Estimation of Statistical Power in a Multicentre MRI study.

Benson Mwangi Irungu

PhD Student,Centre for Neurosciences
Ninewells Hospital and Medical School
University of Dundee.

IPEM Conference
Experiences and Optimisation of Multicentre MRI Research Studies.
Edinburgh Training and Conference Centre.
March 26, 2010.

Why Multicentre Studies?

- A reasonable method of recruiting sufficient subjects especially if the target disorder is particularly rare.

Why Multicentre Studies?

- A reasonable method of recruiting sufficient subjects especially if the target disorder is particularly rare.
- A study protocol requiring a strict inclusion criteria.

Why Multicentre Studies?

- A reasonable method of recruiting sufficient subjects especially if the target disorder is particularly rare.
- A study protocol requiring a strict inclusion criteria.

Statistical power definition:

Why Multicentre Studies?

- A reasonable method of recruiting sufficient subjects especially if the target disorder is particularly rare.
- A study protocol requiring a strict inclusion criteria.

Statistical power definition:

- Power is the probability of rejecting the null hypothesis when it is false.

Problems Associated with Power Calculation In MR Imaging.

- Pooling data from multiple centres introduces scanner related variances.

Problems Associated with Power Calculation In MR Imaging.

- Pooling data from multiple centres introduces scanner related variances.
- Statistical power is bound to change depending on which region of the brain is being analyzed e.g. regions of activation in fMRI or structural differences in sMRI.

Problems Associated with Power Calculation In MR Imaging.

- Pooling data from multiple centres introduces scanner related variances.
- Statistical power is bound to change depending on which region of the brain is being analyzed e.g. regions of activation in fMRI or structural differences in sMRI.
- This makes MRI power analysis rather more difficult than in a typical behavioural measure.

Demographics:

- MRI data acquired in 2 centres (Centre I- Aberdeen , Centre II- Edinburgh).
- All patients had been diagnosed with DSM IV Unipolar major depression.
- Controls were matched based on average age, male/female ratio and estimated IQ.
- Images acquired using 1.5 T General Electric MRI scanners in both centres.

Table: Demographic information

Variable(Mean/SD)	Patients#I	Controls#I	Patients#II	Controls#II	Significance (P-Value)
Age (Yrs)	46.07	40.59	44.66	42.99	p=0.597
Females/Total	9/15	11/18	10/15	7/14	p=0.411
NART	111.6	114.33	115.67	117.69	p=0.566
BDI	22.93	3.15	38	1.14	p<0.001
SP	54.6	30.38	62	28.79	p<0.001
SH	35	51.36	32.13	52.71	p<0.001
Hamilton	23.2	-	27.87	-	p=0.19

Image Processing:

Image processing performed using SPM5.

Image Processing:

Image processing performed using SPM5.
Spatial Normalisation

- Matching the MR volumes into an apriori template using affine transformations followed by non-linear deformations.

Image Processing:

Image processing performed using SPM5.
Spatial Normalisation

- Matching the MR volumes into an apriori template using affine transformations followed by non-linear deformations.

Modulation

- Rescaling the normalized MR volumes by the amount of contraction that may have been introduced during normalization.

Image Processing:

Image processing performed using SPM5.
Spatial Normalisation

- Matching the MR volumes into an apriori template using affine transformations followed by non-linear deformations.

Modulation

- Rescaling the normalized MR volumes by the amount of contraction that may have been introduced during normalization.

Smoothing

- A Gaussian FWHM window of 8-mm used for smoothing.

Voxel-wise Statistical Analysis

An Analysis of Variance (ANOVA) test was performed across the two groups.
Essentially, the following hypothesis were tested:

Voxel-wise Statistical Analysis

An Analysis of Variance (ANOVA) test was performed across the two groups.
Essentially, the following hypothesis were tested:

1. Structural differences between control/patient groups in each centre.

Voxel-wise Statistical Analysis

An Analysis of Variance (ANOVA) test was performed across the two groups.
Essentially, the following hypothesis were tested:

1. Structural differences between control/patient groups in each centre.
2. Scanner related differences- differences between controls from both groups.

