

Neuroimaging data analysis

Family-wise error rate and multiple comparison correction

Sunghyon Kyeong, PhD

Severance Biomedical Science Institute,
Yonsei University College of Medicine

Multiple Comparison Correction

Voxel-level correction (under i.i.d. assumption)

- Bonferroni correction: ex) $0.05/5=0.01$
- False Discovery Rates (FDR) correction (or Benjamini-Hochberg procedure)

Cluster-level correction

- Random Field Theory based (**SPM & FSL**)
- Monte Carlo simulations (**ClustSim**)
- Permutation testing (**SnPM**)

The Bonferroni correction is too conservative because most functional imaging data have some **degree of spatial correlation**.

In this case, **there are fewer independent values** in the statistic volume than there are voxels.

SCIENTIFIC METHOD / SCIENCE & EXPLORATION


UK newspaper
Jul 2, 2016

Software faults raise questions about the validity of brain studies

Interpretation of functional MRI data called into question.

by John Timmer (US) - Jul 2, 2016 3:15pm JST

The Register
Biting the hand that feeds IT

 DATA CENTRE SOFTWARE NETWORKS SECURITY INFRASTRUCTURE DEVOPS BUSINESS HARDWARE

Science

fMRI bugs could upend years of research

This is what your brain looks like on bad data

3 Jul 2016 at 23:59, Richard Chirgwin



1240



725

UK newspaper
Jul 3, 2016

MRI 소프트웨어에 버그 ‘연구 결과도...’

