# Simplified power and sample size calculations using prevalence & magnitude of active peaks.

Joke Durnez[a], Beatrijs Moerkerke[a], Ruth Seurinck[a] and Thomas Nichols[b]

[a] Department of Data Analysis, Ghent University, Belgium

[b] Department of Statistics & Warwick Manufacturing Group, University of Warwick, United Kingdom

UNIVERSITEIT GENT

## Introduction

- There is increasing concern about statistical power in neuroscience research:
  - Low power decreases the chance of detecting a true effect
  - Low power reduces the chance that a statistically significant result indicates a true effect (Ioannidis, 2005).
- ⇒ **A power analysis is a critical component of any study**
- Power analyses for fMRI are difficult: need to specify magnitude, spatial extent and location of a hypothesized effect.
- We present a simple way to characterize the spatial signal in a fMRI study, and a direct way to estimate power based on an existing pilot study.
- We must estimate (or have prior knowledge of):
  - the volume of the brain that is activated
  - the average effect size in activated brain regions
- This allows the calculation of power for given sample size, brain volume and smoothness.

## Methods

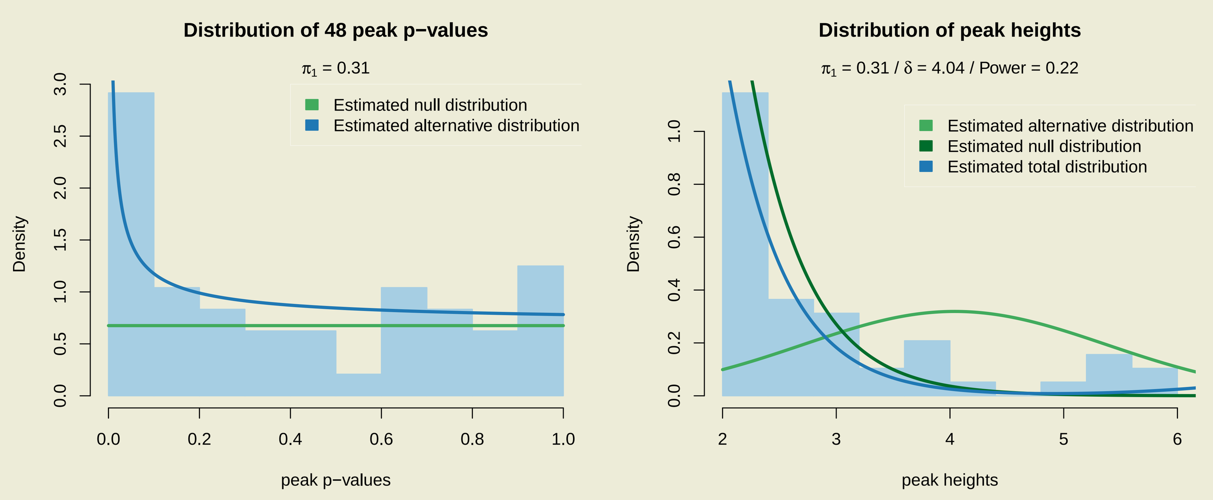
- We start from peak $p$-values in a group level analysis. We estimate $\pi_1$, proportion of peak $p$-values thare are non-null, as described in Durnez, Moerkerke & Nichols (2014).
- We assume that the null distribution of peak values is an exponential distribution (Worsley, 2007).
- We assume that the alternative distribution of peak values is a truncated normal distribution (truncated at excursion threshold $u$).
- Therefore, the distribution of peak values can be written as a mixture:

$$f(x|\pi_0,\mu_1,\sigma_1,u) = (1-\pi_1)u\exp(-u(x-u)) + \pi_1\frac{\frac{1}{\sigma_1}\varphi\left(\frac{x-\mu_1}{\sigma_1}\right)}{1-\Phi\left(\frac{u-\mu_1}{\sigma_1}\right)}$$

