Alina Maria Nechita
Student number: 202104831

Assignment 2

Meta-analysis

Question 1:

Simulate data to setup the analysis and gain insight on the structure of the problem.

Simulate:

- *One data set of 100 studies with a mean effect size of 0.4, average deviation by study of 0.4 and measurement error of 0.8.*
- *N of participants should follow a normal distribution with mean of 20, SD of 10, but no fewer than 10 participants).*

The data should consist of:

a) one row per study, with an effect size mean and standard error.

Then:

b) Build a proper Bayesian model to analyze the simulated data.

c) Then simulate publication bias (only some of the studies you simulate are likely to be published, which?), the effect of publication bias on your estimates (re-run the model on published studies, assess the difference),

d) Discuss what this implies for your model.

e) Use at least one plot to visualize your results.

I have simulated the data using the requirements. I've plotted data in accordance with the publishing bias (the output produced when the code was run for the first time).

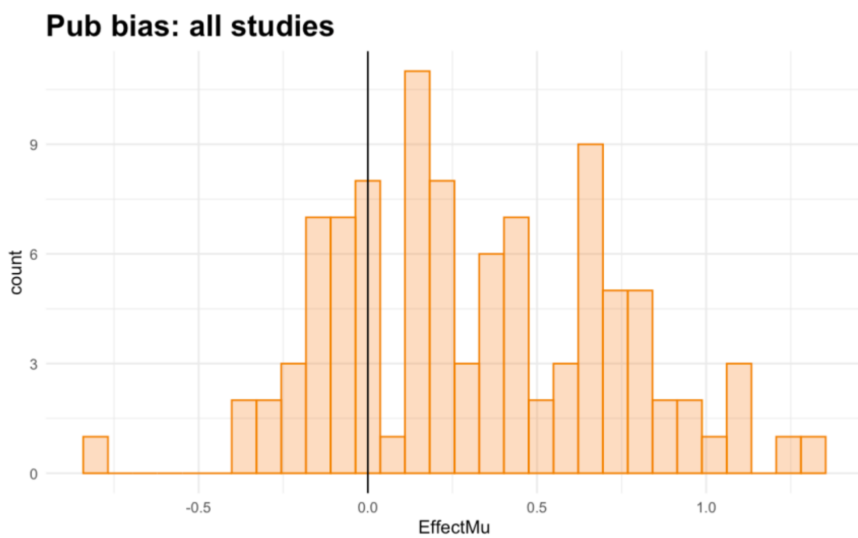


Figure 1: Plot of simulated data: all studies. Includes effect size mean of 0, more studies have positive effect size rather than negative.

Alina Maria Nechita
Student number: 202104831

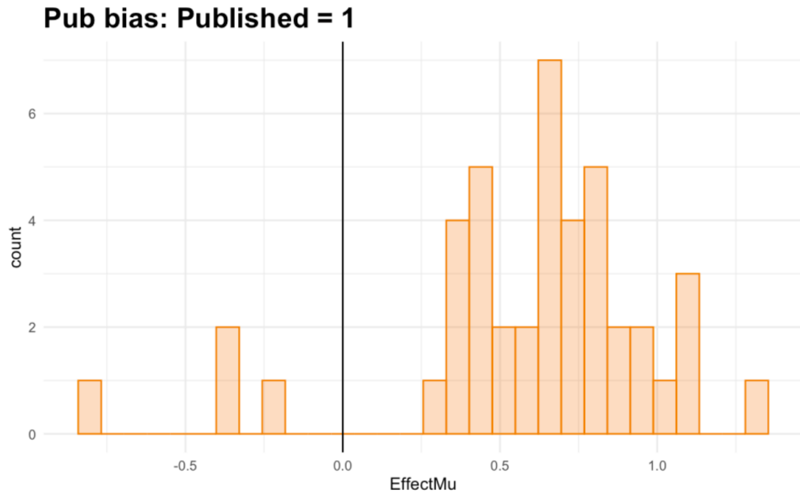


Figure 2: Plot of simulated data: consist of studies that end up being published when the results are significant.

