Daniel, Jesper, Pernille KJ and Astrid

Welcome to the third exciting part of the Language Development in ASD exercise

In this part of the assignment, we try to figure out how a new study should be planned (i.e. how many participants?) in order to have enough power to replicate the findings (ensuring our sample size is adequate, our alpha at 0.05 and our beta at 0.8):

Link to code

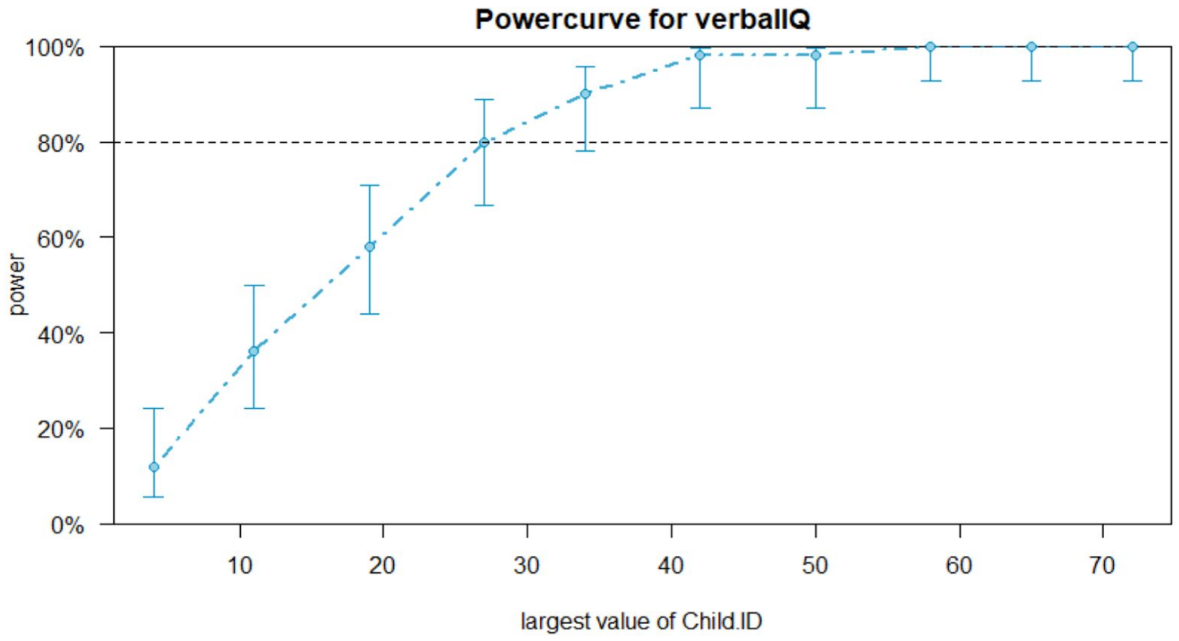
If we trust the estimates of the current study. Report the power analysis and comment on what you can (or cannot) use its estimates for.