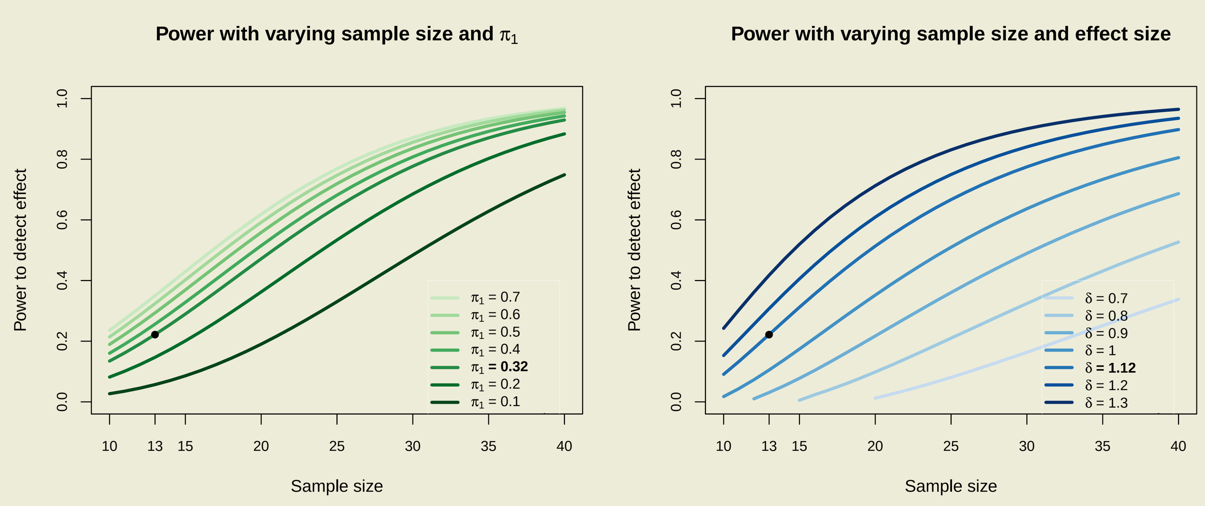
- $\mu_1$ and $\sigma_1$ can be estimated using maximum likelihood, where $\mu_1$ is the expected peak height in activated regions.
- Power can be estimated for a given threshold$t$ as $P(T > t|H_a)$ with $T$ the $T$-statistic of the peak.

## Example

- Data from Seurinck et al. (2011). The original study was analyzed voxelwise with FWER-control at the 5% significance level. We reanalyzed the data peakwise with FDR-control at the 5% significance level.
- Sample size: 13
- Step 1: compute prevalence of activation
- Step 2: estimate truncated distributions

