Maria Mujemula Olsen
12.10.2022
Methods 3

Methods - Assignment 1

Part 1 – simulation

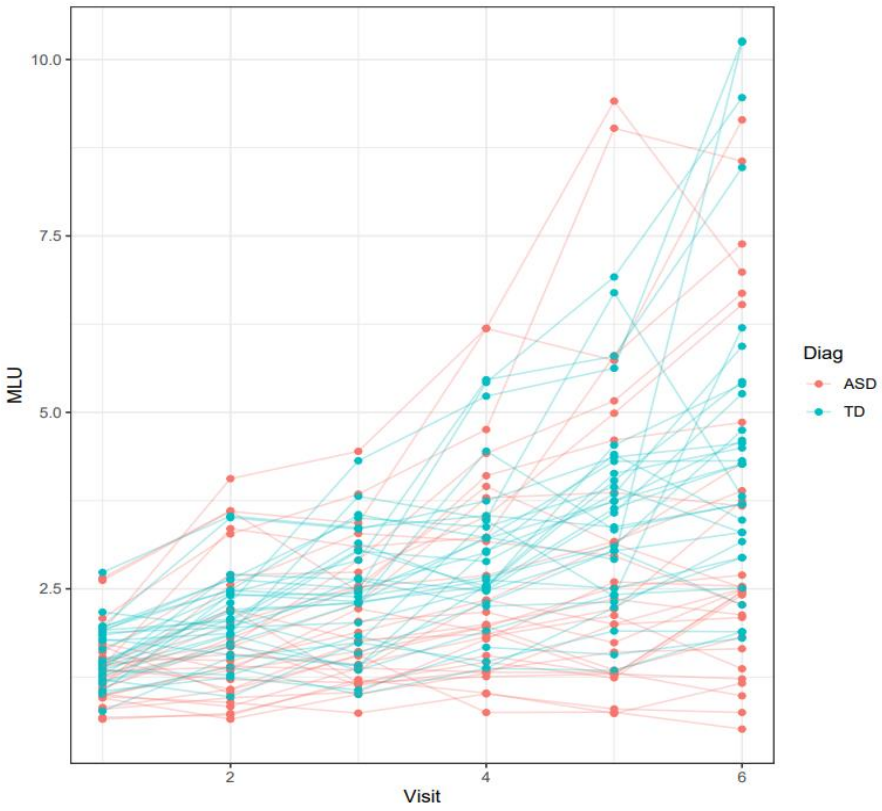
Q1-Briefly describe your simulation process, its goals, and what you have learned from the simulation. Add at least a plot showcasing the results of the simulation. Make a special note on sample size considerations: how much data do you think you will need? what else could you do to increase the precision of your estimates?

The goal of this assignment was to test if there is a difference in the development in language for typically developing children and children diagnosed with autism. The goal of the simulation process was to investigate which model would best describe our real data.

I started by defining parameters for the simulation. The sample size was set to 60 (30 with autism spectrum disorder (ASD) and 30 typically developing (TD), the mean length of utterance (MLU) for visit 1 for both ASD and TD subjects was set to 1.5 with a sd of 0.5 for ASD and 0.3 for TD, the change in MLU for ASD subjects was set to 0.15 with a sd of 0.1, and 0.2 for TD with a sd of 0.08. The error was set to 0.2. The parameters where set on a logscale.

The MLU for each subject for each visit (6 in total) was defined by the formula:

MLU ~ Diagnosis + Diagnosis:Visit + (1 + Visit|ID)



Simulated data visualized.

Then the following 3 formulas where defined in order to investigate which formula would best describe the simulated data:

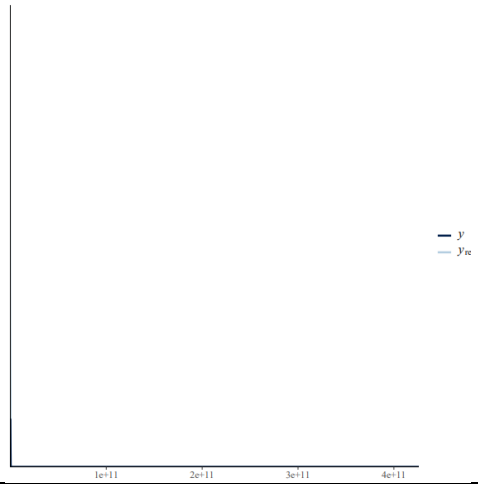
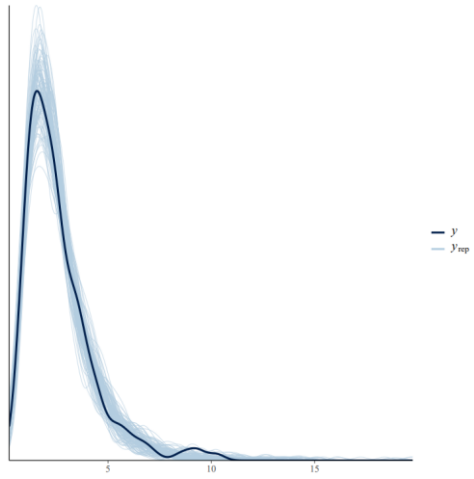
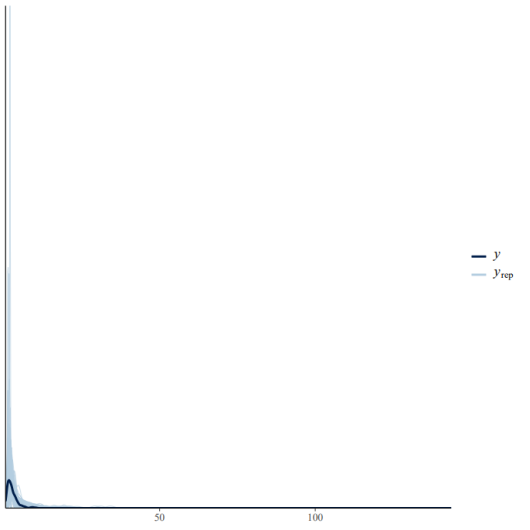
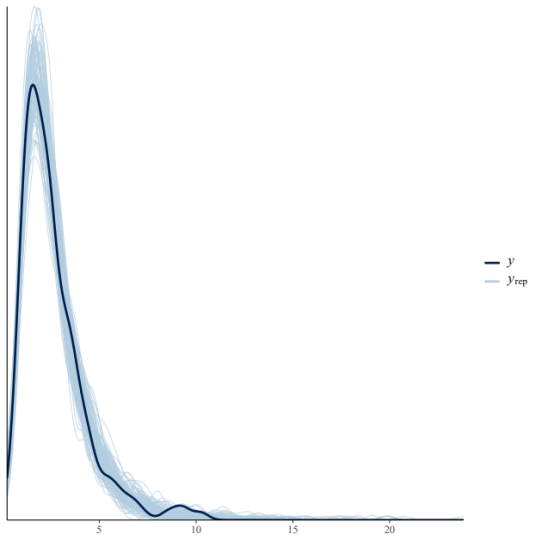
MLU_1 <- bf(MLU ~ 0 + Diagnosis)

