Decision for round #3 : Revision needed
Invitation to revise Stage 1 RR
Thanks for the revised version of your Stage 1 manuscript. Before I can issue IPA, and write a recommendation, there are some details that I need to clarify about your plan. This may require further minor changes to the manuscript.

1) Your 'sample size calculation is based on BFDA (so should probably not be labelled a 'power analysis'). Although you specify the sample size you are aiming for, you only state the total sample size, and not what the size of your two groups will need to be (CP-knowers and subset-knowers). Are there constraints related to participants per group?

2) In the text and design table, you state that you will perform 'two-tailed Bayesian paired t-tests... with factors of Age and CP-status'. I do not understand how you propose to have more than one factor in a t-test. Can you clarify?

3) Your text (and grouping of sub-hypotheses) seems to suggest that BOTH Hypothesis 1a and 1b need to confirmed in order to confirm Hypothesis 1, and similarly that BOTH H2a and H2b need to be confirmed in order to confirm H2. Is this correct? If so, please state explicitly.

4) In your design table, you state: "Moderate BF10 of difference in the left parietal region in CP-knowers compared to subset-knowers will be taken as strong evidence and the anecdotal BF10 of difference in the left parietal region in CP-knowers will be taken as weak evidence.". This seems wrong. Surely BF10 ≥ 10 would be strong evidence, whilst BF10 3-10 would be moderate evidence?

Thanks for further clarifications.

Rob

Dear Prof. McIntosh,
Thank you for the opportunity to resubmit our revised manuscript. We appreciate the time and effort you have invested in providing us with constructive feedback. We have incorporated your suggestions and highlighted them as tracked changes in the manuscript. Please find below our point-by-point responses.

1) Your 'sample size calculation is based on BFDA (so should probably not be labelled a 'power analysis'). Although you specify the sample size you are aiming for, you only state the total sample size, and not what the size of your two groups will need to be (CP-knowers and subset-knowers). Are there constraints related to participants per group?

Following your suggestion, we have renamed the subsection on page 8 from 'Power analysis' to 'Bayes factor design analysis' to avoid further confusion. As for the size of the two groups, we expect the groups to be the same. Please find the changes on page 10.

2) In the text and design table, you state that you will perform 'two-tailed Bayesian paired t-tests... with factors of Age and CP-status'. I do not understand how you propose to have more than one factor in a t-test. Can you clarify?

We apologise for the ambiguity. Following further discussions in response to your comment, we updated the analysis to further reflect our main hypotheses. Instead of paired t-tests, we will run independent sample t-tests. Please find the changes on pages 20-22, 26-27.

We would like to thank you for raising this point as it led us to further literature search for a more straightforward calculation of the sample. Accordingly, we adjusted the BFDA calculation, which was previously based on the paired t-test. For this calculation, we are relying on the web-based BFDA application (http://shinyapps.org/apps/BFDA/), written on a shinyapps platform that utilises R for app development. The web-based BFDA application is developed by Stefan et al. (2019) (https://doi.org/10.3758/s13428-018-01189-8), the same authors as the BFDA package in R. While the methodology behind both tools is the same, the web-based BFDA application provides better outcome analytics (e.g., clearer plot, median of the Distribution of N values) and a more user-friendly design (much faster computation that does not require high-performance computers, no coding skills required) to ensure accuracy in the calculation. Although the effect size remains the same (d=0.35), we have decided to apply informed priors t(μ = 0.35, r = 0.102, df = 3), following the tutorial by Stefan et al. (2019). As we target Cortex journal, we have decreased the expected BF_{10} from 10 to 6. With these calculations, the median of the distribution (a parameter that can provide an accurate idea for the stopping point in a sequential design, according to Stefan et al. (2019)) resulted in 46 participants per group (total of 92 children). Therefore, we will continue data collection until we either reach at least BF of 6 in favour of H1 or H0 or collect 46 subset-knowers and 46 CP-knowers. As testing with 20 participants per group has been shown to highly likely cause misleading evidence (see Stefan et al., 2019; Schönbrodt et al., 2017), we will start consecutive testing for the hypotheses once there are 25 participants per group.

The changes have been reflected on pages 9-10, 20, 26-27 of the manuscript. For convenience, we have included the PDF of the results from the web-based BFDA application reports with this response.

3) Your text (and grouping of sub-hypotheses) seems to suggest that BOTH Hypothesis 1a and 1b need to confirmed in order to confirm Hypothesis 1, and similarly that BOTH H2a and H2b need to be confirmed in order to confirm H2. Is this correct? If so, please state explicitly.

We apologise for the unclear statements. It is correct that both sub-hypotheses should be confirmed in order to confirm both Hypotheses 1 and 2, however, as we explained in the legend of figure 1 on page 8 of the manuscript, the differences for the number word 'four' are not expected to be as strong: "Please note that in both hypotheses, we expect smaller differences between CP-knowers and subset-knowers for the number word 'four' than for the number word 'eight'. While both CP- and subset-knowers understand the number word 'four' and are expected to have a strong parietal response, CP-knowers are more advanced, so their parietal response will be stronger as compared to subset-knowers. However, a larger group difference for the number word 'eight' is expected because subset-knowers do not understand the number word 'eight' yet and are thus not expected to have a strong parietal response. CP-knowers, however, have a much better understanding of the number word 'eight',

and are expected to have a strong parietal response". Therefore, Hypotheses 1 and 2 are expected to be mainly driven by strong evidence for the Hypotheses 1a and 2a (related to number 'eight'), while Hypotheses 1b and 2b (related to number 'four') will probably provide a weaker evidence (in comparison to 1a and 2a). We added this clarification in the revised manuscript on page 21.