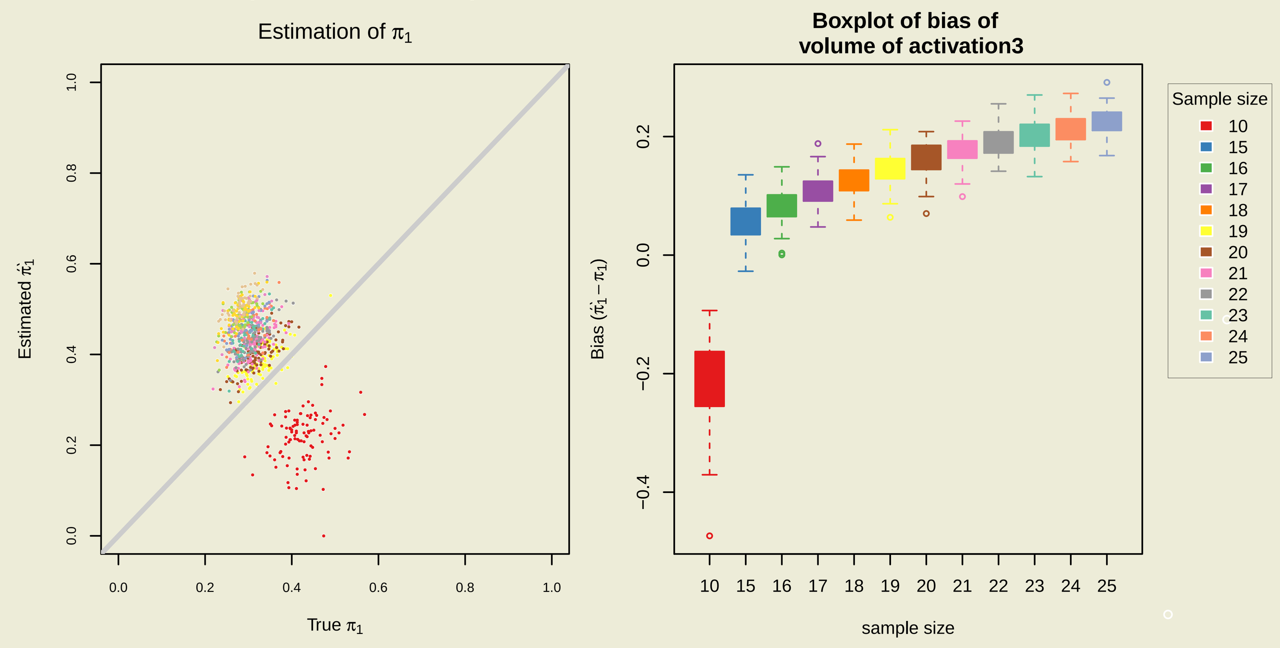



- Expected peak height under activation $\mu_1 = \delta = 4.04$
- Effect size $\mu/\sigma = \delta/\sqrt{n} = 1.13$
- Step 3: estimate power for given sample size, estimated peak height under activation and prevalence of activation ($\pi_1$)