윤신철 칼럼니스트 | 2016년 7월 6일

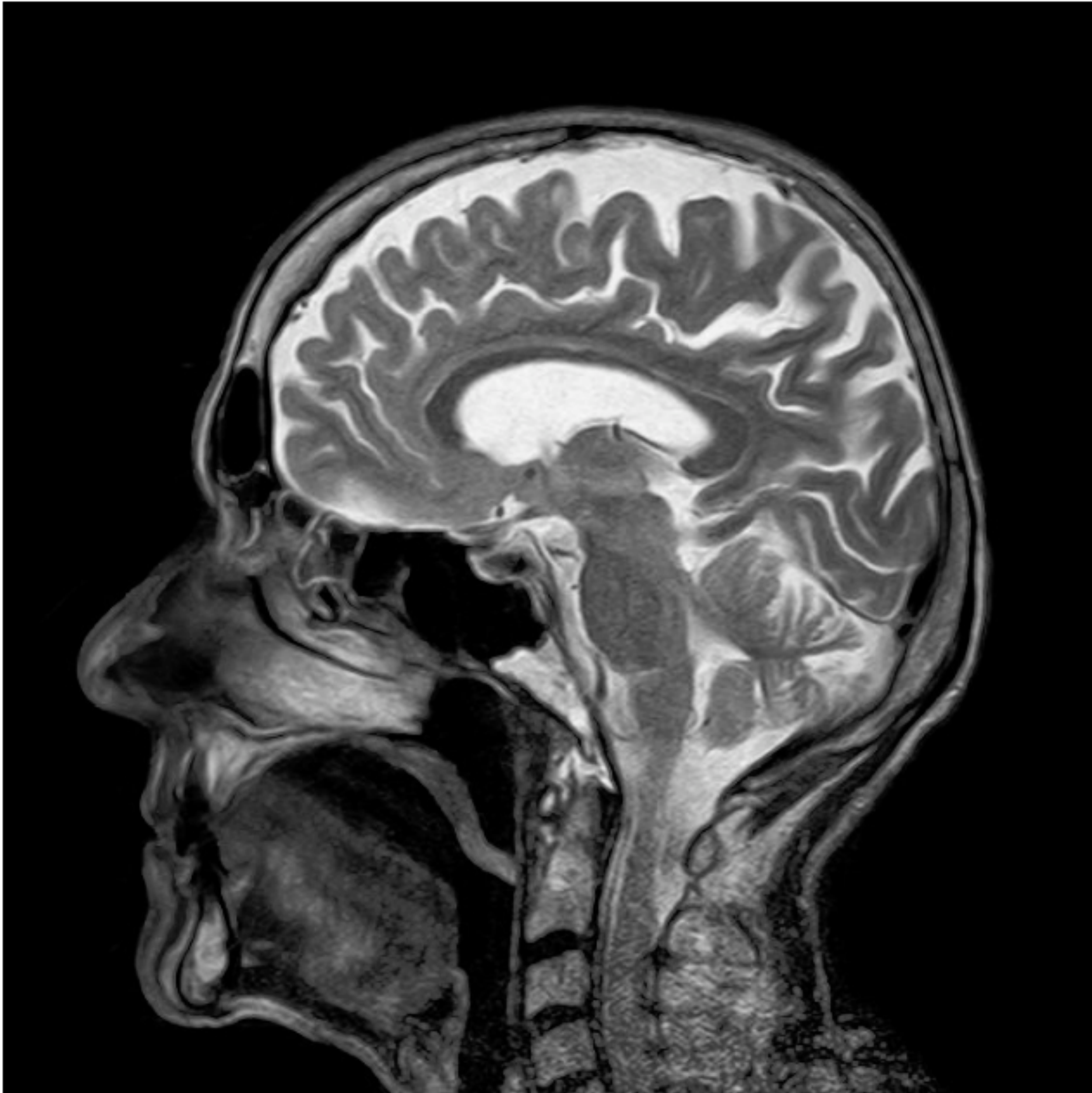
좋아요 166 공유하기 트윗 G+1 0



MRI는 핵 자기공명 현상을 이용해 생체 정보를 이미지로 만들 수 있게 해주는 장치다. CT 검사와 달리 방사선을 이용하지 않기 때문에 방사선에 노출되지 않는다는 장점이 있지만 이런 MRI가 이용하는 소프트웨어에서 버그가 발견되어 4만 건에 달하는 MRI 이미지를 이용한 연구 결과가 무효화될 처지가 됐다고 한다.

기사 링크:
<http://techholic.co.kr/archives/56760>

MRI는 인간의 뇌 단면 이미지를 촬영할 때 자주 쓰인다. 덕분에 MRI를 이용해 지금까지 뇌에 관한 연구에 이용해왔다. 그런데 지난 1년간 MRI 이미지를 이용한 연구 결과가 모두 무효가 될 수 있다는 것.



국제

"뇌질환 fMRI 분석결과, 믿을 수 없다" 의혹제기 논문

기사입력 : 2016-07-07 07:14 | 수정 : 2016-07-08 16:14

바이오펙테이터 김성민 기자





뇌 질환 오진율 70%..4천여 논문 유효성 물음표

국민일보 
www.kmib.co.kr

MRI 검사도 믿을 수 없다 "최대 70%까지 오진 가능성"

입력 2016-07-07 11:30

 Like 0

    인쇄  

(단독)뇌 관찰하는 fMRI 소프트웨어에 치명적 버그...오류 비율 70%

입력: 2016.07.07 11:29 | 수정 2016.07.07 11:29

SNS공유  메일  인쇄  글씨크기  



우리 뇌의 다양한 움직임과 성격을 연구하는데 가장 최적화 된 기구이자 현대 과학의 수준을 끌어올리는데 큰 기여를 한 것으로 인정받는 기능성자기공명영상(fMRI) 장치의 데이터 분석 소프트웨어에 치명적인 '버그'(bug·프로그램의 결함에 의해 오류나 오작동이 일어나는 현상)가 있다는 사실이 15년 만에 밝혀졌다.



박준석 미국 오하이오주립대학 심리학 박사과정

학부와 석사과정에서 심리학을 공부했고 현재 인지심리학 박사과정에 재학중이다. 인간의 마음을 수학적으로 모형화하는 분야인 '인지모델링'을 공부하고 있다. 심리학, 뇌과학, 통계학, 기계학습 등 분야에 관심이 있으며, 과학 재현성(reproducibility) 관련 논의에도 관심이 많다.