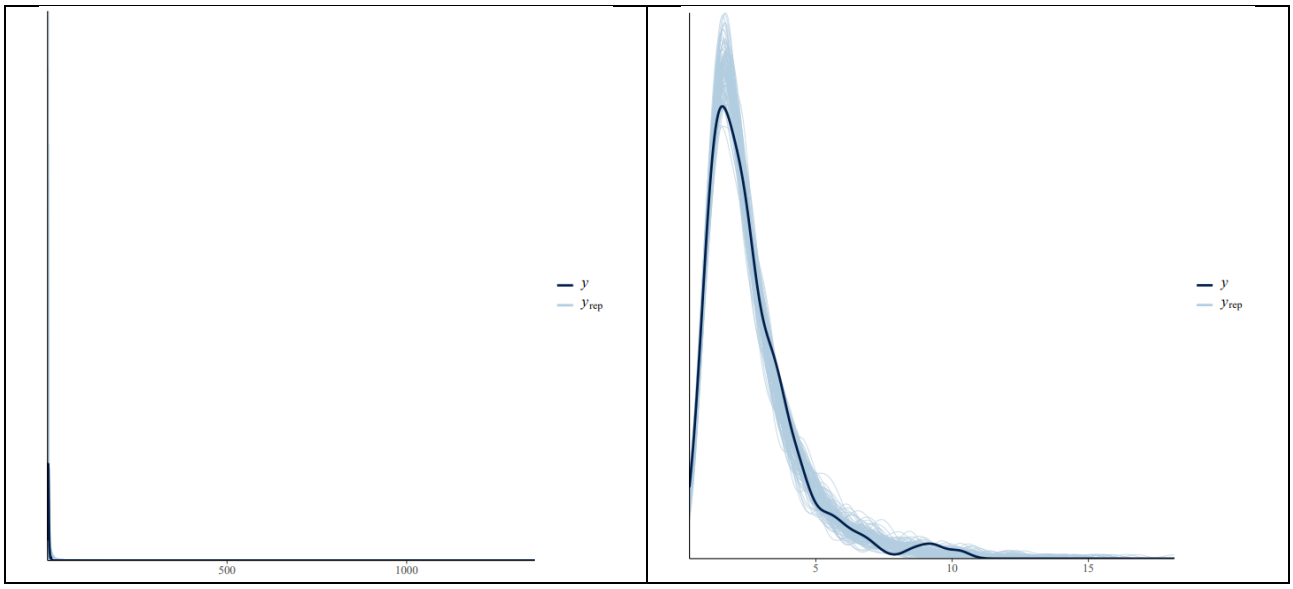
MLU_f2 <- bf(MLU ~ 0 + Diagnosis + Diagnosis:Visit)

MLU_f_3 <- bf(MLU ~ 0 + Diagnosis + Diagnosis:Visit + (1 + Visit|ID))

Weekly informed priors were then set in order to be as least biased as possible and let the data convince us if there truly is a difference in MLU between ASD and TD (if there is a difference in the language development between the two groups).

The models where then fitted and checked:

Prior model 1	**Posterior model 1**
	
Prior model 2	**Posterior model 2**
	
Prior model 3	**Posterior model 3**

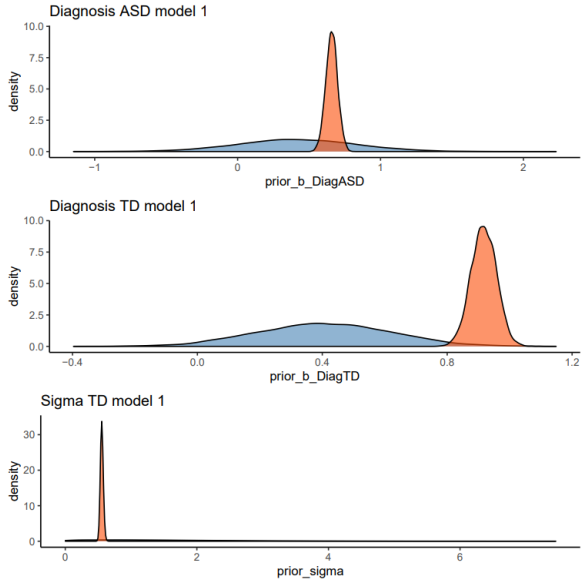


The prior-posterior checks looks fine (even though the spread for all the priors are extremely high)

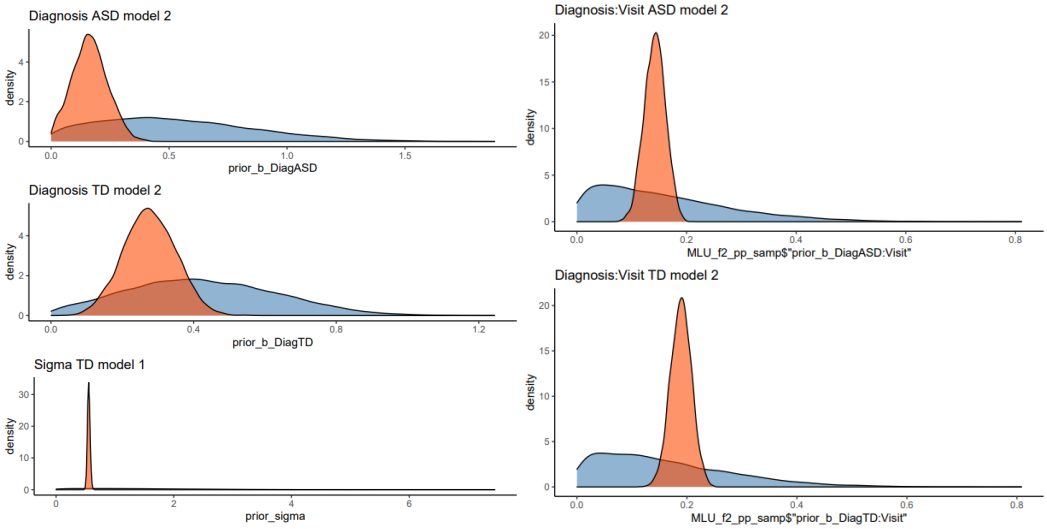
(From the Prior-Posterior checks model 1 and model 3 looks best for the posteriors, as they have the least spread/spread in a range that makes sense [0:15] vs [0:20].)