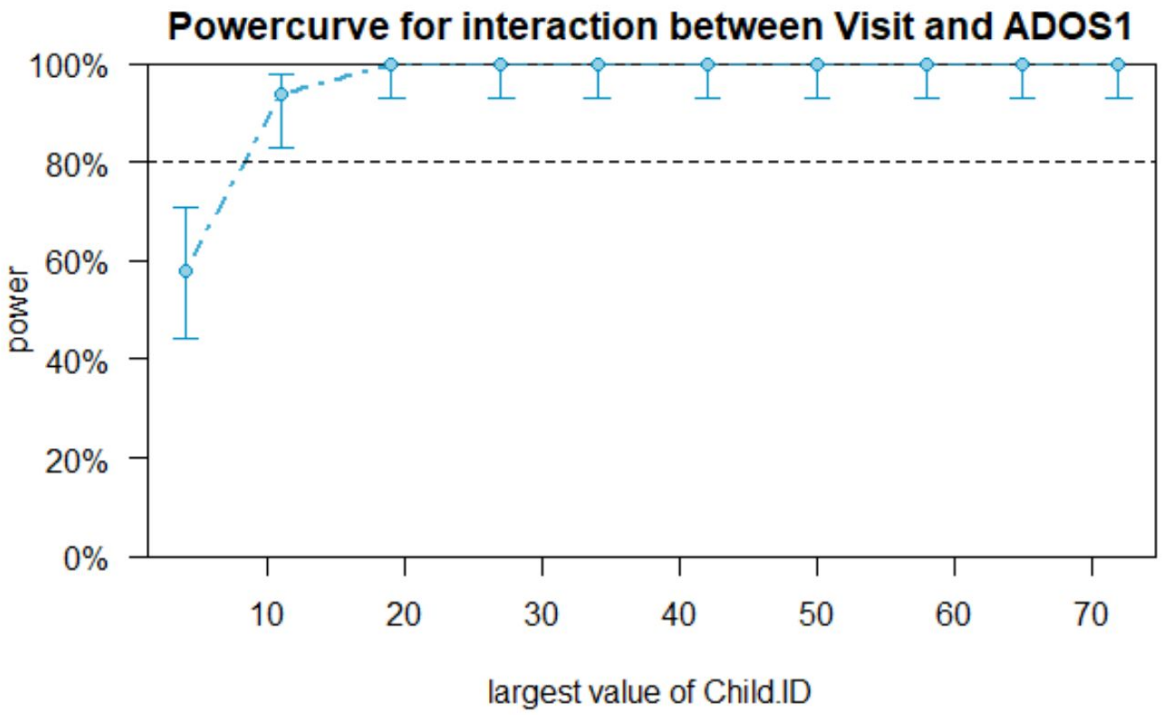
Since our study is concerned with the interaction between the nr. visit and the ADOS score on the child's mean length of utterance (henceforth: MLU), this is what we will interpret, as well as the effect of verbal IQ on the MLU.

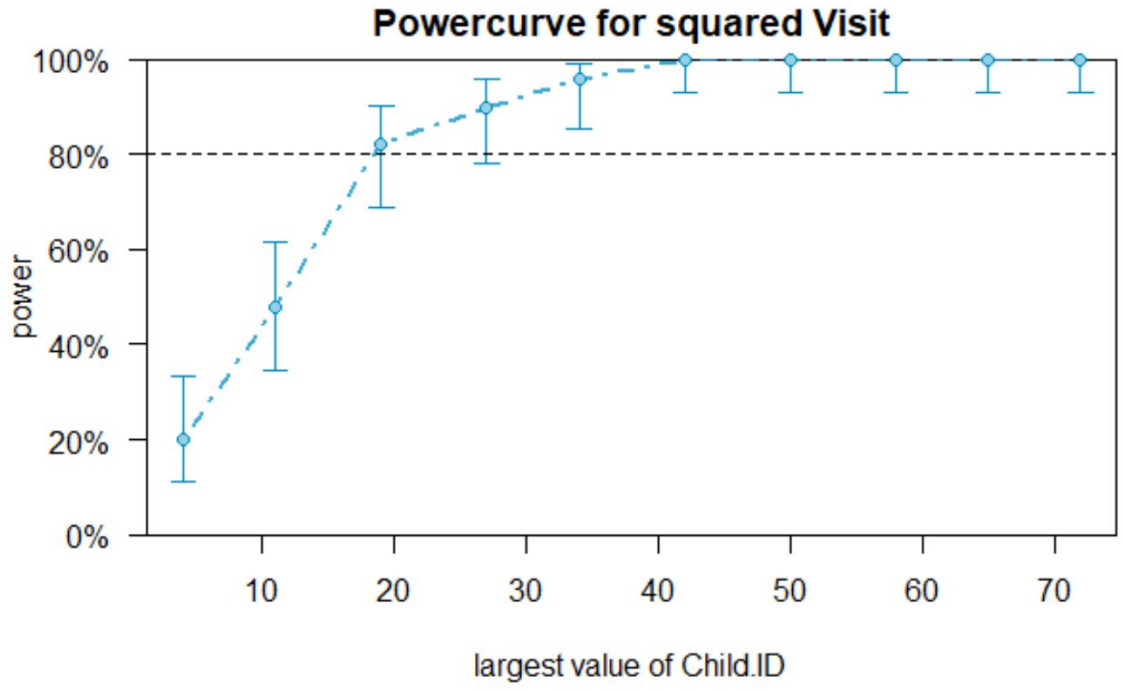
The interaction has a power of 100 % (confidence intervals based on 50 simulations) (92.89-100), with an estimate of -0.02.