Voxel-wise Statistical Analysis

An Analysis of Variance (ANOVA) test was performed across the two groups.
Essentially, the following hypothesis were tested:

1. Structural differences between control/patient groups in each centre.
2. Scanner related differences- differences between controls from both groups.
3. Overall structural differences between controls and patients pooled from both centres.

Voxel-wise Statistical Analysis

Factorial Design:

Design Matrix

	control I	patient I	control II	patient II
control I	1	0	0	0
patient I	0	1	0	0
control II	0	0	1	0
patient II	0	0	0	1

Voxel-wise Statistical Analysis

Factorial Design:

Design Matrix

	control I	patient I	control II	patient II
control I	1	0	0	0
patient I	0	1	0	0
control II	0	0	1	0
patient II	0	0	0	1

Contrasts of Interest:

Voxel-wise Statistical Analysis

Factorial Design:

Design Matrix

	control I	patient I	control II	patient II
control I	1	0	0	0
patient I	0	1	0	0
control II	0	0	1	0
patient II	0	0	0	1

Contrasts of Interest:

- Centre I- [1 -1 0 0], controls>patients
- Centre I- [-1 1 0 0], patients > controls
- Centre II-[0 0 1 -1], controls >patients
- Centre II-[0 0 -1 1], patients > controls

Voxel-wise Statistical Analysis

Factorial Design:

Design Matrix

	control I	patient I	control II	patient II
control I	1	0	0	0
patient I	0	1	0	0
control II	0	0	1	0
patient II	0	0	0	1

Contrasts of Interest:

- Centre I- [1 -1 0 0], controls>patients
- Centre I- [-1 1 0 0], patients > controls
- Centre II-[0 0 1 -1], controls >patients
- Centre II-[0 0 -1 1], patients > controls
  Threshold of significance:
- $p<0.05$ corrected at a whole brain cluster level using a Monte Carlo method (Slotnick et al)