Prior-posterior update plots where then made in order to see how much our models had learned from the simulated data:

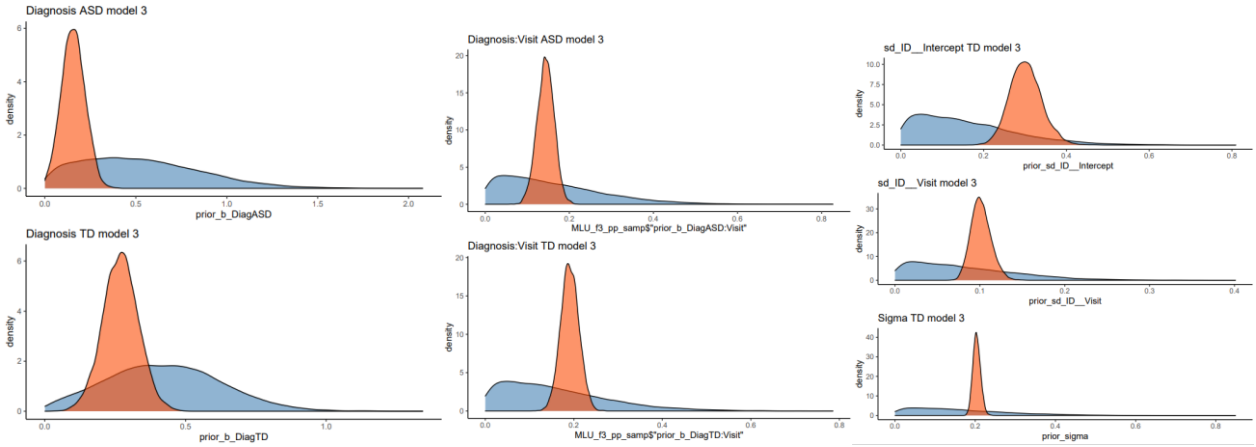
Model 1



Model 2



Model 3



The update plots generally look acceptable for both model 2 and model 3. The update plot for the parameter DiagnosisTD in model 1 has not learned much from the priors.

Looking at the summary output for the three models we can see that all models converged fully because rhat is 1 for all parameters for all the models.

(*Models 1 and 3 have the highest bulk and tail values across parameters, indicating that the these models have explored more samples in the Markov chains than model 2 before converging – meaning that the respective estimates for models 1 and 3 are based on more "evidence". However, the high bulk and tail values for model 1 might reflect that that the model had to look harder to find a connection between the simulated data and the formula.)*

```
 Family: lognormal
  Links: mu = identity; sigma = identity
Formula: MLU ~ 0 + Diag
   Data: d (Number of observations: 360)
  Draws: 2 chains, each with iter = 5000; warmup = 1000; thin = 1;
         total post-warmup draws = 8000

Population-Level Effects:
        Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
DiagASD     0.66      0.04     0.58     0.74 1.00     6587     5252
DiagTD      0.92      0.04     0.84     0.99 1.00     6028     5027

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.56      0.02     0.52     0.60 1.00     5371     4802

Draws were sampled using sample(hmc). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

Summary output for model 1

```
 Family: lognormal
  Links: mu = identity; sigma = identity
Formula: MLU ~ 0 + Diag + Diag:Visit
   Data: d (Number of observations: 360)
  Draws: 2 chains, each with iter = 5000; warmup = 1000; thin = 1;
         total post-warmup draws = 8000

Population-Level Effects:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
DiagASD           0.16      0.07     0.03     0.31 1.00     2993     1540
DiagTD            0.28      0.07     0.13     0.42 1.00     3053     2929
DiagASD:Visit     0.14      0.02     0.10     0.18 1.00     3319     2645
DiagTD:Visit      0.19      0.02     0.15     0.23 1.00     3030     3611

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.47      0.02     0.44     0.50 1.00     4796     5042

Draws were sampled using sample(hmc). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