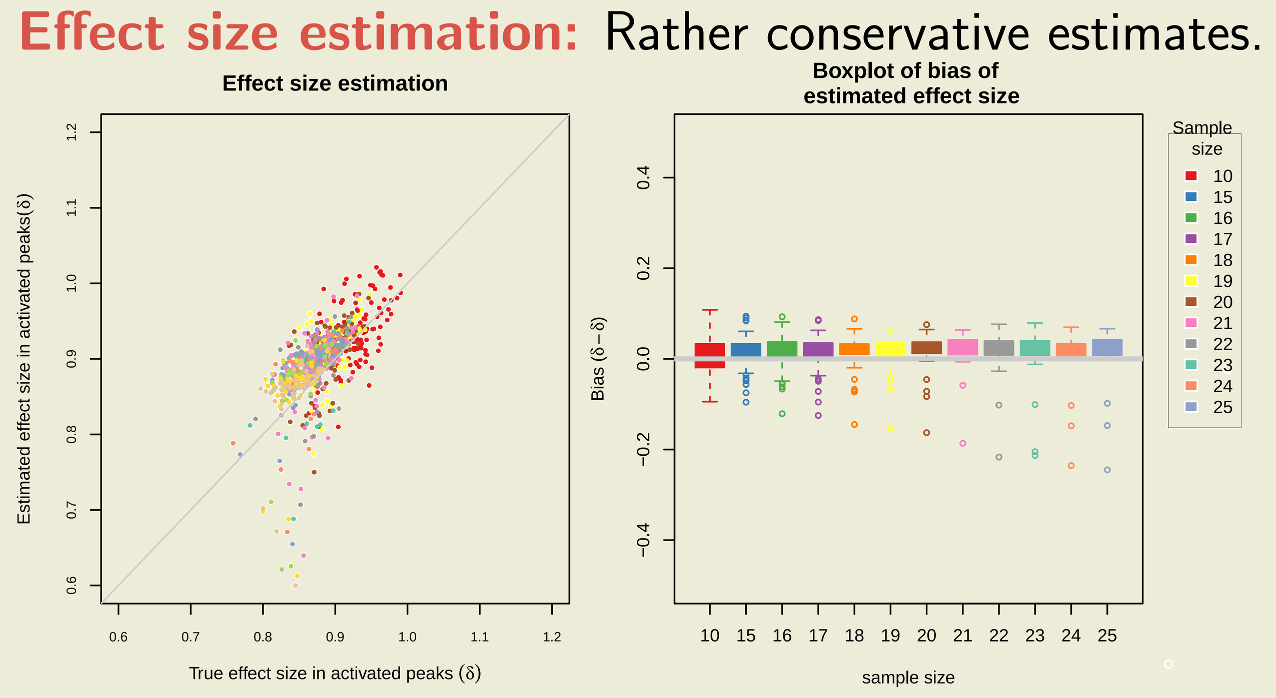



## Simulations

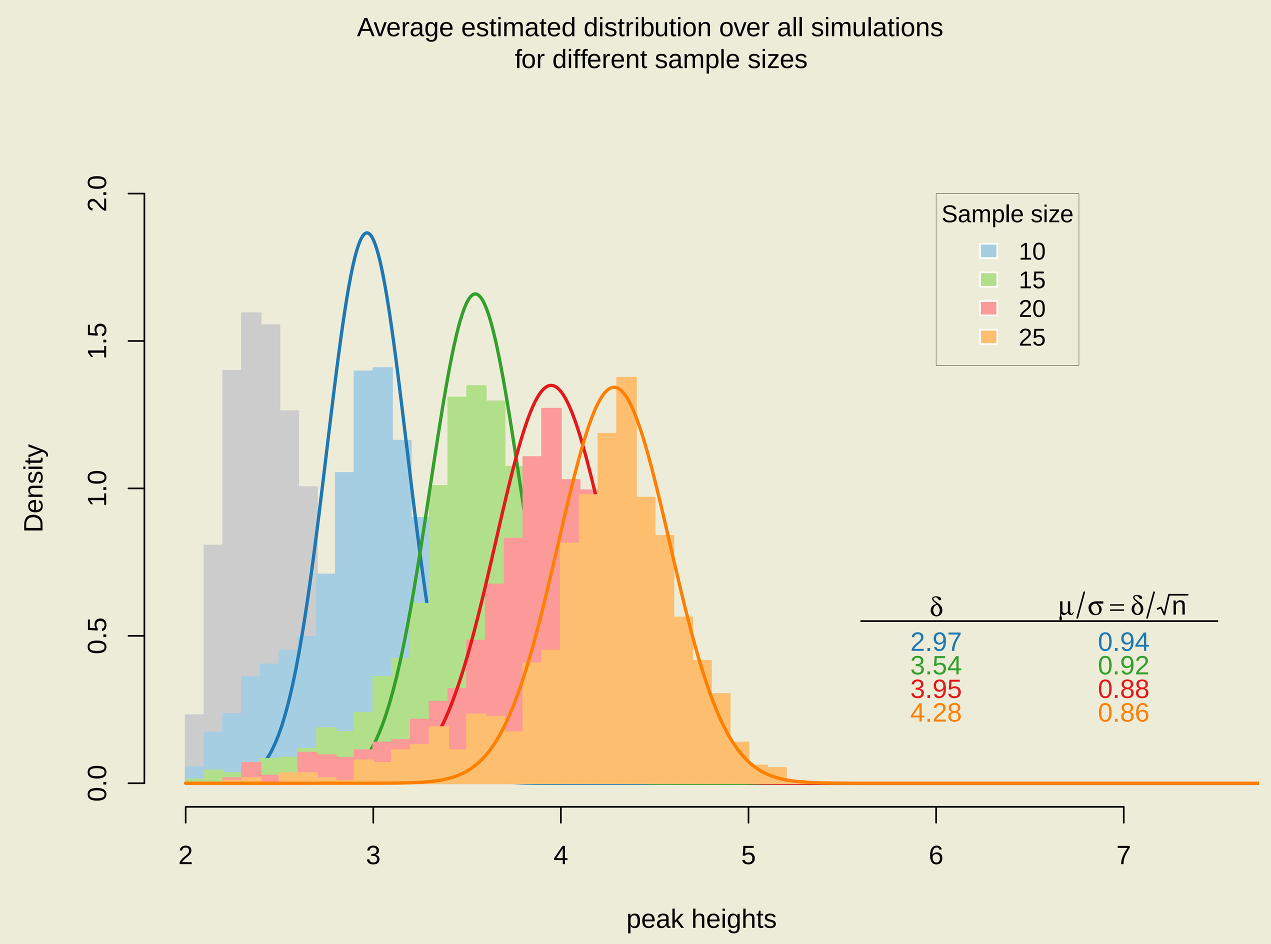
- 100 full-brain datasets with smooth Gaussian noise (3 voxels) superimposed with activation (3 % BOLD change) in 4 foci (3% of total brain volume)
- **Prevalence of activation:** For small sample sizes: very conservative estimates (close to 0).