이메일 : park.1952@osu.edu

뉴스룸

지식창고

전망대

라운지

[H](#) [뉴스룸](#) > [심층](#) > [기사보기](#)

“뇌영상 분석 기법에 심각한 오류 가능성” 논문 파장

박준석 | 2016. 07. 13

추천수 0 [📖](#) [보내기](#)

뉴스 & 해설

“자주 쓰는 fMRI 분석 소프트웨어에 근본 오류 가능성”
“기존 연구 오류 심각” vs “그리 많지 않다” 판단 일러
관행처럼 쓰는 분석·통계기법에 대해 비판적 접근 필요

Fast Random Permutation Tests Enable Objective Evaluation of Methods for Single-Subject fMRI Analysis

Anders Eklund,^{1,2} Mats Andersson,^{1,2} and Hans Knutsson^{1,2}

¹ Division of Medical Informatics, Department of Biomedical Engineering, Linköping University, Linköping, Sweden

² Center for Medical Image Science and Visualization

NeuroImage 61 (2012) 565–578

Correspondence should be addressed to Anders Eklund

Received 19 April 2011; Accepted 14 July 2011

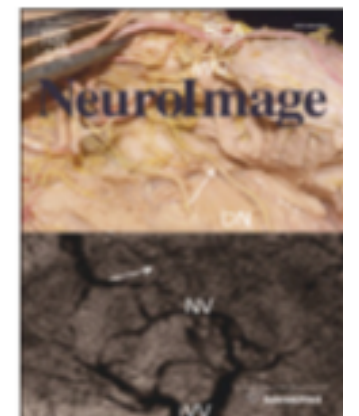
Academic Editor: Yasser M. Kadah



Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



Does parametric fMRI analysis with SPM yield valid results? —An empirical study of 1484 rest datasets

Anders Eklund^{a,b,*}, Mats Andersson^{a,b}, Camilla Josephson^{b,c}, Magnus Johannesson^{c,d}, Hans Knutsson^{a,b}

^a Division of Medical Informatics, Department of Biomedical Engineering, Linköping University, Linköping, Sweden

^b Center for Medical Image Science and Visualization (CMIV), Linköping University, Linköping, Sweden

^c Department of Management and Engineering, Linköping University, Linköping, Sweden

^d Department of Economics, Stockholm School of Economics, Stockholm, Sweden

arXiv:1511.01863v1 [stat.AP] 5 Nov 2015

**CAN PARAMETRIC STATISTICAL METHODS BE TRUSTED
FOR FMRI BASED GROUP STUDIES?**

Anders Eklund^{a,b,c}, Thomas Nichols^d, Hans Knutsson^{a,c}

^aDivision of Medical Informatics, Department of Biomedical Engineering,
Linköping University, Linköping, Sweden

^bDivision of Statistics and Machine Learning, Department of Computer and Information Science,
Linköping University, Linköping, Sweden

^cCenter for Medical Image Science and Visualization (CMIV),

Linkö

^dDepartment of Statistics



Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates

Anders Eklund^{a,b,c,1}, Thomas E. Nichols^{d,e}, and Hans Knutsson^{a,c}

^aDivision of Medical Informatics, Department of Biomedical Engineering, Linköping University, S-581 85 Linköping, Sweden; ^bDivision of Statistics and Machine Learning, Department of Computer and Information Science, Linköping University, S-581 83 Linköping, Sweden; ^cCenter for Medical Image Science and Visualization, Linköping University, S-581 83 Linköping, Sweden; ^dDepartment of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom; and ^eWMG, University of Warwick, Coventry CV4 7AL, United Kingdom

