How important is trial sample size in neuroimaging experiments?

Gang Chen[1], **Daniel Pine**[2], **Melissa Brotman**[3], **Ashley R. Smith**[2], **Paul Taylor**[1], **Robert Cox**[1], **Simone Haller**[3]

1. Scientific and Statistical Computing Core, National Institute of Mental Health, NIH, USA
2. Section on Development and Affective Neuroscience, National Institute of Mental Health, USA
3. Section on Functional Imaging Methods, NIMH, National Institutes of Health, USA

Correspondence: gangchen@mail.nih.gov

Current status of trial sample size in neuroimaging

- Inconvenient parameter when designing an experiment
  - Little attention; chosen for convenience: usually in the range of [10, 40] per condition
  - Implicit assumption: of little importance compared to subject sample size
- Questions
  - Is the lack of attention to trial sample size justifiable?
  - How to quantify the role of trial sample size?
  - How to effectively leverage trial sample size to gain statistical efficiency?
  - How is the trial sample size related to the summary-statistics modeling approach?

Conventional power analysis

- Largely focusing on the number of subjects
  - Tools: fMRIPower; Neurodesign
  - Rarely used; mainly to pacify reviewers
- Difficulties of power analysis
  - Poor practice: effect magnitude usually not reported in literature
  - Not one, but many regions involved
  - Complexity: multiple testing adjustment
- Justifications for Big Data initiatives
  - Small effects
  - Large heterogeneity across subjects
  - Straightforward: more subjects always improve statistical efficiency
  - Only and most cost-effective approach to increase statistical efficiency?

Overview

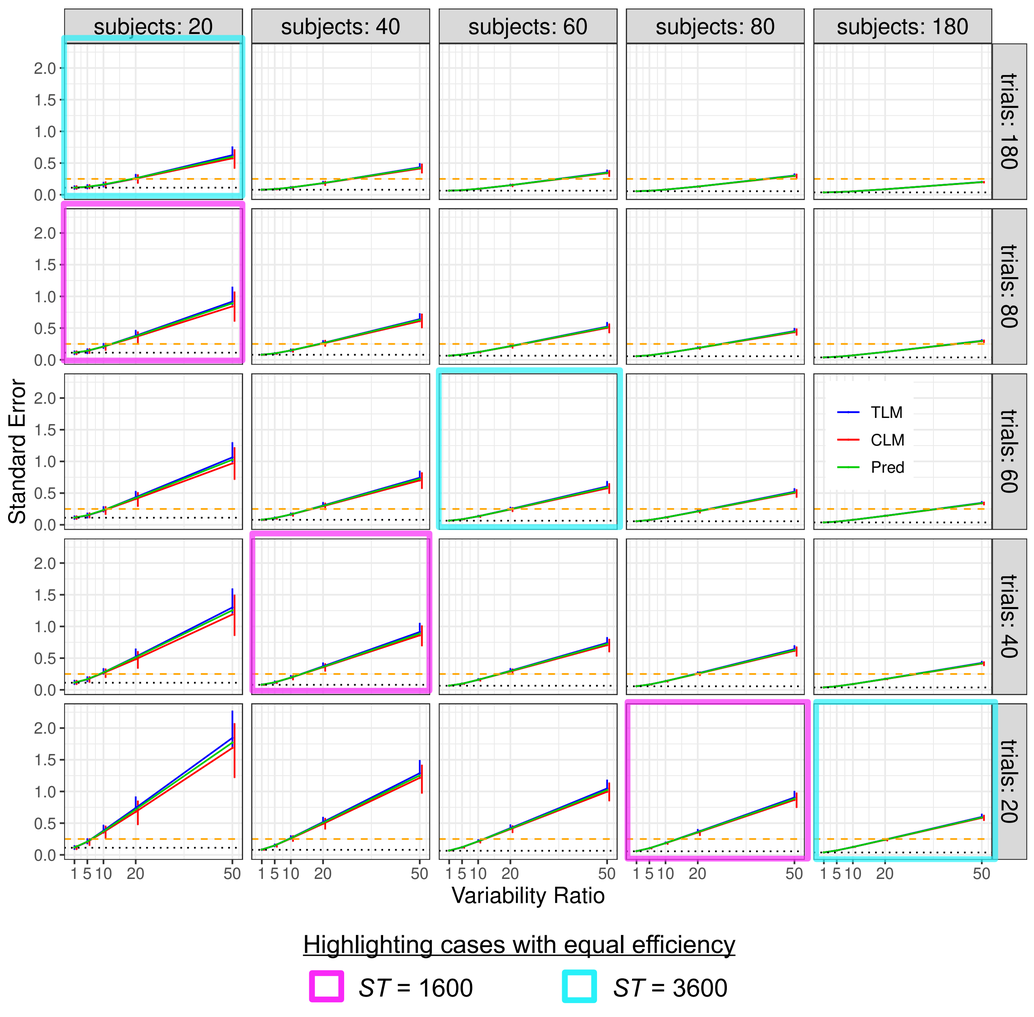
- Adopt a hierarchical model to accurately map the data across multiple levels
- Derive theoretical relationship between trial and subject sample size
- Use simulations to gain insights
- Validate simulation results through an experimental dataset
- Make recommendation for future experimental designs