Structural and scanner Differences-Results

Centre I: Controls vs Patients

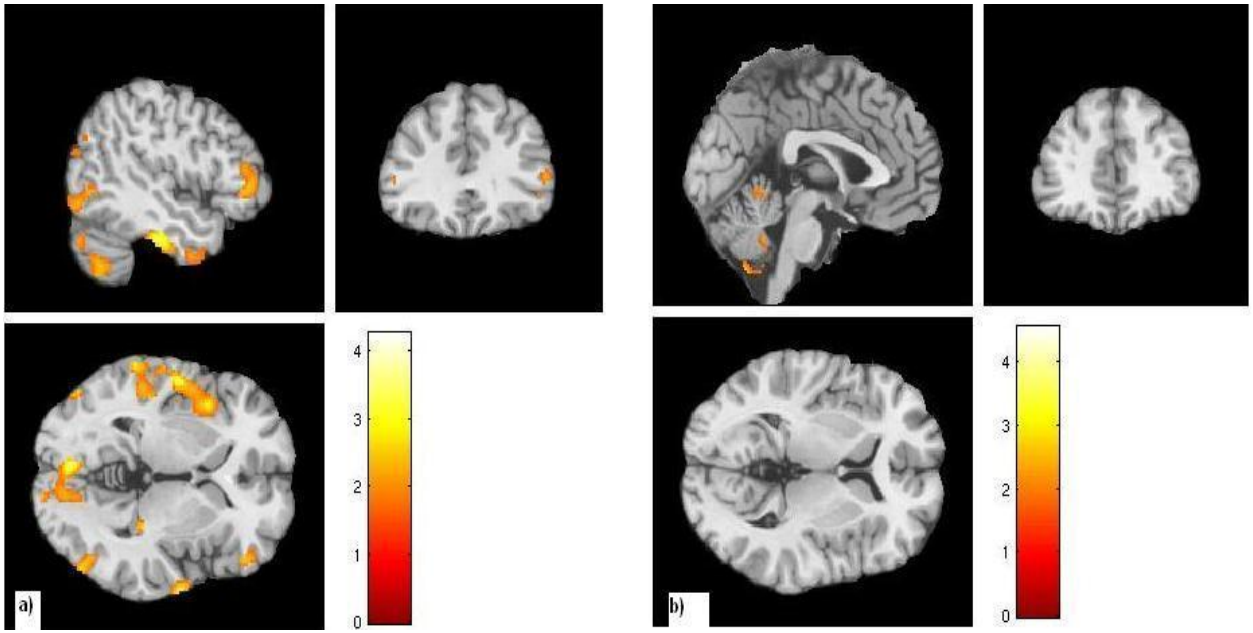


Figure: Structural differences between patients and controls a) Patients had reduced grey matter density, b) no evidence of increase in grey matter density in the patient group.

Structural and scanner Differences-Results

Centre II: Controls vs Patients

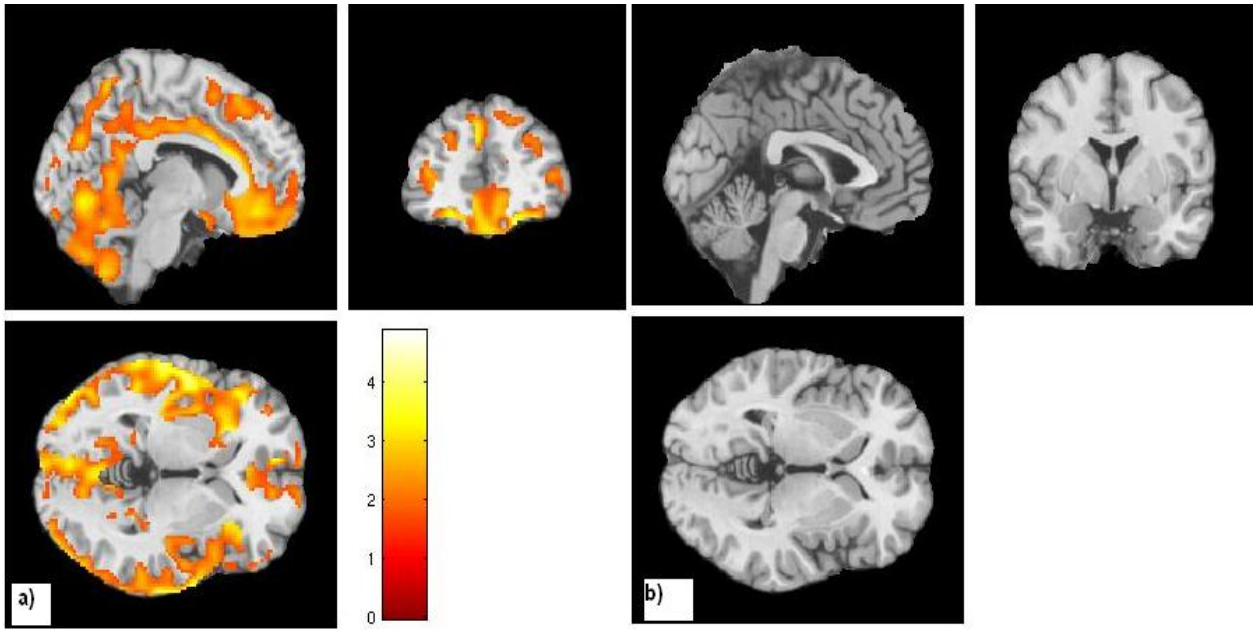


Figure: Structural differences between patients and controls a) Patients had reduced grey matter density, b) no evidence of increase in grey matter density in the patient group.

Structural and scanner Differences-Results

Scanner Differences:Centre I>Centre II.