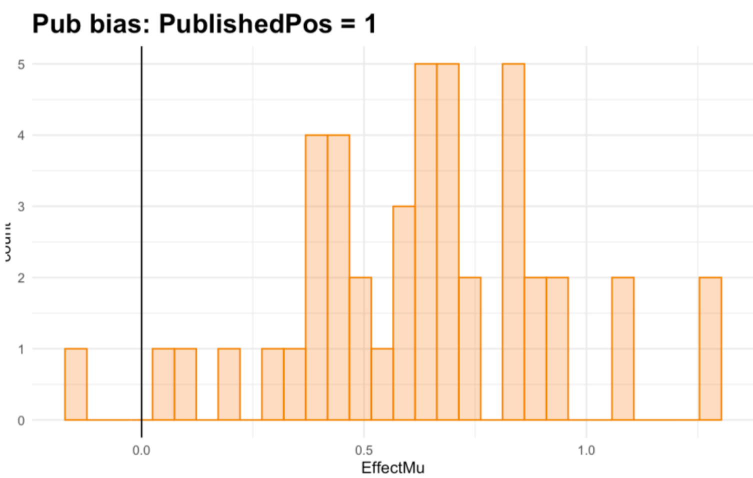


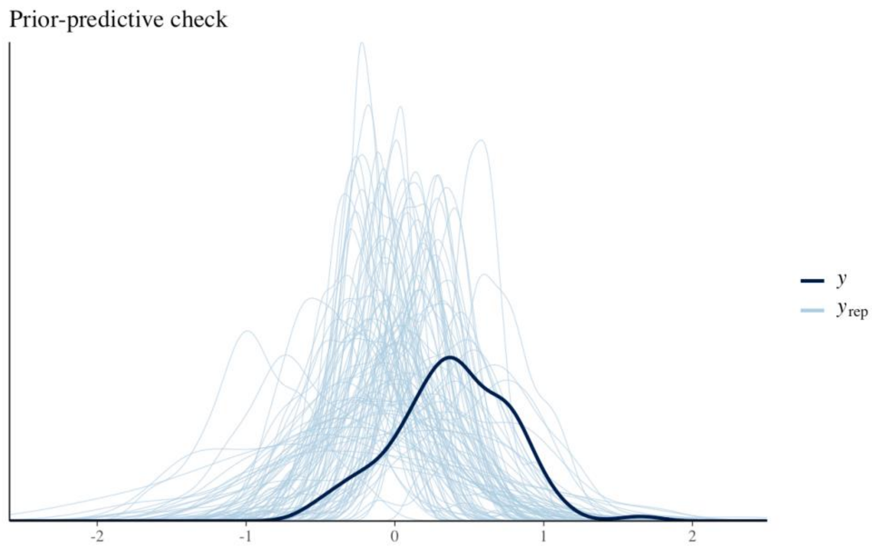
Figure 3: Plot of simulated data: consist of studies that end up being published when the results are significant and outcome is the one as expected.

This formula has been utilized for the Bayesian model to produce a meta-analytic multilevel modeling: (EffectMu | se(EffectSigma) ~ 1 + (1 | Study)),

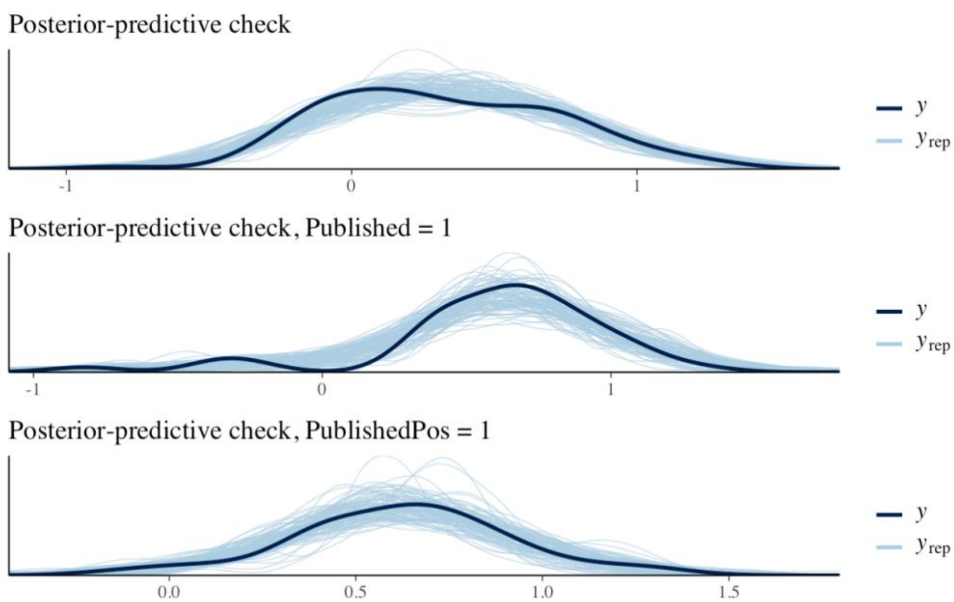
It indicates that the impact size mean and standard error vary between studies, proving that the result is not simply one single point but rather the distribution.

