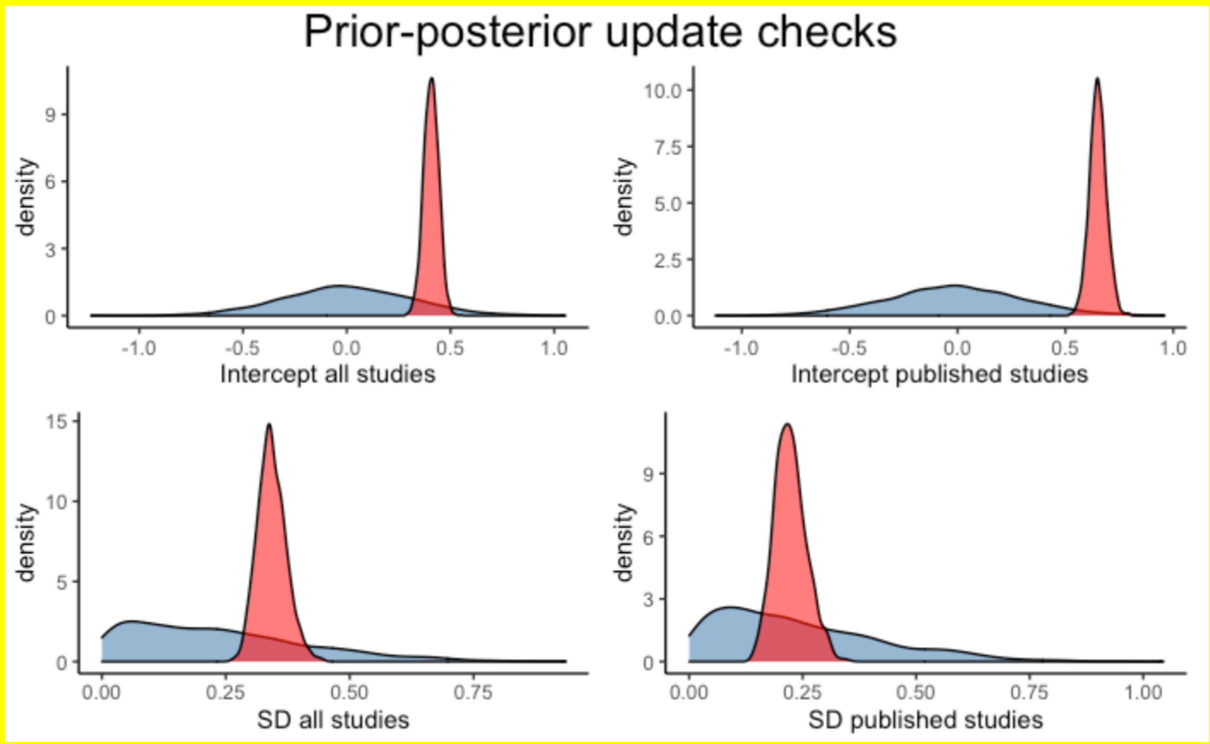

Figure 3: Prior-posterior update checks

From the prior-posterior update checks, the blue distributions are our priors and the red distributions are our posteriors. We did two prior-posterior update checks for both the unbiased-data (all studies in the graph) and the biased-data (published studies in the graph), one for the intercept and one for the SD. Across all prior-posterior update checks we can see that the priors are all relatively flat and thereby not very informative. Additionally, we see quite confident posterior distributions across conditions as well, meaning that they have learned from the data.

The prior-posterior update check for the intercept generated by unbiased-data shows that the posterior is slightly skewed to the right, and this in combination with the flat priors tells us that we should expect more positive outcome values.

The prior-posterior update check for the intercept generated by biased-data shows a confident posterior distribution which is even more skewed to the right and pushing towards the border of the prior.

The prior-posterior update check for the SD generated by unbiased-data shows a very confident posterior distribution, which is situated around the middle of the prior distribution, indicating that it has learnt from the data.

The prior-posterior update check for the SD generated by biased-data shows a slightly less confident posterior distribution, which is situated more to the left than the one generated by the unbiased data, indicating that there is less variance in the biased-data. This is consistent with the biased-data not including the negative studies.

1.4 What we have learned from the simulation

To sum up, we ran the prior predictive checks, posterior predictive checks and prior-posterior update checks with both the unbiased-data and the publication-bias-data, to investigate how the publication bias would impact our findings. One of the key findings from the comparison between the prior- and posterior predictive checks was that the range of outcomes changes depending on which dataset is used, which is expected since virtually all negative outcomes have been eliminated by the publication bias. We learned that the publication-bias-dataset overall generated distributions which suggested higher values and smaller deviations compared to the distributions generated by the unbiased-data.