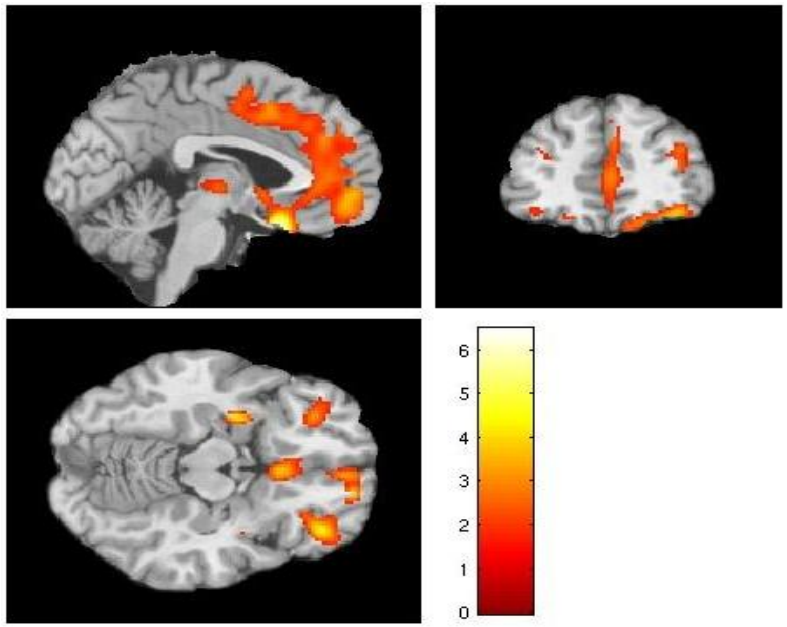


Figure: Scanner differences between scanner I and scanner II and whilst it might be a subject effect we think it's more likely a scanner effect.

Data Pooling:

Do Z values go up when we pool data and more importantly does the statistical power increase?

Data Pooling:

Do Z values go up when we pool data and more importantly does the statistical power increase?

Factors which would influence the power outcome include:

Data Pooling:

Do Z values go up when we pool data and more importantly does the statistical power increase?

Factors which would influence the power outcome include:

1. Sample size

Data Pooling:

Do Z values go up when we pool data and more importantly does the statistical power increase?

Factors which would influence the power outcome include:

1. Sample size
2. Effect size

Data Pooling:

Do Z values go up when we pool data and more importantly does the statistical power increase?

Factors which would influence the power outcome include:

1. Sample size
2. Effect size
3. Type I error e.g. $\alpha = 0.05$- (The odds that the observed result is due to chance)

Data Pooling

Do Z values go up?

Table: Significant cluster peaks : Centre I subjects

Region	Brodmann Area	MNI Coordinates	Z-value
Superior temporal gyrus	22	-64,-14,6	3.04
Precentrall gyrus	6	-50,-2,8	2.99
Superior temporal gyrus	38	-48,8,-14	3.20
Inferior frontal gyrus	45	-50,22,14	3.08
Middle frontal gyrus	47	48,38,0	2.32

Data Pooling

Do Z values go up?

Table: Significant cluster peaks : Centre I subjects

Region	Brodmann Area	MNI Coordinates	Z-value
Superior temporal gyrus	22	-64,-14,6	3.04
Precentrall gyrus	6	-50,-2,8	2.99
Superior temporal gyrus	38	-48,8,-14	3.20
Inferior frontal gyrus	45	-50,22,14	3.08
Middle frontal gyrus	47	48,38,0	2.32

Table: Significant cluster peaks: Centre II subjects

Region	Brodmann Area	MNI Coordinates	Z-value
Middle temporal gyrus	21	64,-4,-6	4.44
Inferior frontal gyrus	11	-20,28,-24	4.28
Superior temporal gyrus	38	48,12,-14	3.66
Superior temporal gyrus	38	-56,4,-6	3.84
Superior frontal gyrus	9	-8,50,34	3.94