Afterwards, the priors were selected such that they wouldn't be overly restrictive. The prior-predictive check model appears as follows:

Alina Maria Nechita
Student number: 202104831



I have not changed any of the initial priors I established because the range of result values seems to be pretty plausible (within the order of magnitude). The model was then created for the real data:



The first graphic, labeled "Posterior-predictive check," suggests that the model is doing a good job of collecting the data. The center plot suggests that symmetric publication bias is being analyzed, by excluding studies that do not get published (no significant outcome). The final plot suggests that the studies, that have a significant and expected outcome, get published.

The model is attempting to "smooth-out" the result of 0 because the publishing bias is not built into the model (the middle plot). The drawings (blue lines) in all three plots closely resemble the actual data (black line).

Alina Maria Nechita
Student number: 202104831

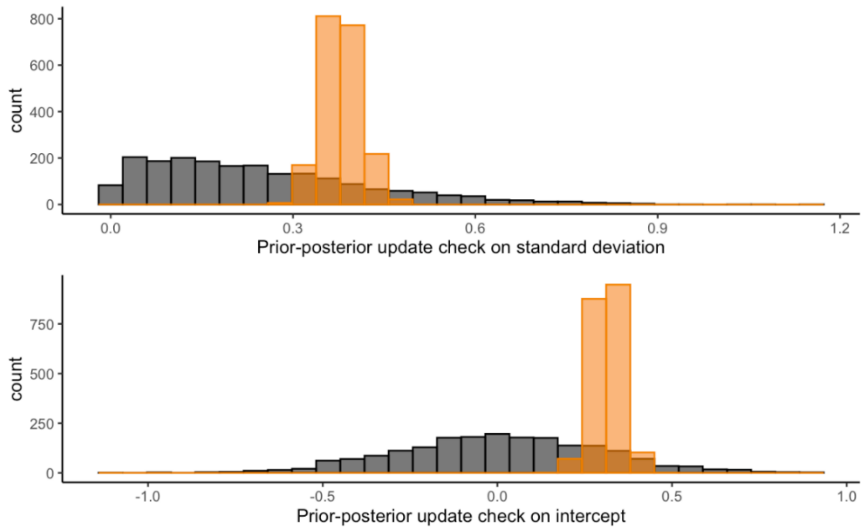


Figure 4: The model is attempting to "smooth-out" the result of 0 because the publishing bias is not built into the model (the middle plot). The drawings (blue lines) in all three plots closely resemble the actual data (black line).

The graphic above shows the prior-posterior update check on simulated data. It incorporates all research without consideration to publication bias. The posterior exhibits confidence and appears to have learnt from the priors.

Question 2:

2. What is the current evidence for distinctive vocal patterns in schizophrenia?

2.1) Describe the data available (studies, participants).

*2.2) Using the model from question 1 analyze the data, visualize and report the findings**: population level effect size; how well studies reflect it; influential studies**, publication bias.*

The 50 research included in the data from Parola et al (2020) were published between 1977 and 2018. Each research had, on average, 13 men and 11 women in the control group and 17 men and 7 women with schizophrenia as participants. Participants with schizophrenia are 35.95 years old on average, with an 8.6 standard deviation. Participants in the control group had an average age of 35.34 and 9 as a standard deviation. The fundamental frequency of pitch (F0), which is the focus variable for pitch variability, changes by 21.6 SD on average in the control group whereas by 26.8 SD in schizophrenia subjects.