Summary output for model 2

```
 Family: lognormal
  Links: mu = identity; sigma = identity
Formula: MLU ~ 0 + Diag + Diag:Visit + (1 + Visit | ID)
   Data: d (Number of observations: 360)
  Draws: 2 chains, each with iter = 5000; warmup = 1000; thin = 1;
         total post-warmup draws = 8000

Group-Level Effects:
~ID (Number of levels: 60)
                    Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)           0.30      0.04     0.23     0.38 1.00     4789     5561
sd(Visit)               0.10      0.01     0.08     0.13 1.00     2102     3067
cor(Intercept,Visit)   -0.26      0.15    -0.54     0.06 1.00     1344     1818

Population-Level Effects:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
DiagASD          0.16      0.06     0.03     0.28 1.00     4368     3908
DiagTD           0.27      0.06     0.14     0.39 1.00     5312     4771
DiagASD:Visit    0.14      0.02     0.10     0.18 1.00     2702     4044
DiagTD:Visit     0.19      0.02     0.15     0.23 1.00     2722     3827

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.20      0.01     0.19     0.22 1.00     5683     5327

Draws were sampled using sample(hmc). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

Summary output for model 3

I therefore used the leave one out method for comparing the three models. The elpd_diff value favors model 3 as the best model with having an expected log pointwise predictive density 243.3 higher than model 2 and 300.6 higher than model 1. Looking at the weight of the models in the model comparison, model 3 also has the highest weight (1.00) (explaining most of the variance?).

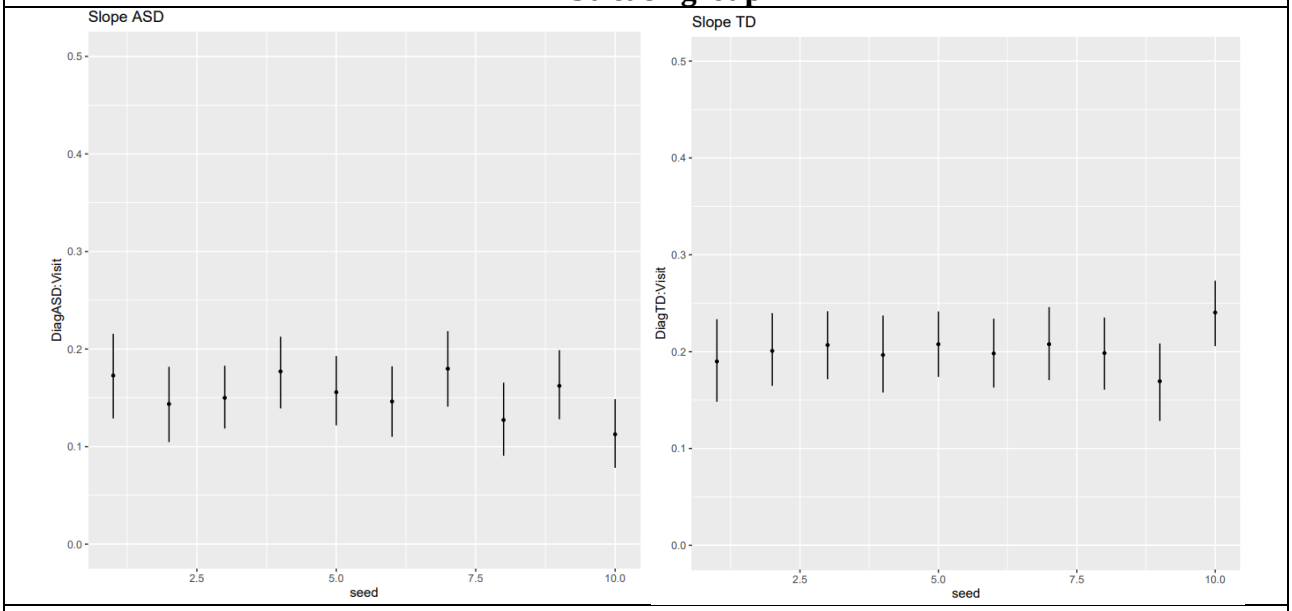
```
                      elpd_diff se_diff
MLU_f3_prior_posterior    0.0       0.0
MLU_f2_prior_posterior -242.3      18.3
MLU_f1_prior_posterior -300.6      17.0
Method: stacking
------
       weight
model1 0.000
model2 0.000
model3 1.000
```

From this simulation process model 3 (MLU ~ Diagnosis + Diagnosis:Visit + (1 + Visit|ID)) seems like the model that best describes the simulated data. Therefore model 3 will be used in part 2 for analyzing the true data.

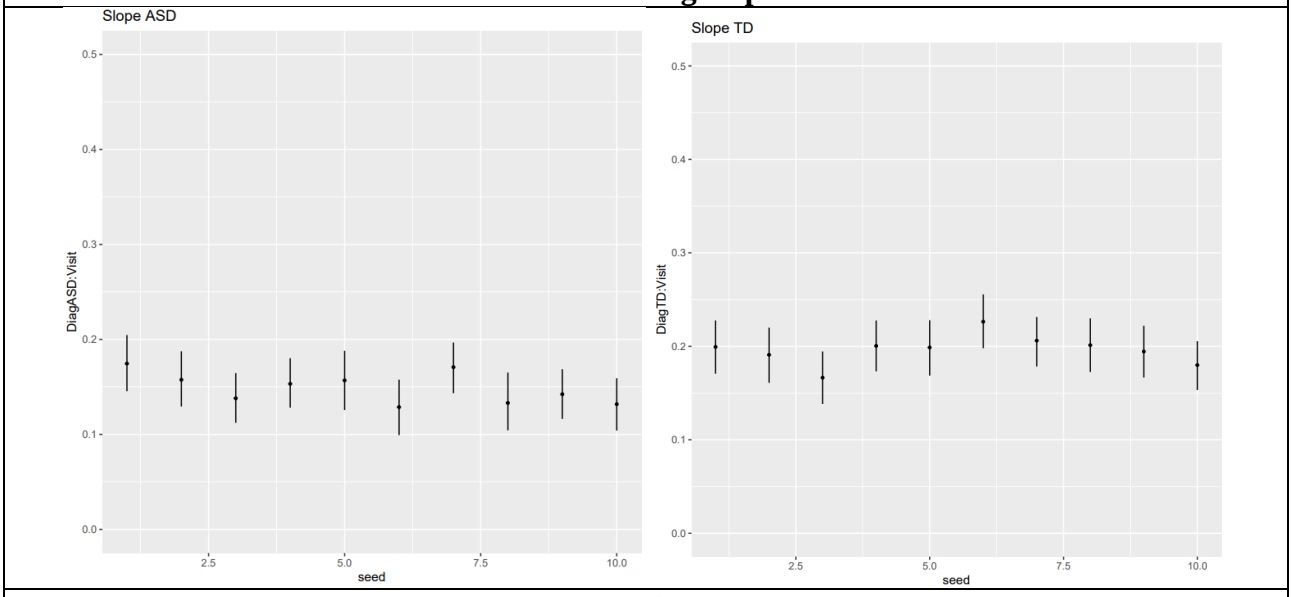
Maria Mujemula Olsen
12.10.2022
Methods 3

Now that model 3 has been chosen, a precision analysis was run to asses which sample size would be optimal. Data with a sample size of 30, 50, 70 and 100 subjects in each group was simulated and run through 10 different seeds each.

N=30 each group

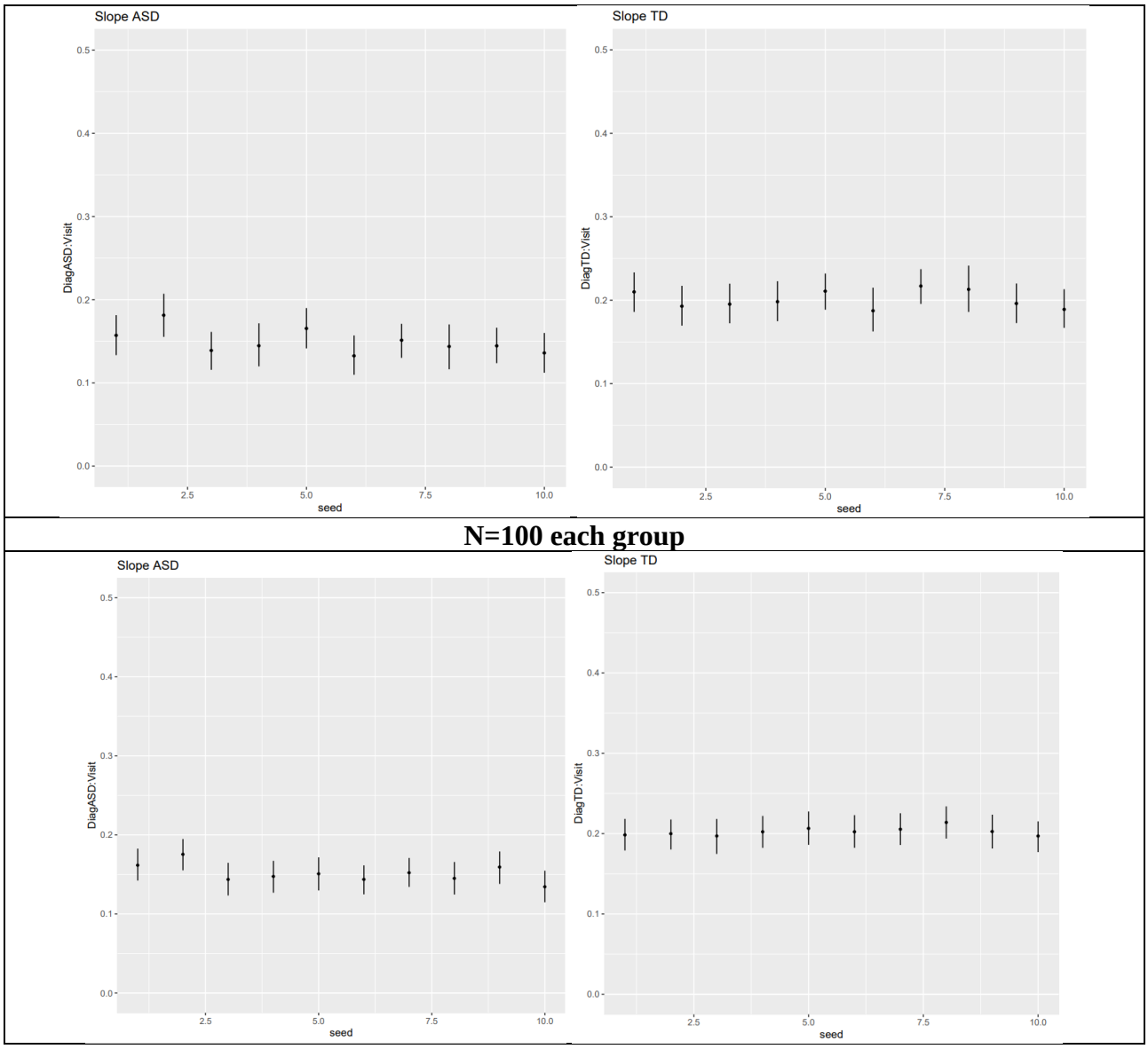


N=50 each group



N=70 each group

Maria Mujemula Olsen
12.10.2022
Methods 3



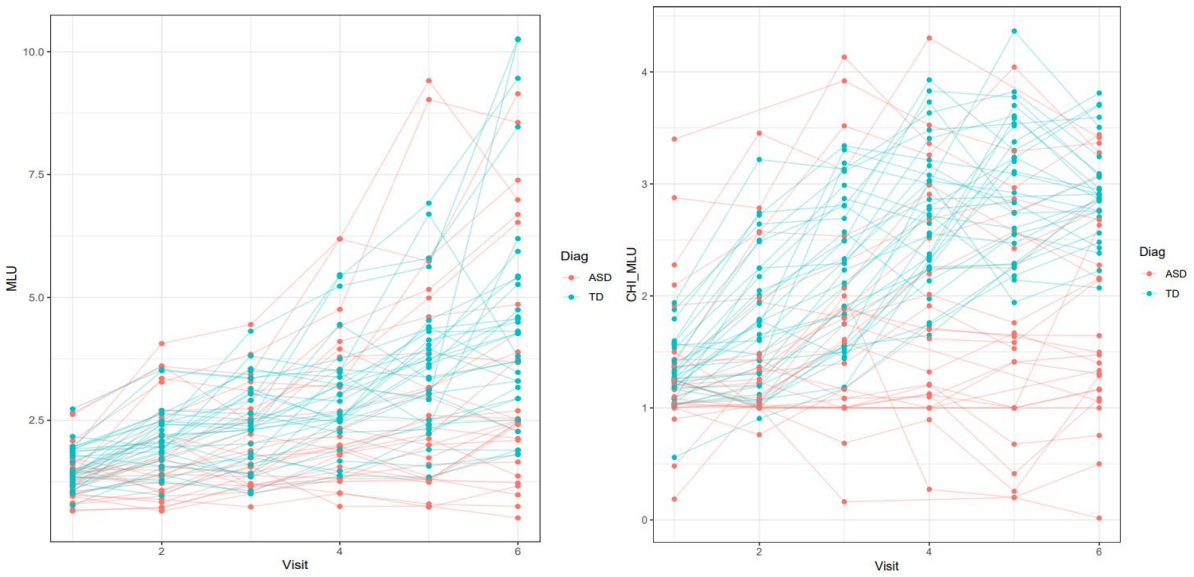
Sample size each group	Sample size total	Slope for	Number of simulations with true effect within ci-95%	Precent
30	60	ASD	10/10	100%
50	100	ASD	10/10	100%
70	140	ASD	9/10	90%
100	200	ASD	9/10	90%
30	60	TD	9/10	90%
50	100	TD	9/10	90%
70	140	TD	10/10	100%
100	200	TD	10/10	100%