VerbalIQ has a power of 100% (92.89-100), with an estimate of 0.07.

Visit^2 has a power of 100% (92.89-100), with an estimate of -0.047. see power curves below:







Since the power is high we are likely to avoid making a type 2 error (false negative). However, conventionally, we should aim for a power of 80% to avoid wasting resources. Therefore we consider this study overpowered.

If we are skeptical of the current study. Report the power analysis and comment on what you can (or cannot) use its estimates for.

The estimates are not any worse because the study is overpowered. The higher the power, the more we can trust the estimates. However, we are necessarily a lot more sure of these estimates, compared to how many more resources has been used to gain this extra power. It is always a trade-off and depends on the context. For our purposes, it was probably not worth it. However, had this data been based on i.e. web scraping or a meta-study it would not be a problem that it was overpowered.

Identify and justify a minimum effect size for each of your relevant effects. How would you perform a more conservative power analysis?

Here it would make sense to look at prior research, i.e. this present study (read: Riccardo's). Ideally, we should look at other studies as well. We would like to take a conservative approach and set our minimum effect size at half the size of the one used in the present study. Therefore a minimum effect size for verbal IQ is 0.035, the interaction between ADOS score and visit is -0.01, and visit^2 is -0.02.

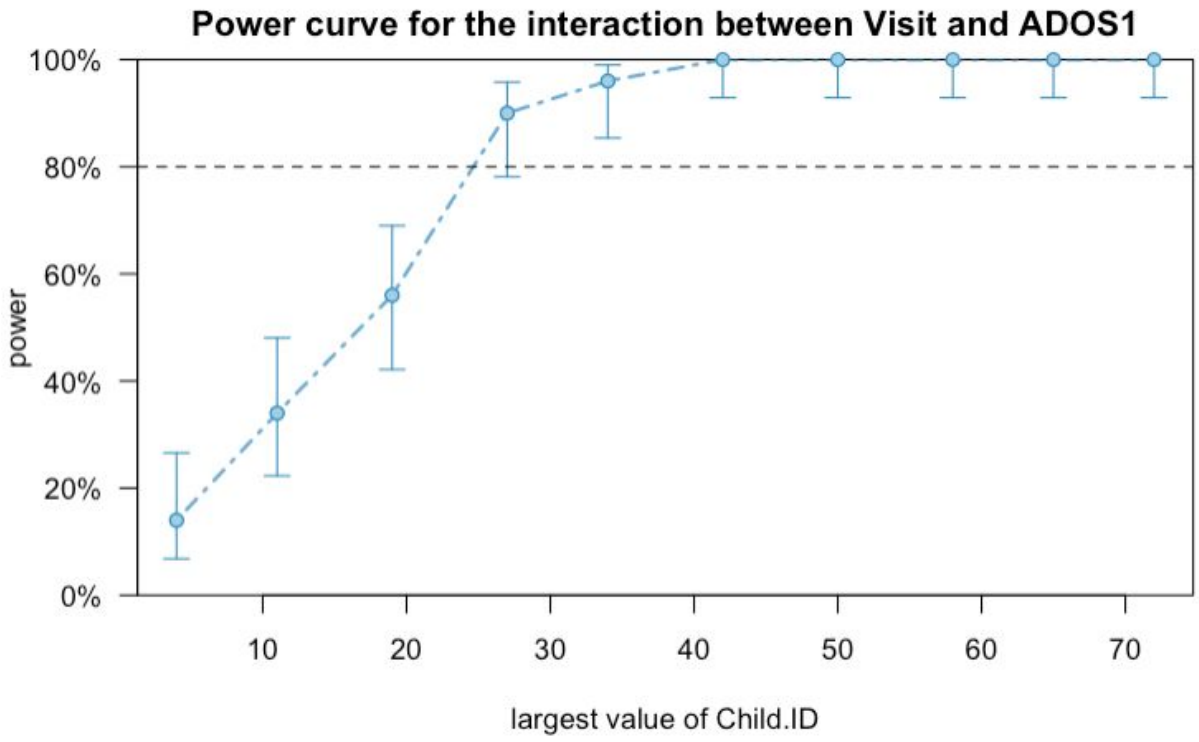
The risk of setting a too low minimum effect size is that our power gets too high considering the available resources. We would rather risk this than risk having too low power since this is a common problem in psychological research and leads to false negatives.