The model from the first section of the assignment (with modified variable names) was utilized to examine the data. The results were translated to Cohen's D since various scales for the variables of interest might be used by different papers.

As a result, the pitch F0 mean and standard deviation values for the control and schizophrenia groups were modified.

The following formula is used in the pitch analysis model: yi | se(vi) ~ 1 + (1 | StudyID),

where “yi” is the effect size and the “vi” is standard error of each study, converted to the Cohen’s D scale.

The model from part 1's model was applied, using the identical priors. Thus, actual data were used for the prior and posterior prediction checks:

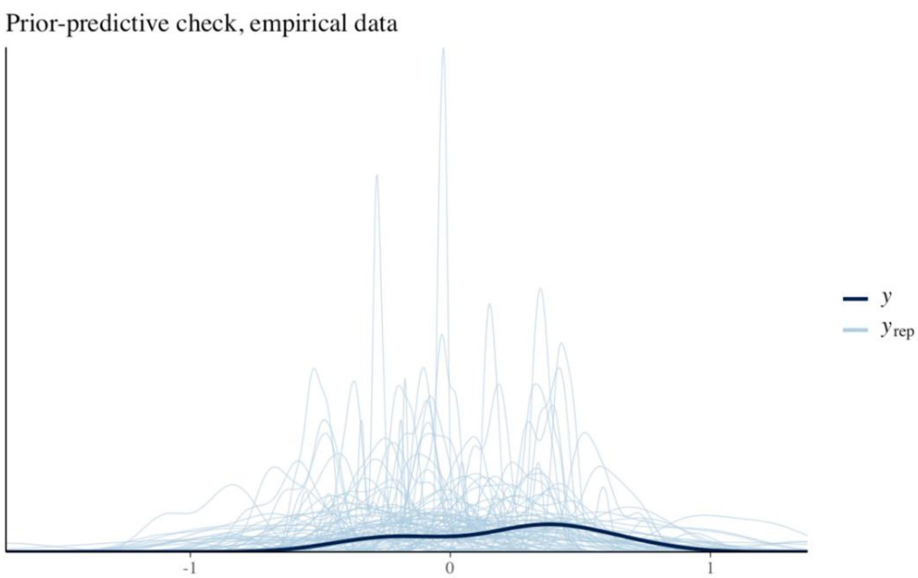
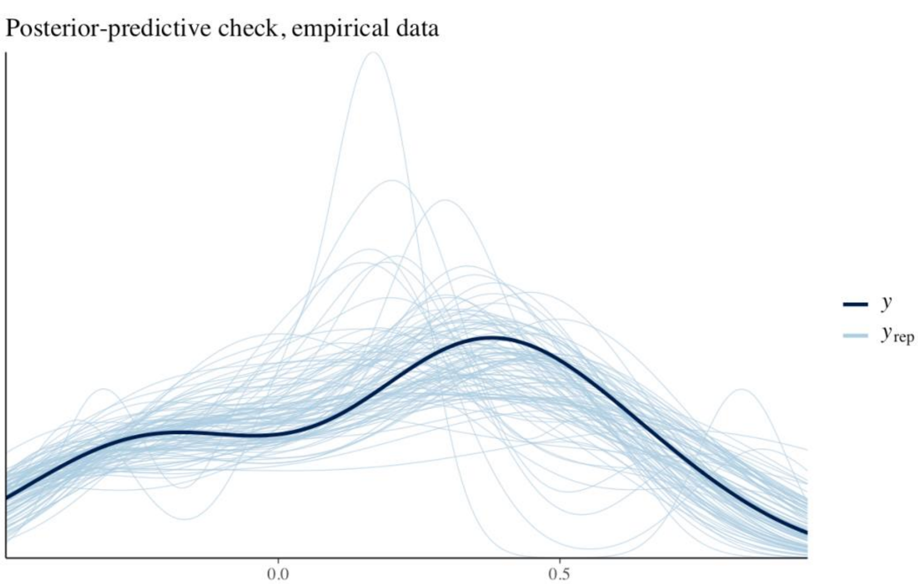


Figure 5: Prior values appear to be in the correct order of magnitude.



Although there is some uncertainty in the model's predictions, it appears that it closely represents the general distribution of the relevant variable. To determine how successfully the model updates from the priors after being exposed to the real data, a prior-posterior update check may be helpful.