With a sample size of 30 and 50 in each group the true effect of the slope for ASD is within the 95% confidence interval in 10 out of 10 simulations and 9 out of 10 for TD, where this is true for 9 out of 10 simulations for sample sizes with 70 and 100 in each group for ASD and 10 out of 10 for TD. However, the estimates are closer to the true effect and the spread of the 95% intervals are smaller for sample sizes with 70 and 100 pr. group. (It is worth noting that exploring more seeds would probable have given an even better intuition of which sample size is the most reliable, but this wasn't done due to lack of time to run trough all simulations). From these findings, a bigger sample size of 140 or 200 would be preferable, but taking recourses into account, a sample size of 60 is acceptable as the chance of the true value being within the 95% confidence interval would still be 90% (TD) and 100% (ASD).

Part 2 – Analyse actual data

Q2-Briefly describe the empirical data and how they compare to what you learned from the simulation(what can you learn from them?). Briefly describe your model(s) and model quality. Report the findings: how does development differ between autistic and neurotypical children (N.B. remember to report both population and individual level findings)? Which additional factors should be included in the model(cue: model comparison!)? Add at least one plot showcasing each of your key findings.

The empirical data includes the MLU for 66 children (31 ASD, 35 TD) for 6 visits. The empirical data shows that the subjects generally have a lower MLU than in the simulated data (4<10).



Left: MLU pr. visit, simulated data. Right: MLU pr. Visit, empirical data

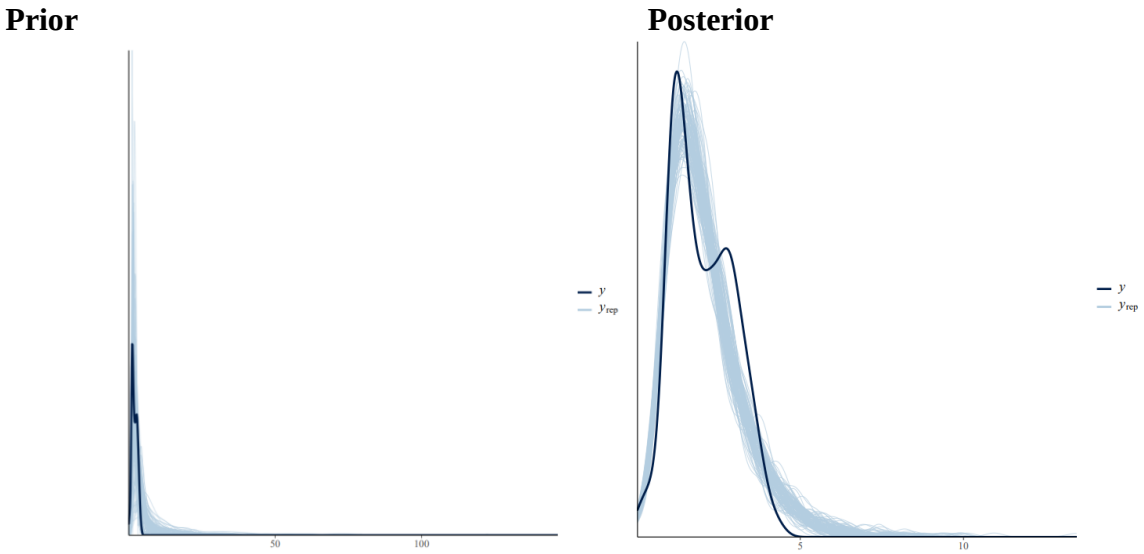
As we learned from the simulated data, a sample size of 100 or over would be preferable but based on the power analysis from the simulated data, the chance of the true effect being within the 95% confidence interval is still 100% for this sample size.

Maria Mujemula Olsen
12.10.2022
Methods 3

The following model was chosen for the analysis (based on the simulation process):

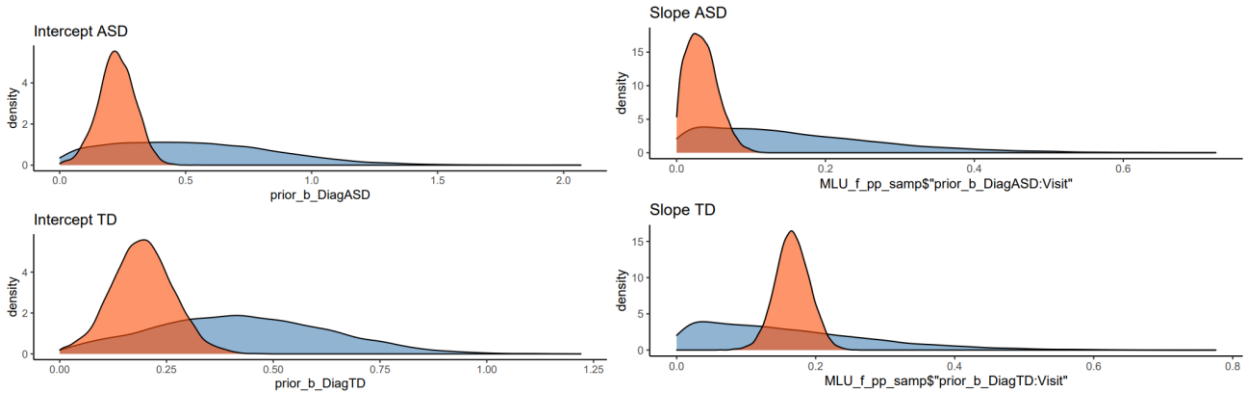
bf(CHI_MLU ~ 0 + Diagnosis + Diagnosis:Visit + (1 + Visit|Child.ID))

The same weekly informed priors from the simulation process were set for this model. Then the prior and the posterior distribution was made.

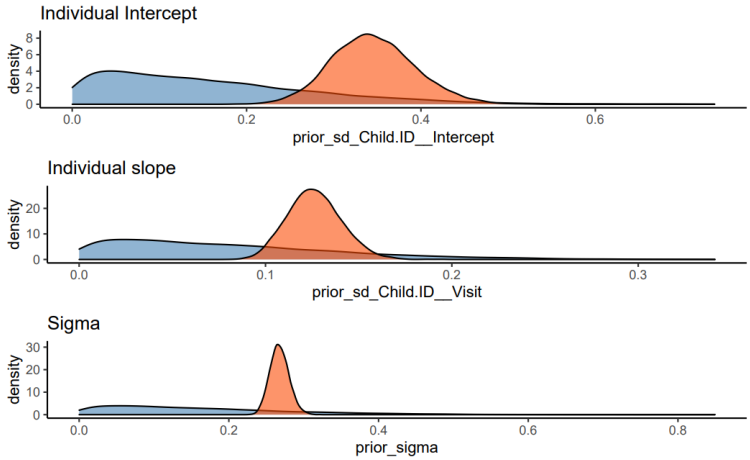


PP_checks for the prior and the posterior

Looking at the pp_check for the posterior distribution, it looks like there are two distributions. This could indicate that the distributions of MLU for ASD subjects and TD subjects are different. Prior-posterior update plots for the parameters were then made to see how much our data had learned from the priors as well as to see underlying distributions.



Maria Mujemula Olsen
12.10.2022
Methods 3



Prior-posterior update plots with the priors in blue and the posteriors in orange. Top left: distributions of the intercept for ASD and TD. Top right: distributions of the slope for ASD and TD. Bottom: distributions of the varying intercepts and slops as well as distribution for the error.

All the plots look fine. The distributions for the individual intercepts and slopes as well as sigma could look better. We could go back and make more informed priors by changing the mu or lowering the sd – however for now we just want to let the empirical data talk for itself. Interestingly there is not much overlap between the distributions of the posteriors for the slopes for ASD and TD subjects with TD subjects having a higher slope than ASD subjects. This explains the pp_check for the posterior's underlying distributions.