Data Pooling

Do Z values go up?-Pooled data:

Table: Significant cluster peaks: Pooled data

Region	Brodmann Area	MNI Coordinates	Z-value
Superior temporal gyrus	22	-58,-6,0	4.69
Middle temporal gyrus	21	64,-4,-4	4.49
Superior temporal gyrus	22	-54,0,-4	4.34
Middle Occipital gyrus	19	-50,-60,-8	4.42
Inferior Occipital gyrus	18	-38,-86,-12	3.95

Data Pooling

Do Z values go up?-Pooled data:

Table: Significant cluster peaks: Pooled data

Region	Brodmann Area	MNI Coordinates	Z-value
Superior temporal gyrus	22	-58,-6,0	4.69
Middle temporal gyrus	21	64,-4,-4	4.49
Superior temporal gyrus	22	-54,0,-4	4.34
Middle Occipital gyrus	19	-50,-60,-8	4.42
Inferior Occipital gyrus	18	-38,-86,-12	3.95

- Clearly, by pooling Centre I and Centre II data the Z values did go up!

Data Pooling

Do Z values go up?-Pooled data:

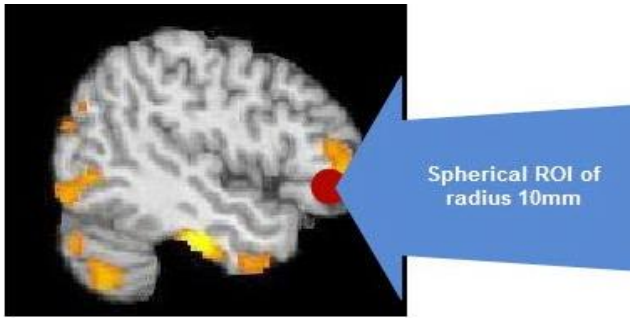
Table: Significant cluster peaks: Pooled data

Region	Brodmann Area	MNI Coordinates	Z-value
Superior temporal gyrus	22	-58,-6,0	4.69
Middle temporal gyrus	21	64,-4,-4	4.49
Superior temporal gyrus	22	-54,0,-4	4.34
Middle Occipital gyrus	19	-50,-60,-8	4.42
Inferior Occipital gyrus	18	-38,-86,-12	3.95

- Clearly, by pooling Centre I and Centre II data the Z values did go up!
- Increased power is suggested by higher Z values.

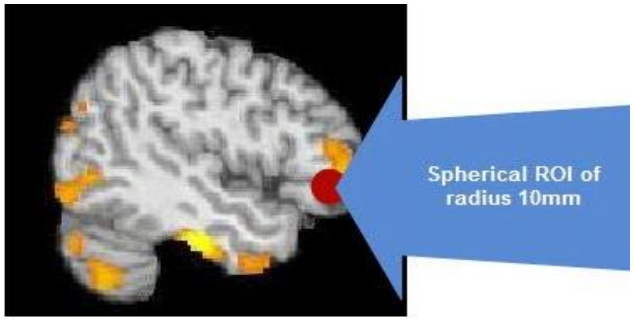
Effect Size and Power calculation

Steps:



Effect Size and Power calculation

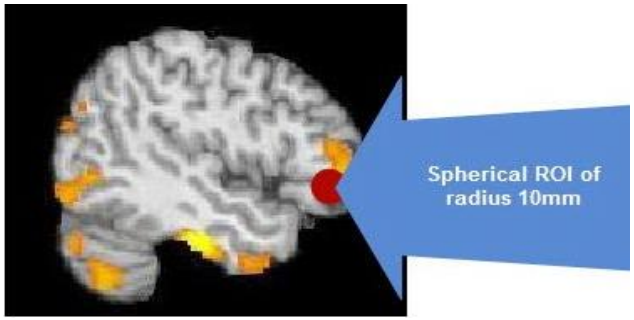
Steps:



1. A spherical region of interest (ROI) is selected by centering around a particular cluster point where significant structural differences were detected.