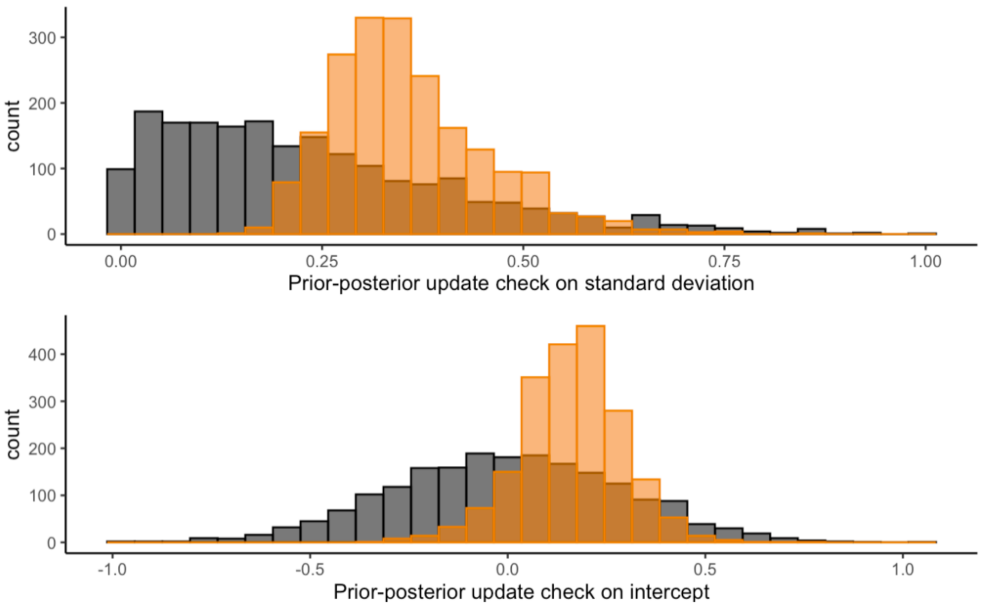
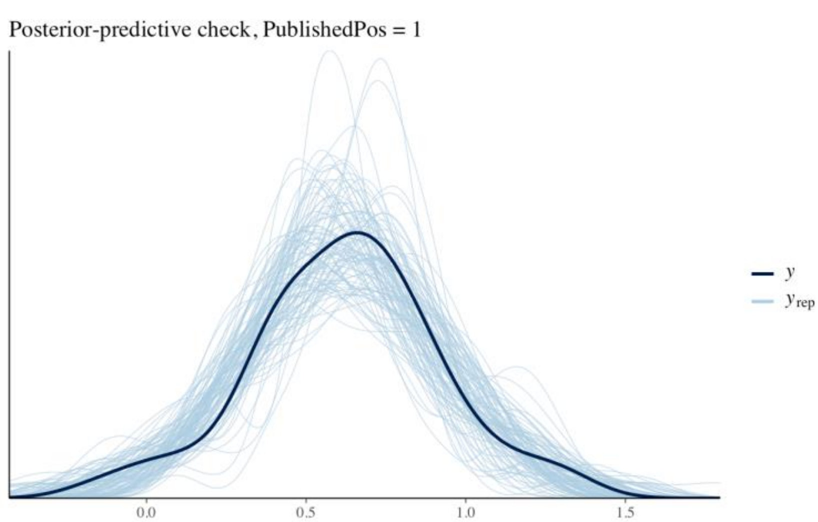
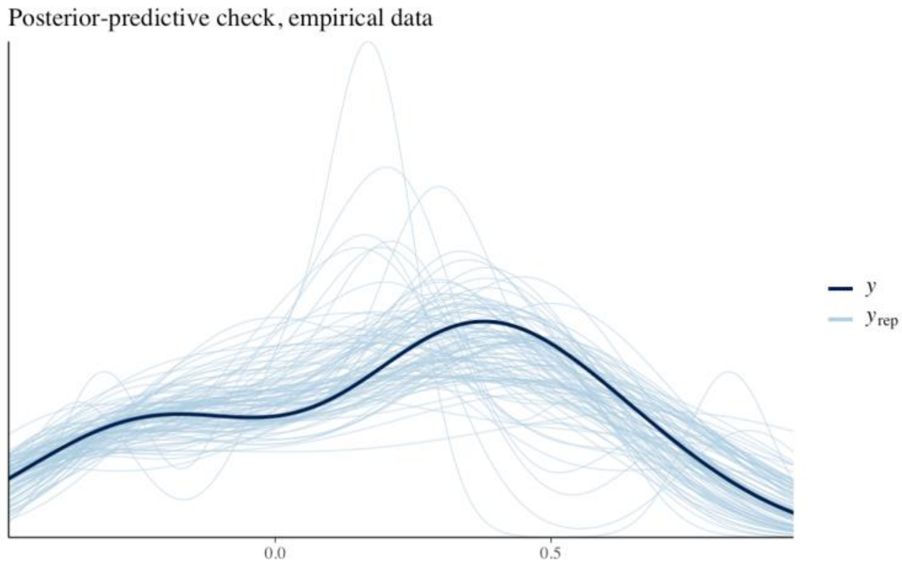