Assess the power curve by Child.ID, identifying an ideal number of participants to estimate each effect

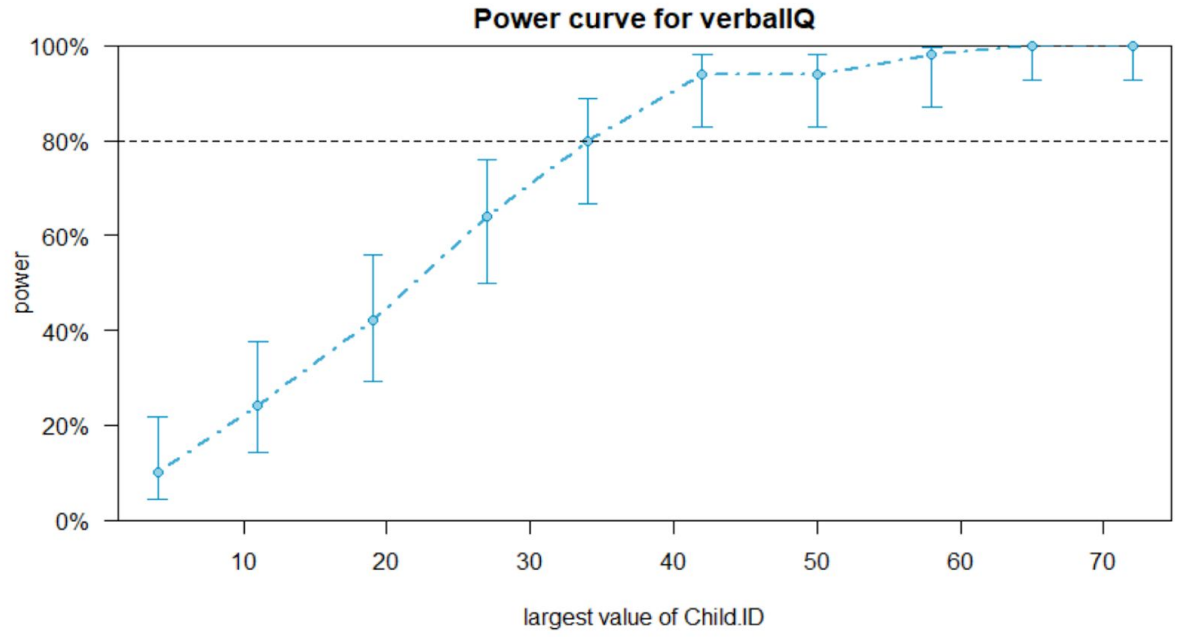
See the power curves below. These have been completed with 50 data simulations, expect for visit^2, here we ran 100 data simulations because the power was close to 80% (with a confidence interval that was both under and over 80%) for 50 simulations. As to be more confident in our judgment we ran a higher number of simulations.

The following number of participants are needed to find our minimum effect sizes on MLU for each effect with a power of 80%:

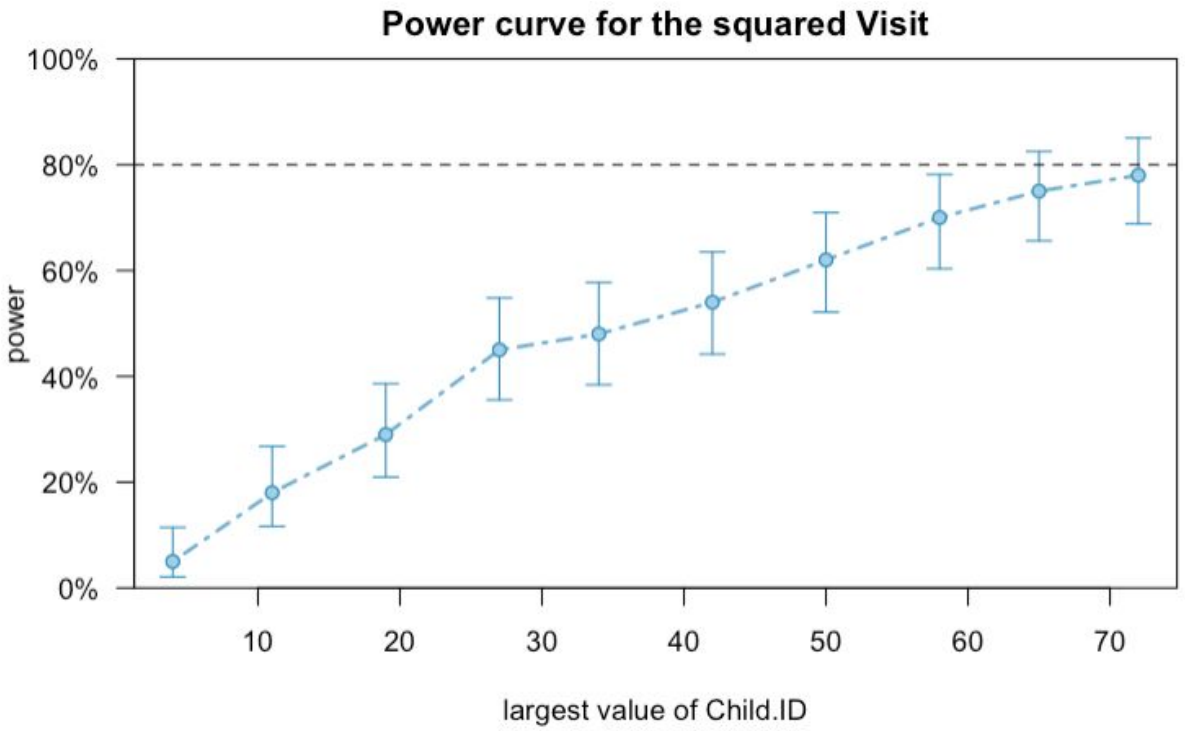
Interaction between visit nr. and ADOS: ca. 25 participants



Verbal IQ: ca. 35 participants



Visit^2: More than 70 participants



Report the power analysis and comment on what you can (or cannot) use its estimates for.

The power for the interaction and for the verbal IQ is unchanged.

Visit^2 has a power of 82% (ci: 73.05-88.97). This is still an acceptable power, with a more conservative effect size.

We can use our estimates just as well as with the previous study while accepting a lower effect size. This means that if there is a difference, but it is smaller than what has been found in the study (by Riccardo), we are still likely to detect it at this power.

If we only have access to 30 participants. Identify the power for each relevant effect and discuss whether it's worth to run the study and why

We're still using our more conservative minimum effect sizes.