Theoretical framework

- Data y_{cst}: trial-level effect with c, s and t indexing conditions, subjects and trials
- Model formulation

$$y_{cst} \sim \mathcal{N}(\mu_c + \pi_{cs},\ \sigma_\tau^2);$$
$$\begin{bmatrix}\pi_{1s}\\ \pi_{2s}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma_{\pi_1}^2 & \rho\sigma_{\pi_1}\sigma_{\pi_2}\\ \rho\sigma_{\pi_1}\sigma_{\pi_2} & \sigma_{\pi_1}^2\end{bmatrix}\right);\ c = 1, 2;\ \ s = 1, 2, \ldots, S;\ \ t = 1, 2, \ldots, T.$$

- Model parameters
  - μ_c: cth condition effect
  - π_{cs}: sth subject's effect under cth condition
  - $\sigma_{\pi_c}^2$: cross-subject variances under cth condition
  - σ_τ^2: cross-trial variances
  - ρ: correlation between the two conditions
- Variance for effect of interest (contrast between two conditions)
  - $\sigma^2 = 2(1-\rho)\frac{\sigma_\pi^2}{S} + \frac{2\sigma_\tau^2}{ST}$
  - Revealing relationship between the two sample sizes S and T: hyperbolic
  - Variability ratio $R_v = \sigma_\tau/\sigma_\pi$ - indicator for relative magnitude of cross-trial variability



- Insights from simulations
  - Subject sample size always plays a crucial role in statistical efficiency
  - Impact of trial sample size: varying
    - negligible: small cross-trial variability (eg, $R_v \lesssim 1$): implicitly assumed in practice
    - as important as subjects: large cross-trial variability (eg, $R_v \gtrsim 10$): symmetric hyperbola
    - moderate but less impactful than subjects when cross-trial variability is moderate
- Crucial question: what is the typical magnitude of cross-trial variability in real data?

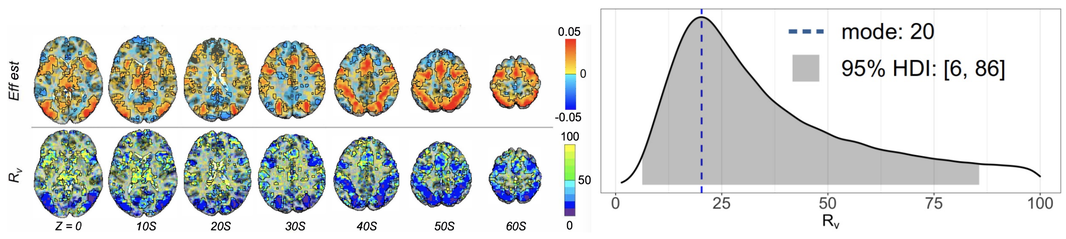
Validation of hyperbolic relationship with experimental dataset