Figure 6: Plots show that both posterior distributions grow significantly more certain and learn from the prior. The priors may not be changed since it is not being pushed at the tails of the prior distribution.

By examining the distribution of effect sizes across studies, the distribution resulting from the model using real data resembles that of the simulated data with asymmetric publication bias:

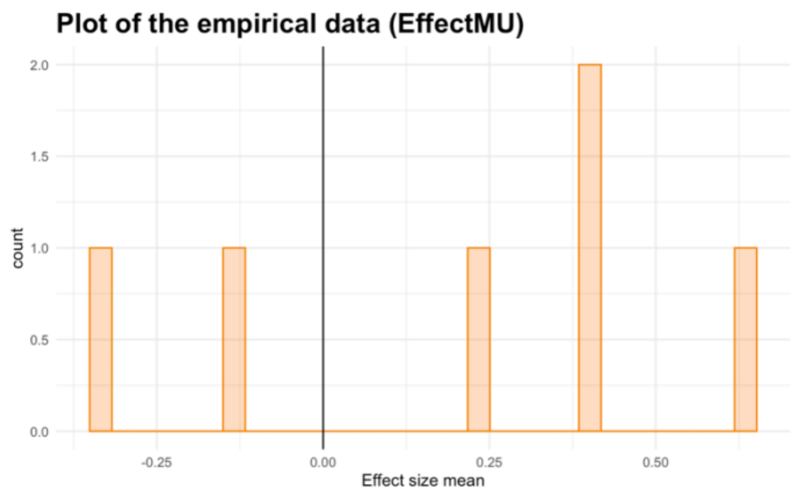




Even though the amount of potential values in the model using simulated data is bigger than that in the model using empirical data, it may be plausible to assume that the empirical data also contains asymmetric publication bias. The published research may be positively biased, thus we should exercise caution when analyzing their findings if the genuine data is in fact subject to publication bias and many studies are not accepted for publication. In this particular instance, the anticipated pitch results are made public.

Figure 7: The distribution of effect sizes in the simulated data is shown in the histogram above, where only experiments that produce significant and anticipated findings are reported.

Alina Maria Nechita
Student number: 202104831



The effect sizes of the studies in the empirical data set are shown in the second histogram. As the data set contains several NAs, only 6 mean impact sizes are calculated in this instance.

However, the actual data distribution resembles the simulated studies of figure 7 the most, which may once again point to this kind of publishing bias (of course, bearing in mind that the data on the left is purely simulated).