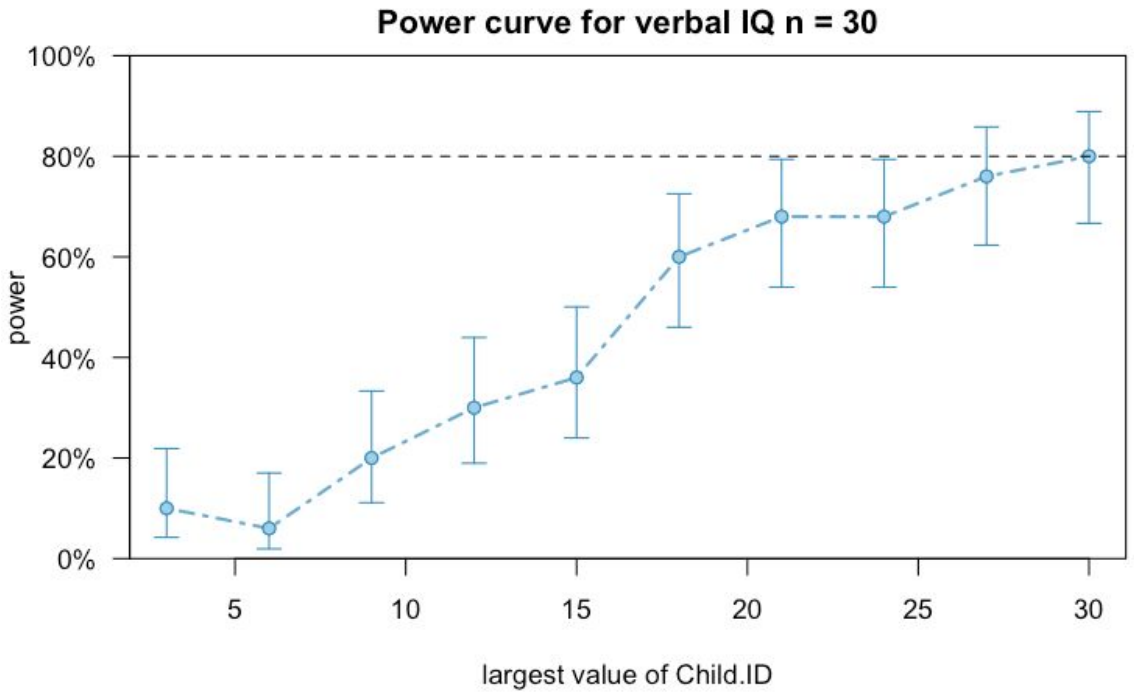
We wanted to avoid selecting a biased group from our data using a subset where we specifically select i.e. the first 30 participants. It could be there is some effect of the order the participants are in.

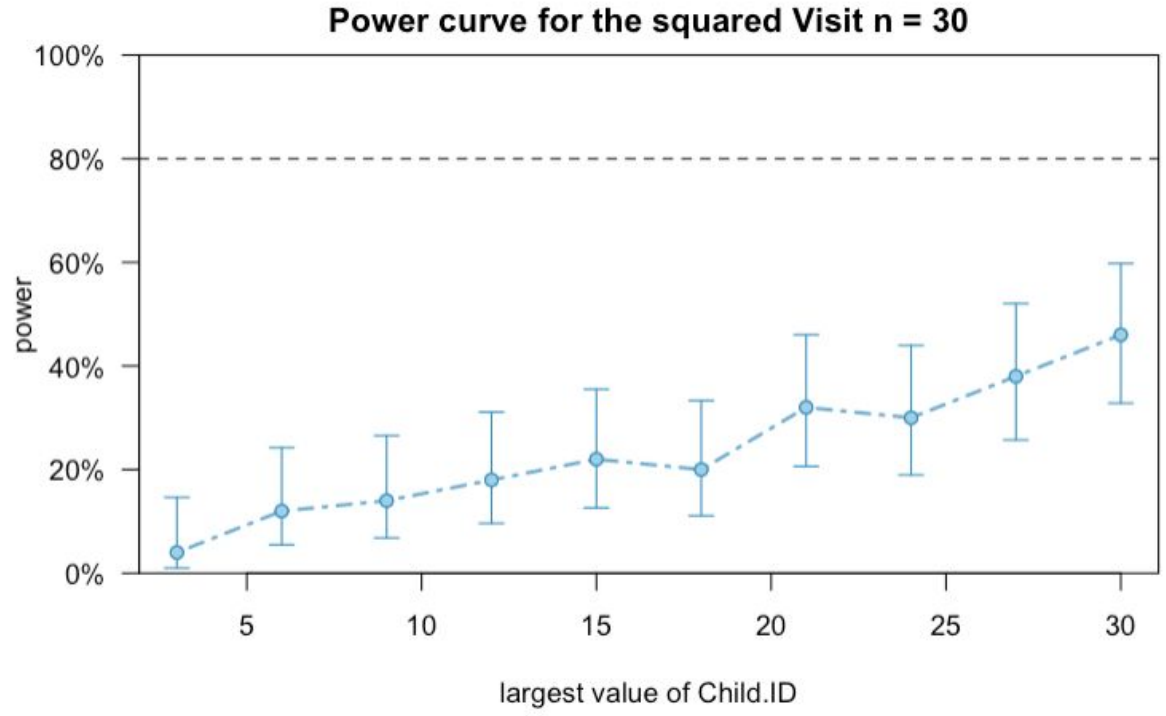
Therefore, we chose instead to simulate a new group of participants using extend(), based on the data. This group may not consist of exactly 15 ASD and 15 TD kids.