Q2: The empirical data

2.1 The empirical data and how it compares to the simulated data

The empirical data from the meta-analysis consists of a sample containing 46 articles (55 studies). The meta-analysis has a total of 1254 patients with schizophrenia, 788 of those were male and 466 female. For the control condition, there are 699 patients, 376 of those were male and 323 female. These numbers include only the relevant and not repeated participants from all the studies.

To be able to use our Bayesian model on the empirical data, we applied the *escalc()* function from the *metafor* package to calculate the study effects and their corresponding standard deviations. We did a posterior-predictive check on the empirical data to see how well our model fitted the data.

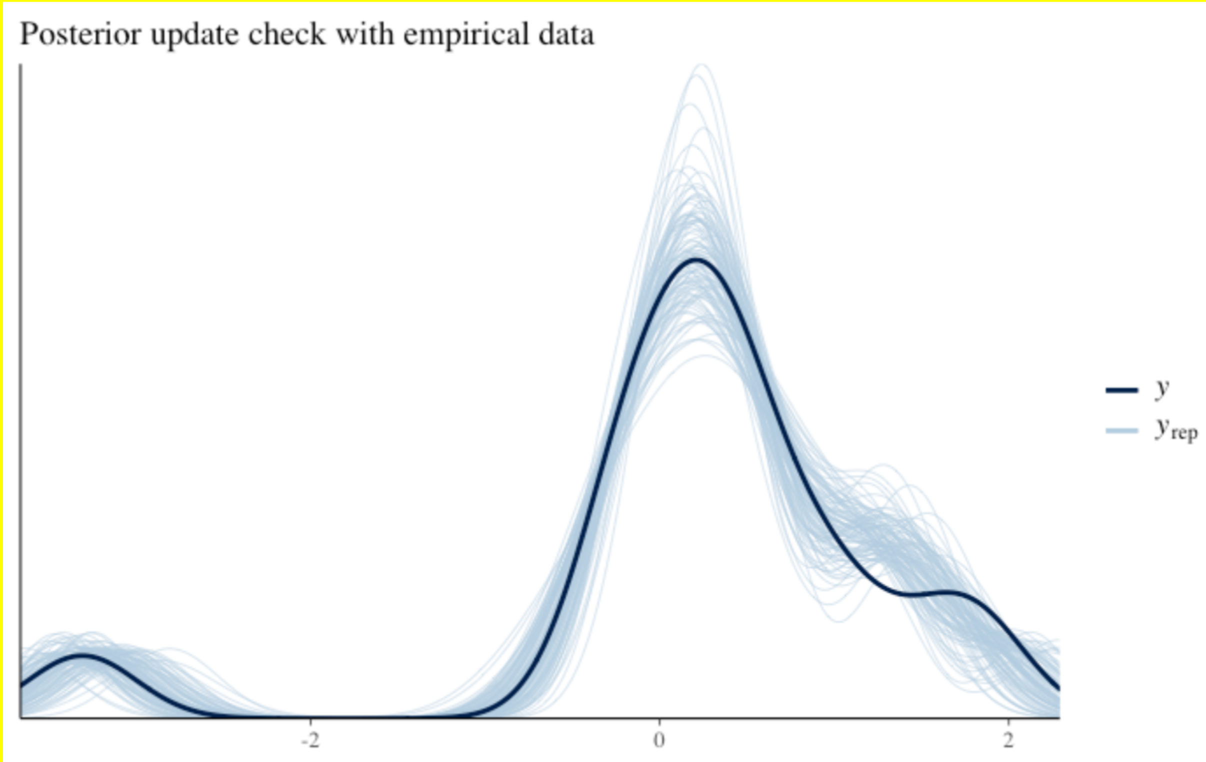

Figure 4: Posterior update check with empirical data

The posterior update check shows that the model fits the empirical data quite well with the exception of values around 2 where there is more noise and the model deviates more from the empirical data. The population level effect is 0.11. This suggests that the schizophrenic group tends to have higher pitch compared to the control group.

However from part 1 of the assignment we know that a publication bias can skew the estimated population effect. We know that there generally is a publication bias, which favors positive and significant results, this means that non-significant and negatively significant results are left underreported(i.e. the file drawer effect). With this in mind it makes sense to be cautious about the calculated population effect from a meta-analysis, as we can assume that there are unpublished negative/non-significant results “missing”, which would have changed the outcome of the meta-analysis.