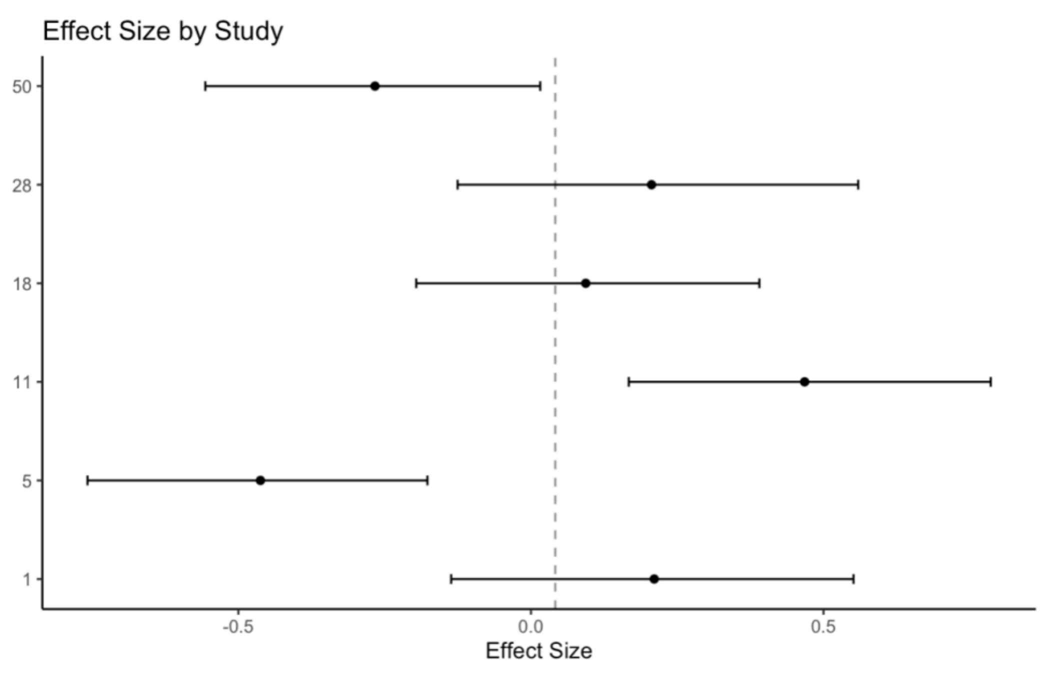


Figure 8: The graphic above shows the predicted impacts of the model (of empirical data).

Each estimated study impact is shown by the mean and upper/lower quantiles, while the dashed line shows the mean effect size.

The fifth, eleventh, and fifty-first investigations appear to miss the overall mean. The fifth and eleventh studies stand out from the others the most.

The posterior prediction check is as follows after running the model on the data without the use of significant studies:

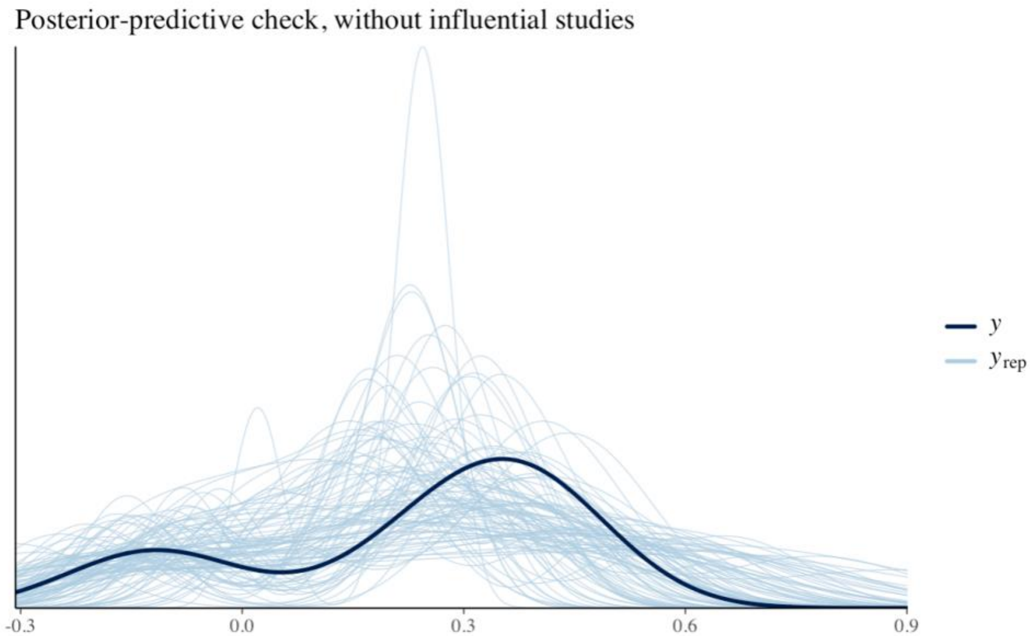


Figure 9: No matter if the outcomes are what was anticipated or not, the output now resembles the simulation of just published research.