```
 Family: lognormal
  Links: mu = identity; sigma = identity
Formula: CHI_MLU ~ 0 + Diag + Diag:Visit + (1 + Visit | Child.ID)
   Data: d_real_no_0 (Number of observations: 349)
  Draws: 2 chains, each with iter = 5000; warmup = 1000; thin = 1;
         total post-warmup draws = 8000

Group-Level Effects:
~Child.ID (Number of levels: 61)
                  Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)         0.35      0.05     0.26     0.44 1.00     2801     4900
sd(Visit)             0.13      0.01     0.10     0.16 1.00      878     2222
cor(Intercept,Visit)  -0.46      0.14    -0.70    -0.15 1.00      651     1426

Population-Level Effects:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
DiagASD          0.23      0.07     0.08     0.37 1.00     2251     1738
DiagTD           0.19      0.07     0.05     0.34 1.00     2186     1359
DiagASD:Visit    0.03      0.02     0.00     0.08 1.00     1823     2816
DiagTD:Visit     0.17      0.02     0.12     0.21 1.00     1338     2130

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.27      0.01     0.24     0.29 1.00     4200     5830

Draws were sampled using sample(hmc). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

Summary output for fitted model

The summary output for the fitted model gives us the effects for the parameters of the models. All the parameters have a rhat value of 1, telling us that all then parameters converged which is good. The bulk and tail values for all the parameters are also quite high (the brm model was set with 5000 iterations and 2 chains) with most of the values being above 1000. This is good as this means that the model has explored quite a lot before converging.