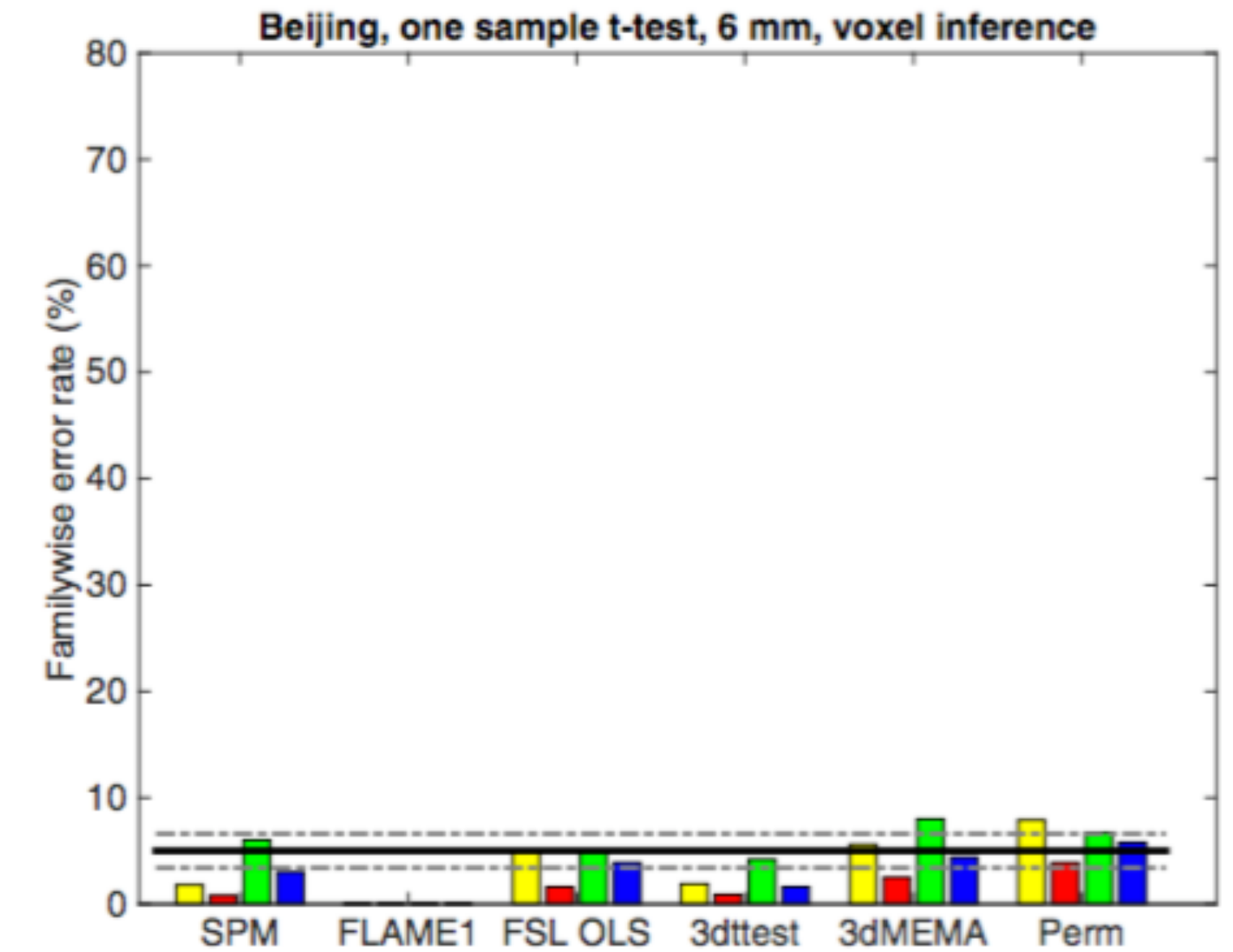
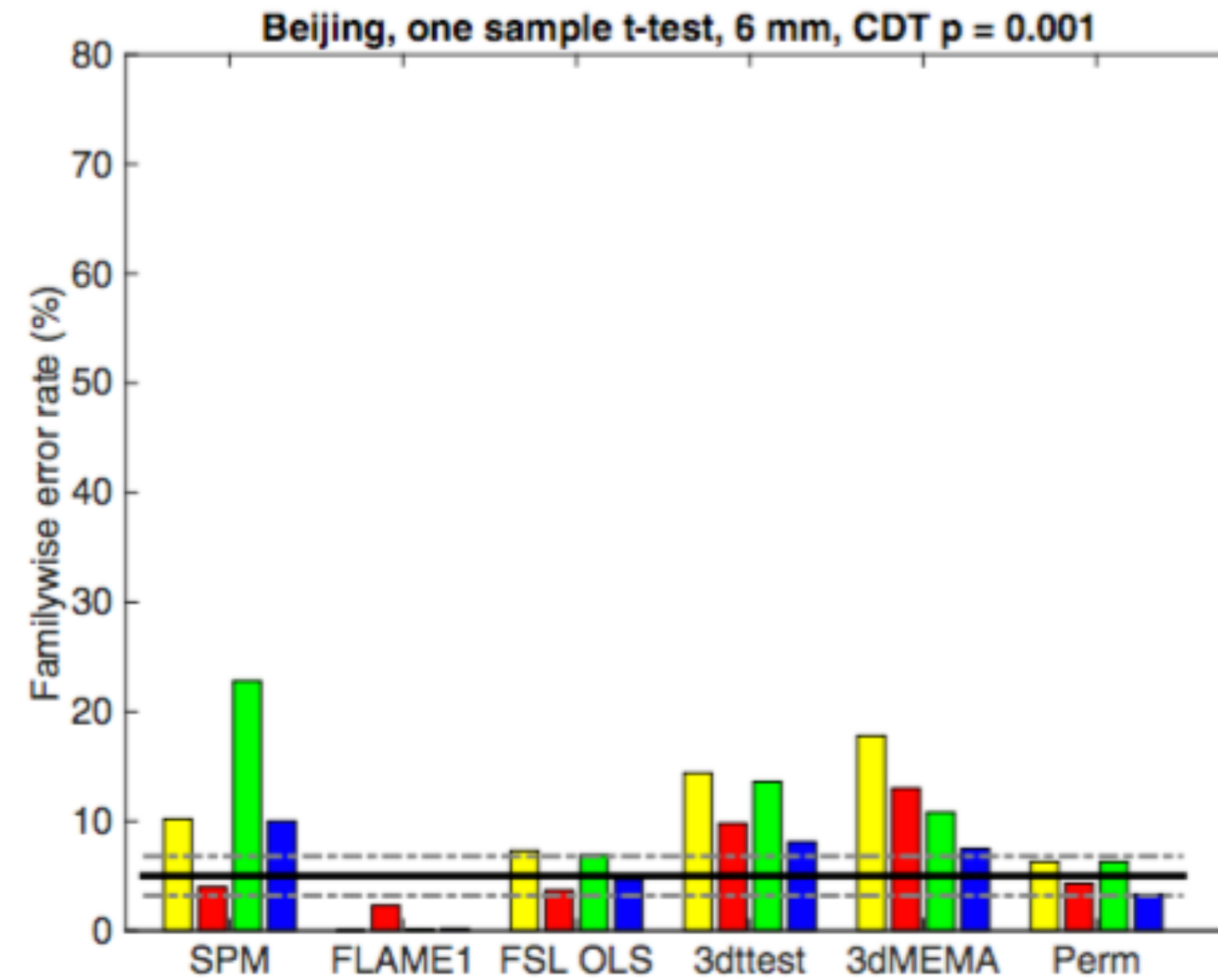