4) In your design table, you state: "Moderate BF10 of difference in the left parietal region in CP-knowers compared to subset-knowers will be taken as strong evidence and the anecdotal BF10 of difference in the left parietal region in CP-knowers will be taken as weak evidence.". This seems wrong. Surely BF10 ≥ 10 would be strong evidence, whilst BF10 3-10 would be moderate evidence?

We apologise for the incorrect wording. The text was revised and now reads as follows: "BF10 3-10 of difference in the left parietal region in CP-knowers compared to subset-knowers will be taken as moderate evidence and the BF10 1-3 of difference in the left parietal region in CP-knowers compared to subset-knowers will be taken as anecdotal evidence". We added this clarification in the revised manuscript on pages 27-28.

This is a dynamic report of your Bayes Factor Design Analysis for the sequential design. The analysis was conducted on 2024-06-26 00:10:40.

Selected parameters

Prior on Effect Size:

- Default Prior: Cauchy$(\mu = 0, r = \sqrt{2}/2)$
- Informed Prior: t$(\mu = 0.35, r = 0.102, df = 3)$

Data Generating Process: ES = 0.35

Decision Boundaries:

- Lower Boundary: 0.17
- Upper Boundary: 6

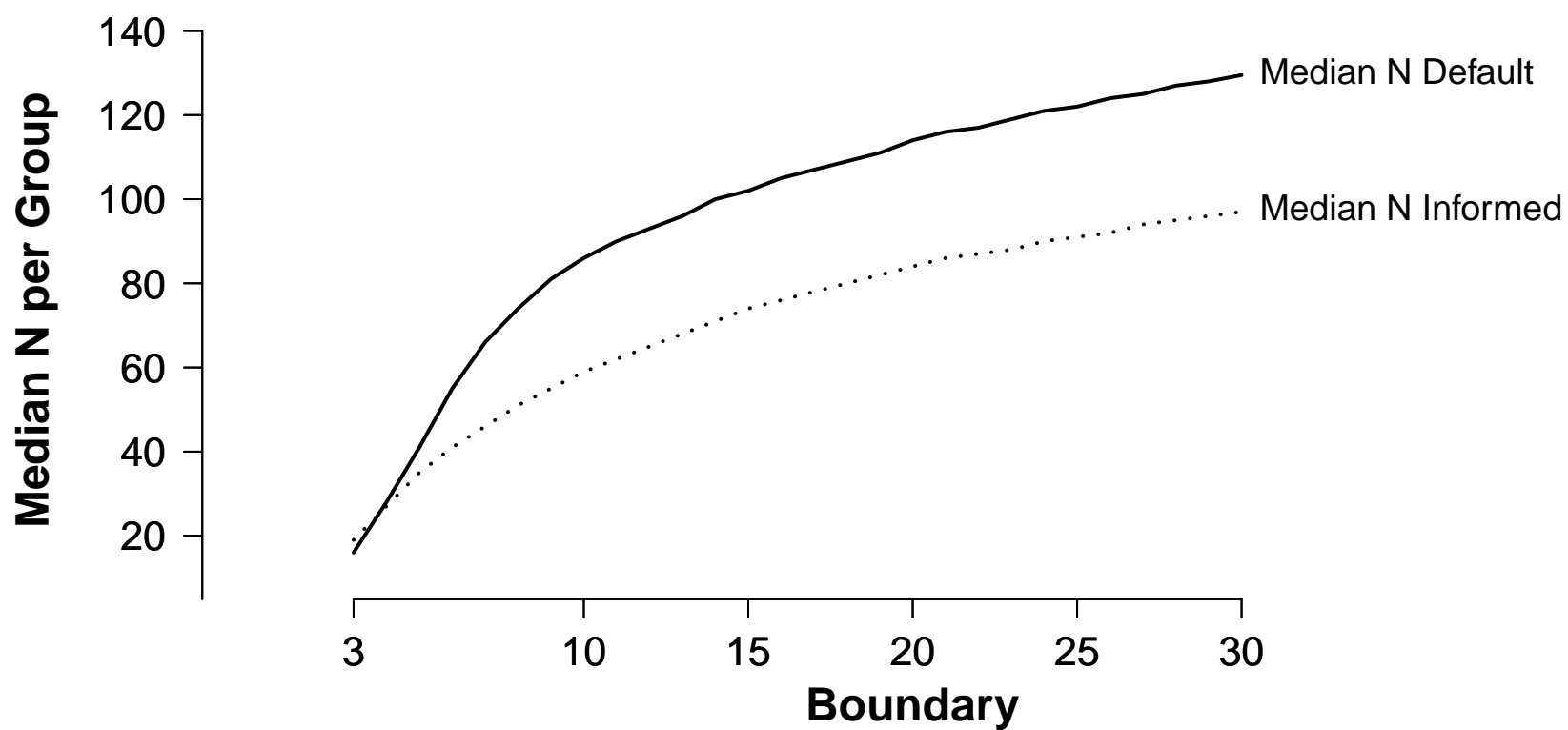### Expected N per group (Data Generating Process: δ = 0.35)