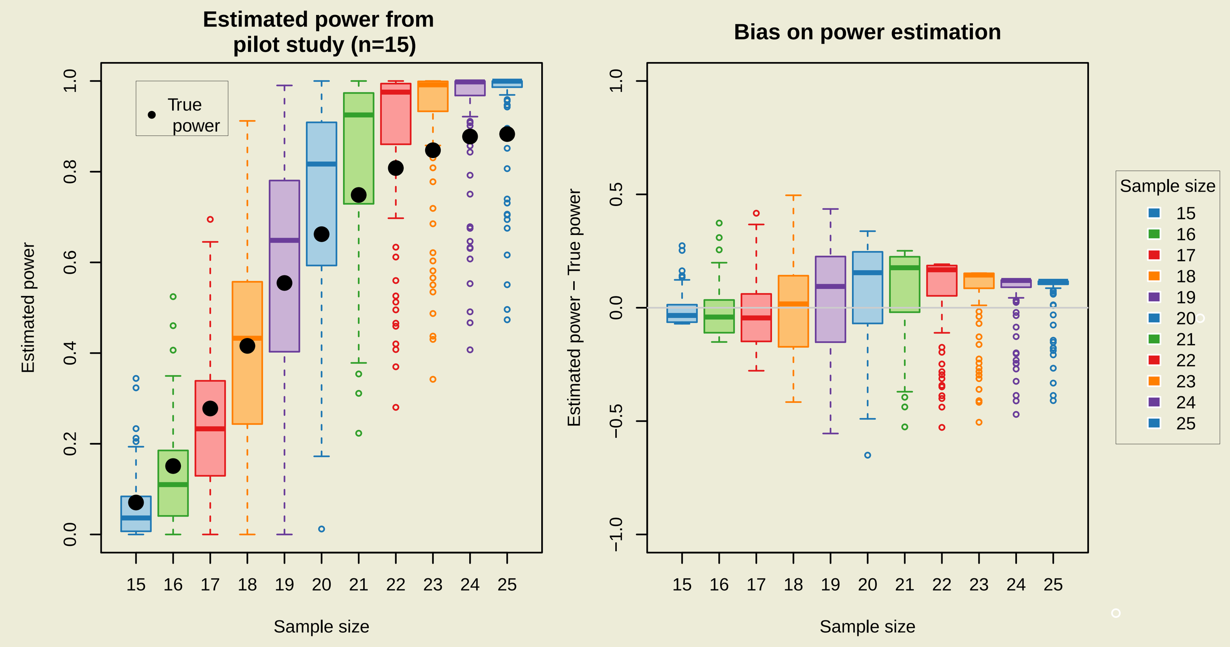

- **Effect size estimation:** Rather conservative estimates.




- **Alternative distribution:** Constant effect size ($\mu/\sigma$) over different sample sizes to be used in power analyses.




- **Power calculations:** Good estimates for realistic values of power with FWER control at the 5% significance level.




## References and acknowledgements

Durnez, Moerkerke, Nichols (2014). NeuroImage, 84.

Ioannidis (2005). PLOS Medicine, 2:6.

Seurinck, de Lange, Achten, Vingerhoets (2011). Journal of Cognitive Neuroscience, 23:6.

Worsley (2006). Random Field Theory in Friston, Ashburner, Kiebel, Nichols and Penny, Statistical Parametric Mapping. Elsevier, London

**This work was carried out using the STEVIN Supercomputer Infrastructure at Ghent University, funded by Ghent University, the Flemish Supercomputer Center (VSC), the Hercules Foundation and the Flemish Government â department EWI.**