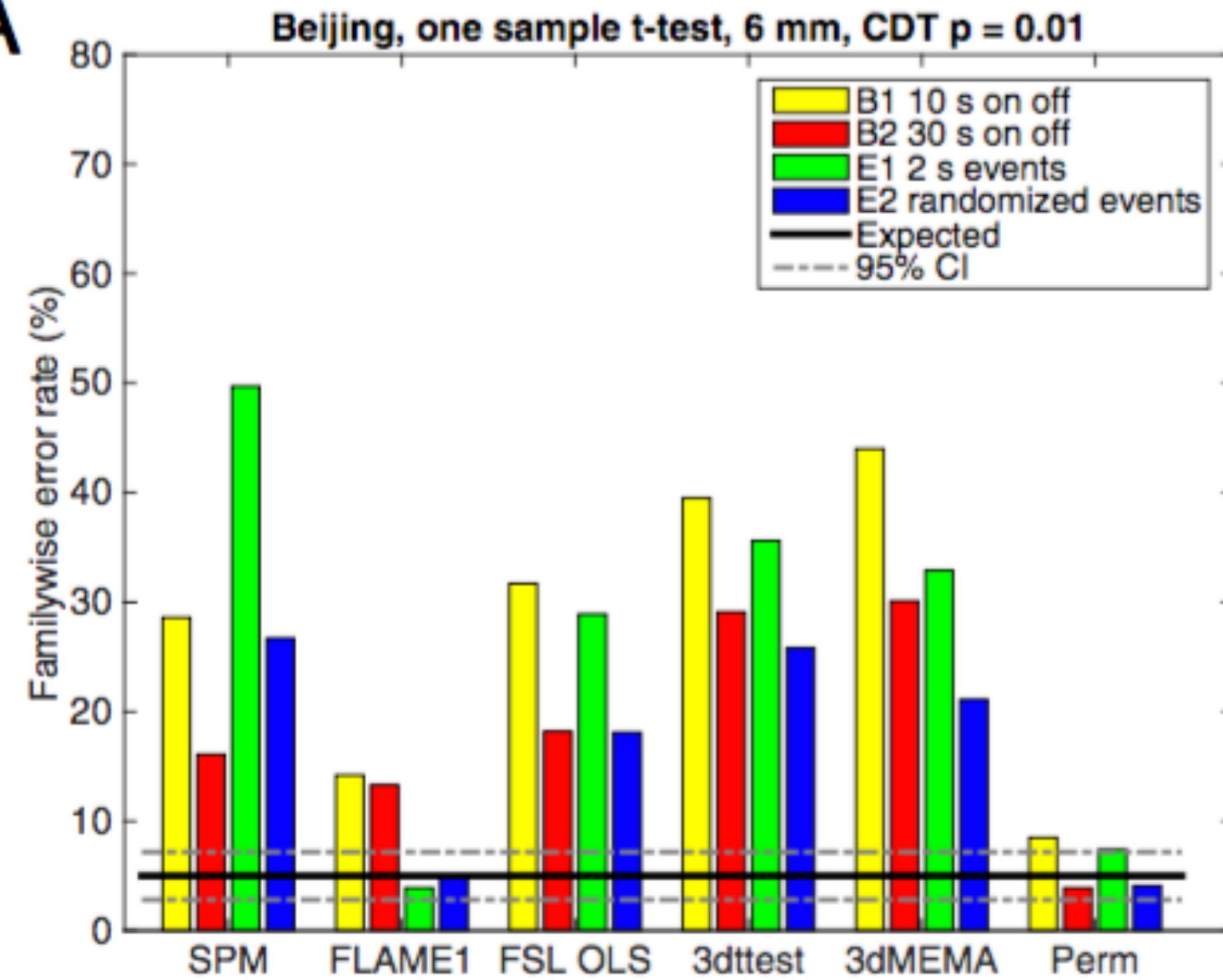
Edited by Emery N. Brown, Massachusetts General Hospital, Boston, MA, and approved May 17, 2016 (received for review February 12, 2016)