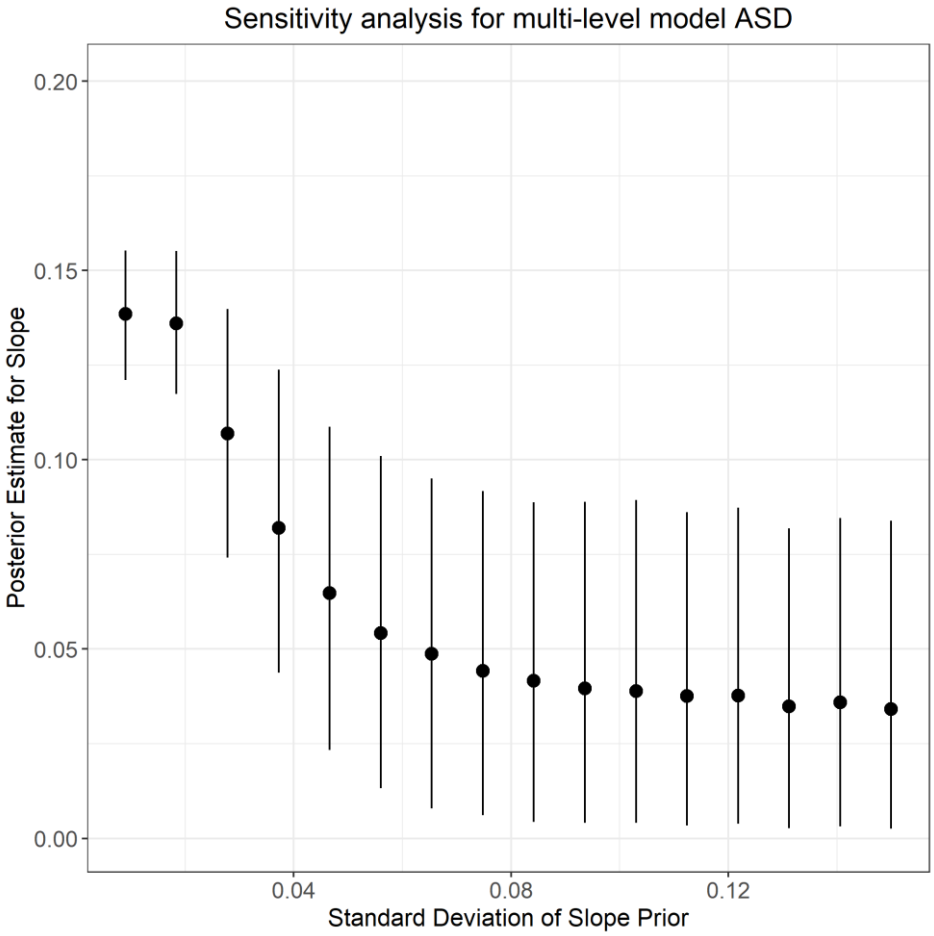
At the group level effects we see that the estimated sd for the intercept is 0.35, this meaning that the individual intercepts vary from the estimated intercept (DiagASD=0.23, DiagTD=0.19) with 0.35 on a logscale. The estimate for the slope is 0.13, meaning that the individual slope varies with 0.13. On the population level, the estimate MLU for DiagASD is 0.23 and for DiagTD is 0.19 on a logscale, meaning that the estimated MLU for an ASD child at visit 1 is 1.47 and for a TD child it is 1.66. It also tells us that for each visit the MLU increases with 03% for ASD children where it increases with 17% for TD children. This supports our hypothesis that children with autism spectrum disorder have a slower language development than typically developing children.

We then test our hypothesis that the change in MLU for children with ASD will be lower than that of TD (DiagASD<DiagTD):

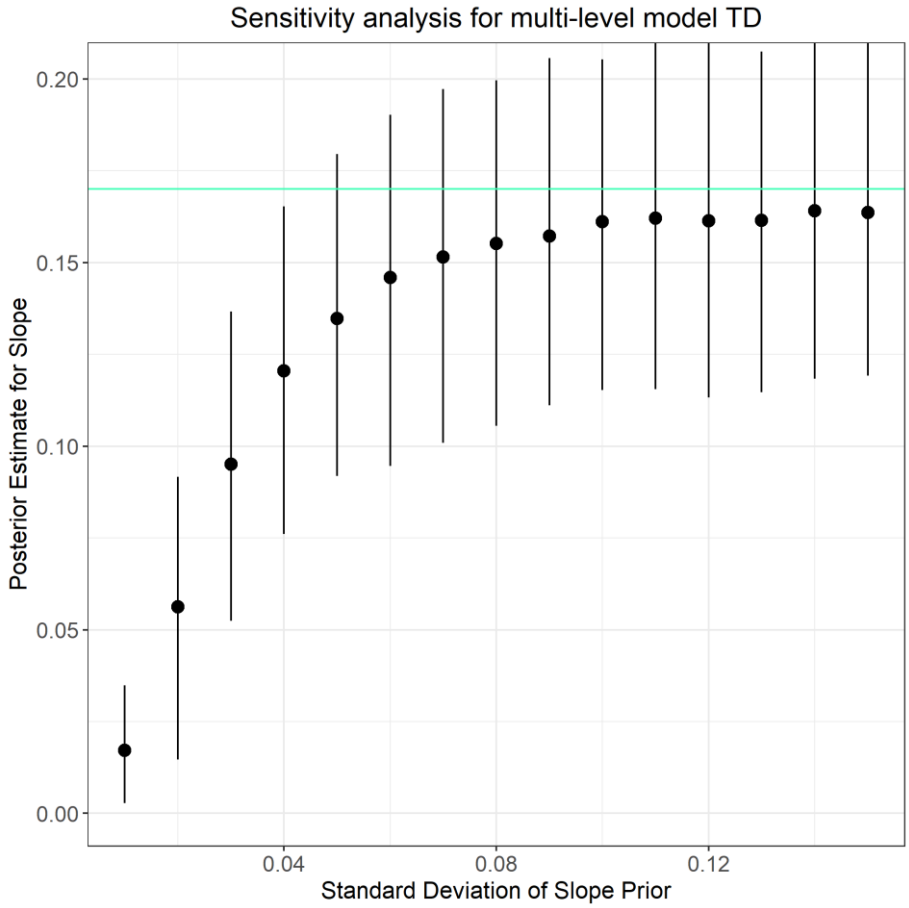
Hypothesis <chr>	Estimate <dbl>	Est.Error <dbl>	CI.Lower <dbl>	CI.Upper <dbl>	Evid.Ratio <dbl>	Post.Prob <dbl>	Star <chr>
(DiagASD:Visit)-(... < 0	-0.13	0.03	-0.18	-0.08	Inf	1	*

Looking at the posterior probability (Post.Prob) we see that given the model, with the priors we set, and given the data, the probability of the change in MLU of a child with ASD being lower than that of a TD child is 1 – in other words, all the ASD children's change in MLU is lower than that of the TD children. This is also reflected in the prior-posterior update plot. The estimated difference is -0.13, meaning that ASD children have are estimated to be 0.13 percentage points lower than that of TD children.

If we were to make more informed priors based on the estimates from the fitted model (model 3) in the simulation process, the following robustness checks showcases how the prior would affect the posterior distribution for the slopes (change in MLU for each visit) for ASD and TD children, based on which sd value the prior has.