Data information

- 42 subjects
- Eriksen Flanker task with two trial types: congruent and incongruent
- 8 runs, ~380 trials per condition per subject (350 ± 36 incongruent, 412 ± 19 congruent trials)
- 4 trial sample sizes
  - 12.5%: ≈ 48 trials from the first run
  - 25%: ≈ 95 trials from two runs
  - 50%: ≈ 190 trials from four runs
  - 100%: ≈ 380 trials from full dataset
- 2 modeling approaches
  - Hierarchical framework with trial-level modeling (TLM)
  - Conventional summary statistics approach through condition-level modeling (CLM)
- Questions to be addressed
  - Relative importance of trial sample size compared to subjects
  - Relative magnitude of cross-trial variability: variability ratio R_v
  - Performance of conventional summary statistics approach

Cross-trial variability R_v

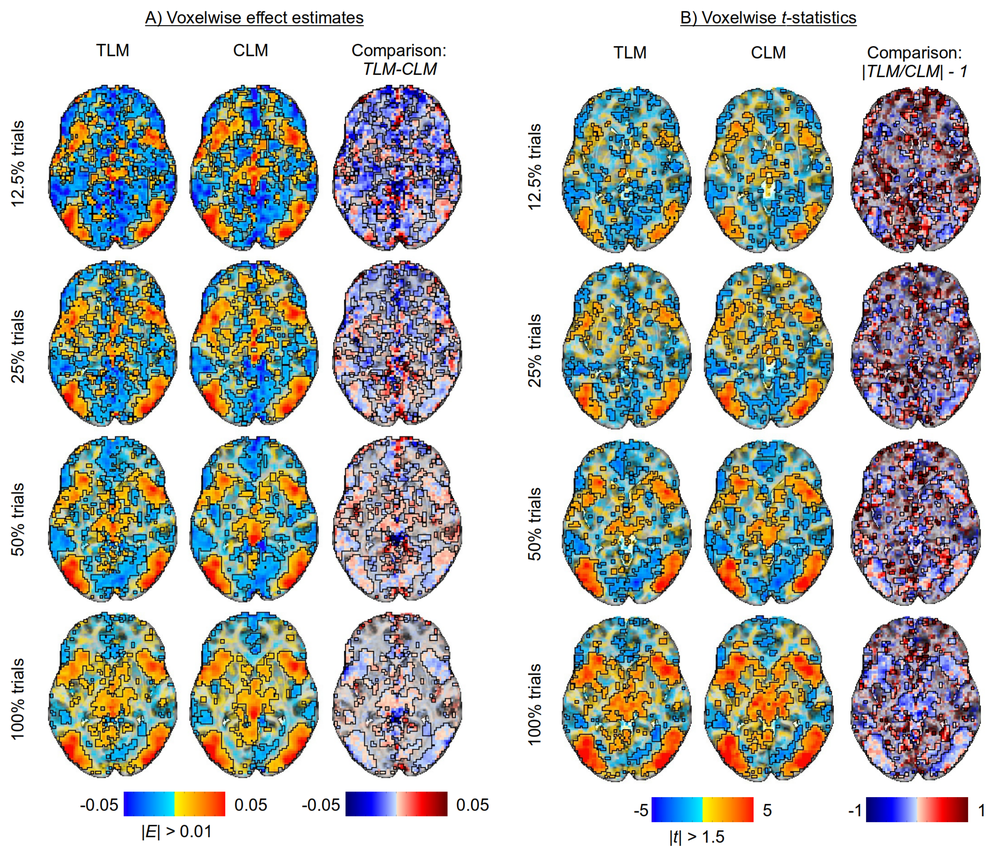
- Typical range of cross-trial variability: $R_v > 5$



Impact of trial sample size

- Hyperbolically symmetric: With moderate to large R_v, trials almost as important as subjects
- More cost-effective to increase both trials and subjects
- TLM vs CLM: quite similar statistical results in most regions

Revelations from experimental data



Conclusions

- Trial sample size: nearly similar impact as subject sample size on statistical efficiency
- More effectively to increase both trials and subjects simultaneously
- Trials can be leveraged alongside subjects to improve cost-effectiveness
- Trial-level modeling is preferred especially when trial sample size is small

Acknowledgments

The research was supported by the NIMH & NINDS Intramural Research Programs of the NIH.

References

Chen et al, 2022. Hyperbolic trade-off: The importance of balancing trial and subject sample sizes in neuroimaging. NeuroImage 247, 118786.