Effect Size and Power calculation

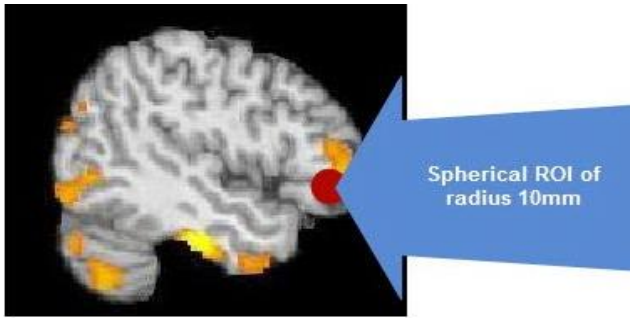
Steps:



1. A spherical region of interest (ROI) is selected by centering around a particular cluster point where significant structural differences were detected.
2. The mean representing all voxels around the ROI is computed.

Effect Size and Power calculation

Steps:



1. A spherical region of interest (ROI) is selected by centering around a particular cluster point where significant structural differences were detected.
2. The mean representing all voxels around the ROI is computed.
3. Computation of effect size and study power for the contrasts of interest using Gpower 3

Effect Size and Power Calculation

Effect Size:

- Effect size $F = \sigma_m/\sigma$ where σ_m is the standard deviation of the effect and σ is the standard deviation within each group.

Effect Size and Power Calculation

Effect Size:

- Effect size $F = \sigma_m/\sigma$ where σ_m is the standard deviation of the effect and σ is the standard deviation within each group.

- The standard deviation of the effect (σ_m) is given by: $$\frac{|C|}{\sqrt{N\sum_{i=1}^{k} c_i^2/n_i}}$$

Effect Size and Power Calculation

Effect Size:

- Effect size $F = \sigma_m/\sigma$ where σ_m is the standard deviation of the effect and σ is the standard deviation within each group.
- The standard deviation of the effect (σ_m) is given by: $$\frac{|C|}{\sqrt{N\sum_{i=1}^{k} c_i^2/n_i}}$$
- $|C| = \sum i = 1 \ldots k \mu_i c_i$; μ_i=mean and c_i is the ANOVA contrast of interest- e.g. [1 -1 0 0].

Effect Size and Power Calculation

Effect Size:

- Effect size $F = \sigma_m/\sigma$ where σ_m is the standard deviation of the effect and σ is the standard deviation within each group.
- The standard deviation of the effect (σ_m) is given by: $$\frac{|C|}{\sqrt{N\sum_{i=1}^{k} c_i^2/n_i}}$$
- $|C|=\sum i = 1...k \mu_i c_i$; μ_i=mean and c_i is the ANOVA contrast of interest- e.g. [1 -1 0 0].
- N = total sample size (e.g. 62), n= sample size in group i (e.g. 4).

Effect Size and Power Calculation

Power Calculation: Gpower3

- Gpower3- A software dedicated to statistical power analysis (Faul, F., et al).

Effect Size and Power Calculation

Power Calculation: Gpower3

- Gpower3- A software dedicated to statistical power analysis (Faul, F., et al).
- Offers five different types of statistical power analysis i.e. apriori, compromise, criterion, **post hoc** and sensitivity.

Effect Size and Power Calculation

Power Calculation: Gpower3

- Gpower3- A software dedicated to statistical power analysis (Faul, F., et al).
- Offers five different types of statistical power analysis i.e. apriori, compromise, criterion, **post hoc** and sensitivity.
- Inputs to the software include: Effect size, Alpha error probability, Total sample size, numerator degrees of freedom and number of groups.