Table 1: Median of the Distribution of N

	Default	Informed
DGP: H1	55	41
DGP: H0	43	46

This is a dynamic report of your Bayes Factor Design Analysis for the sequential design. The analysis was conducted on 2024-06-26 00:10:40.

Selected parameters

Prior on Effect Size:

- Default Prior: Cauchy($\mu = 0$, $r = \sqrt{2}/2$)
- Informed Prior: t($\mu = 0.35$, $r = 0.102$, $df = 3$)

Data Generating Process: ES = 0.35

Decision Boundaries:

- Lower Boundary: 0.17
- Upper Boundary: 6
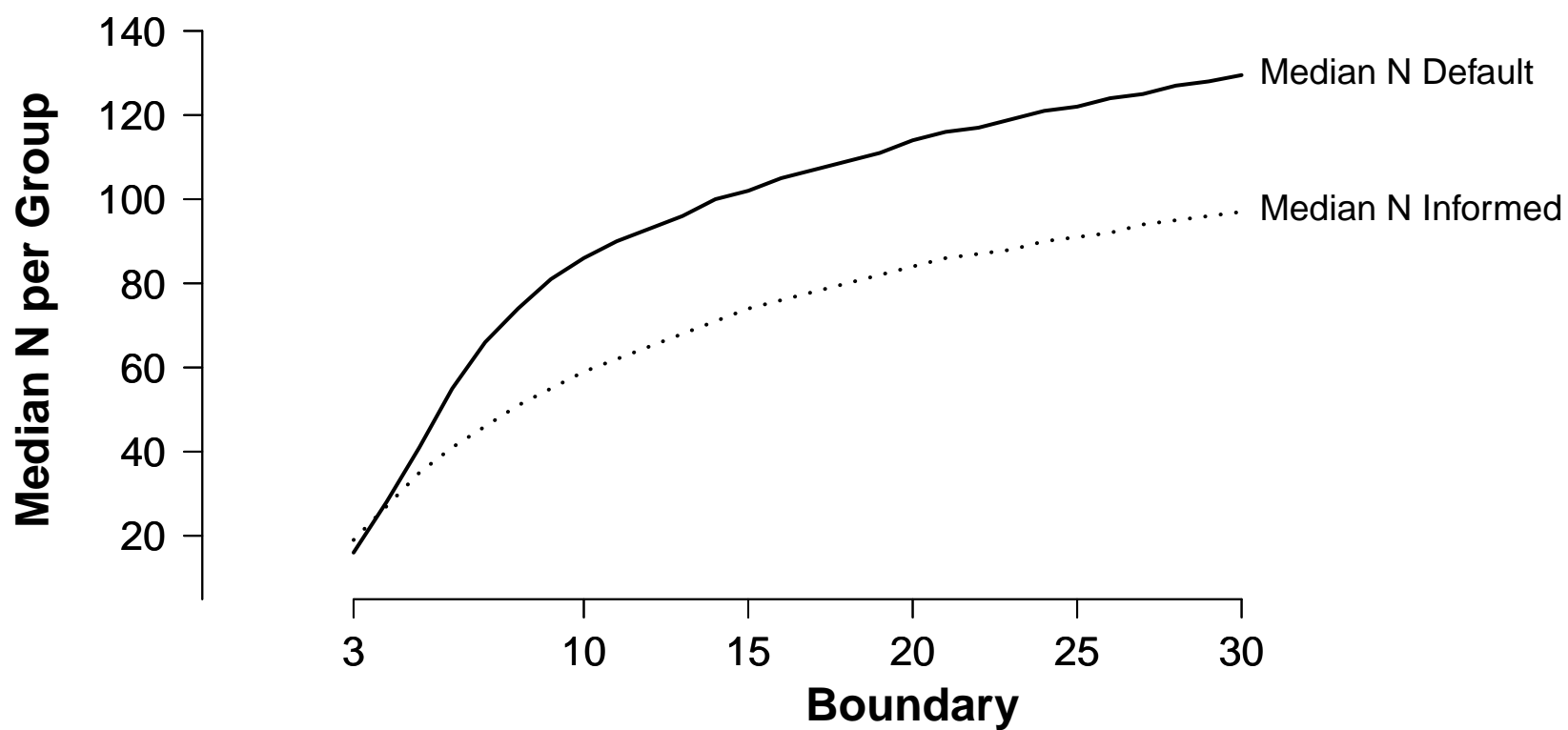
Expected N per group (Data Generating Process: δ = 0.35)

Table 1: Median of the Distribution of N

	Default	Informed
DGP: H1	55	41
DGP: H0	43	46