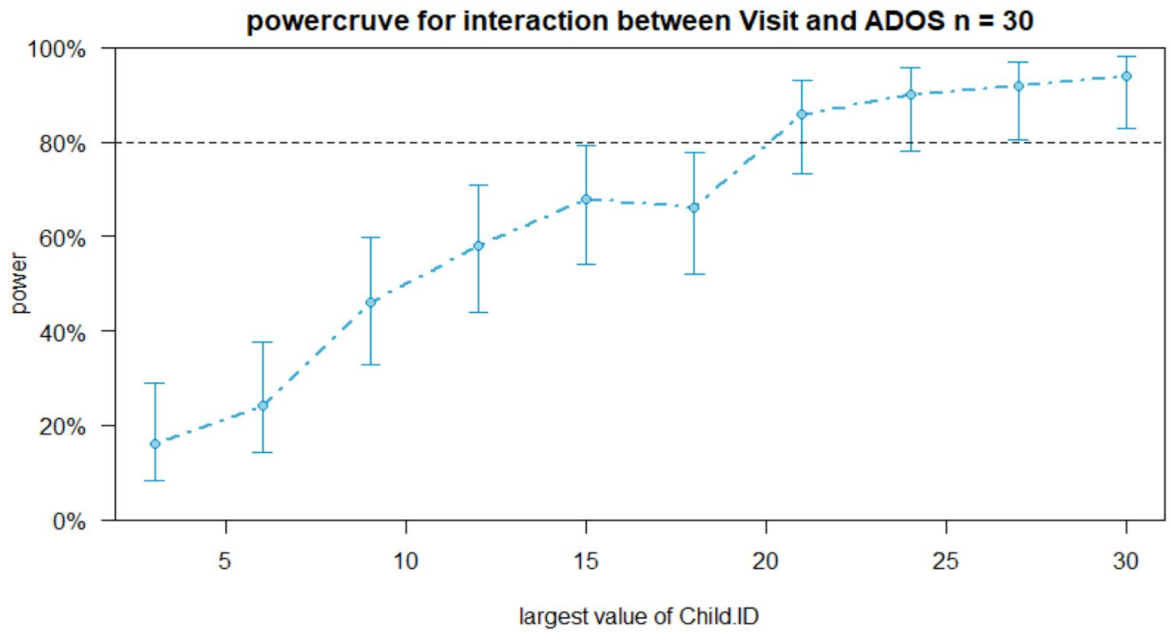
The powers were as follows:

interaction 98% (89.35 - 99.95)
visit^2 28% (16.23 - 42.49)
verbalIQ 86% (73.26 - 94.18)

Power curves on next page.


Power curve for verbal IQ n = 30
100%
80%
60%
40%
20%
0%
power
5
10
15
20
25
30
largest value of Child.ID


Power curve for the squared Visit n = 30
100%
80%
60%
40%
20%
0%
power
5
10
15
20
25
30
largest value of Child.ID



Whether it is worth running this study depends on the hypothesis we are interested in. For the interaction and the verbal IQ, the power is high enough to be worth it. But the visit^2 power was too low to be worth it. It is too likely that we would have a false negative.

Who did what?

We completed the portfolio while working all together in the same room. All problems and approaches were discussed amongst the group. Sometimes we split up i.e. googling-tasks for fixing code-specific problems or looking up theoretical issues. Most tasks were complete on all computers. We split up some of the computational heavy runs, to speed up the process. Daniel ran most of the simulations. The code we linked to is mostly based on Jesper's code and commented by Jesper. The report is mostly written by Astrid.