Sensitivity analysis for the slope of ASD children. (prior mu: 0.14)

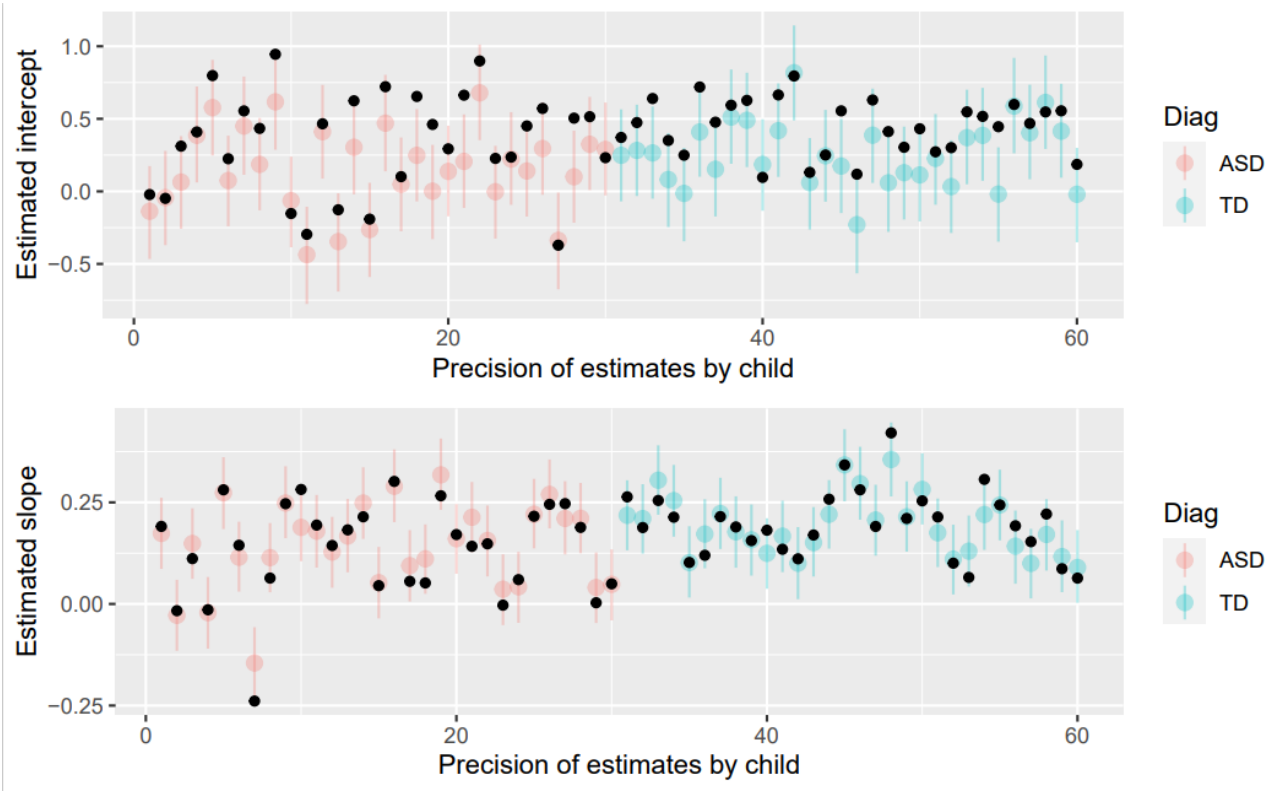


Sensitivity analysis for the slope of TD children. (prior mu: 0.17)
Note: Sara Kjær Kristensen from my study group made this plot

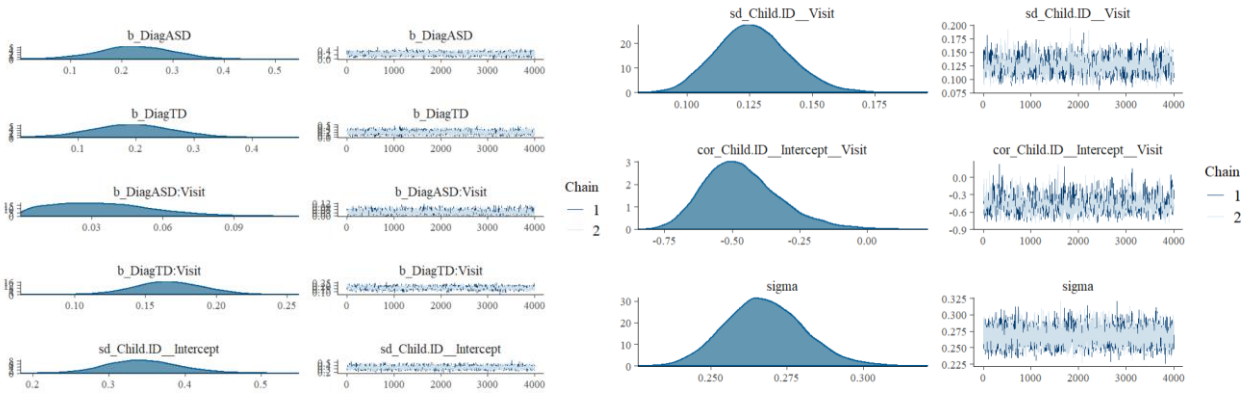
From the plots it is seen that setting a standard deviation of about 0.06 the priors start having a bigger and more noticeable affect on the posterior. In this analysis the weakly informed priors for both slopes set sd=0.2. From the robustness check, it then looks like these priors do not affect the posterior that much. In this analysis, this is fine as we want the empirical data to talk for itself. If we wanted the empirical data to work harder to convince us of a correlation different from our assumption/prior knowledge, making informed priors with a sd<0.06 would be worth considering.

Maria Mujemula Olsen
12.10.2022
Methods 3

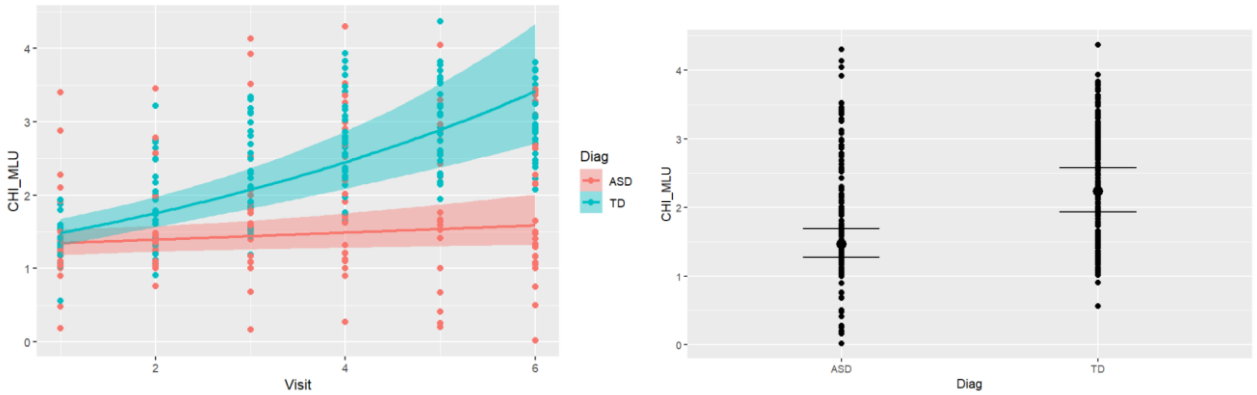
Appendix



Estimate plot for simulated data



Traceplots for model on emperical data



Conditional effects plots