Effect Size and Power Calculation

Power Calculation Results centre I , centre II:

Table: Power calculation results:Centre I

Region	Brodmann Area	ROI centre MNI Coordinates	Effect size	Statistical Power
Middle temporal gyrus	38	-42,4,-44	0.2725	0.5596
Insula	13	-44,14,0	0.2691	0.5393
Superior temporal gyrus	22	64,-14,6	0.3021	0.6477

Table: Power calculation results:Centre II

Region	Brodmann Area	ROI centre MNI Coordinates	Effect size	Statistical Power
Middle temporal gyrus	38	-42,4,-44	0.3435	0.758
Insula	13	-44,14,0	0.2055	0.3563
Superior temporal gyrus	22	64,-14,6	0.2938	0.6236

Effect size and power calculation

Power Calculation Results Pooled Data:

Table: Power calculation results:Pooled data

Region	Brodmann Area	ROI centre MNI Coordinates	Effect size	Statistical Power
Middle temporal gyrus	38	-42,4,-44	0.4348	0.9202
Insula	13	-44,14,0	0.3364	0.7404
Superior temporal gyrus	22	64,-14,6	0.4229	0.9050

As a result:

- Cohen defined effect sizes as *small d=0.2*, *medium d=0.5* and *large d=0.8*.

Effect size and power calculation

Power Calculation Results Pooled Data:

Table: Power calculation results:Pooled data

Region	Brodmann Area	ROI centre MNI Coordinates	Effect size	Statistical Power
Middle temporal gyrus	38	-42,4,-44	0.4348	0.9202
Insula	13	-44,14,0	0.3364	0.7404
Superior temporal gyrus	22	64,-14,6	0.4229	0.9050

As a result:

- Cohen defined effect sizes as *small d=0.2*, *medium d=0.5* and *large d=0.8*.
- The effect sizes for all three planned ANOVA contrasts are therefore in the small to medium range.

Effect size and power calculation

Power Calculation Results Pooled Data:

Table: Power calculation results:Pooled data

Region	Brodmann Area	ROI centre MNI Coordinates	Effect size	Statistical Power
Middle temporal gyrus	38	-42,4,-44	0.4348	0.9202
Insula	13	-44,14,0	0.3364	0.7404
Superior temporal gyrus	22	64,-14,6	0.4229	0.9050

As a result:

- Cohen defined effect sizes as *small d=0.2*, *medium d=0.5* and *large d=0.8*.
- The effect sizes for all three planned ANOVA contrasts are therefore in the small to medium range.
- There is a benefit to pooling data in all regions of interest and the likelihood of rejecting the null hypothesis increases with pooled data.

Conclusion

- Scanner related variances were outweighed by increase in power to detect group differences due to larger sample sizes.

Conclusion

- Scanner related variances were outweighed by increase in power to detect group differences due to larger sample sizes.
- The results depend on the ROI selected.

Conclusion

- Scanner related variances were outweighed by increase in power to detect group differences due to larger sample sizes.
- The results depend on the ROI selected.
- Efforts towards scanner harmonisation should be encouraged.

Conclusion

- Scanner related variances were outweighed by increase in power to detect group differences due to larger sample sizes.
- The results depend on the ROI selected.
- Efforts towards scanner harmonisation should be encouraged.
- Initial results from two of the multicentre based studies indicate that multivariate pattern recognition techniques are not largely affected by scanner related variability kloppel., et al 2008-Alzheimer's disease study ,Mwangi, B., et al, 2010- Unipolar major depression study.

References

- Slotnick, S.D., Schacter, D.L., (2004) A sensory signature that distinguishes true from false memories, Nature Neuroscience 7,6, 664-672.
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175-191.
- Murphy K., Myors., B.,Statistical Power Analysis. Lawrence Erlbaum Associates.2004.

Acknowledgements

- The first author is pursuing a PhD part-financed by the Scottish Imaging Network: A platform for scientific excellence (SINAPSE).

- Many thanks to Prof. Douglas Steele and Prof. Keith Matthews both of the University of Dundee.