PNAS

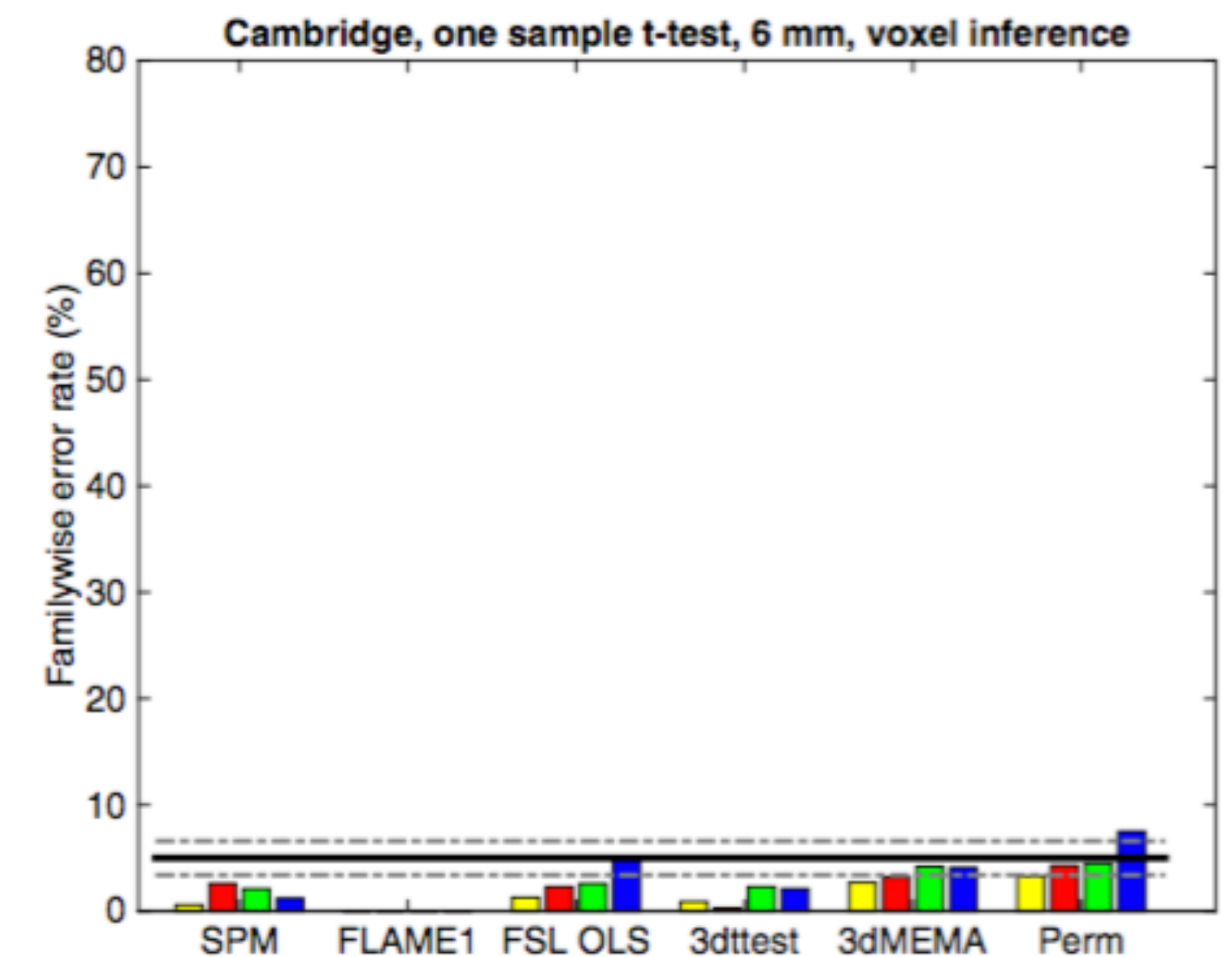
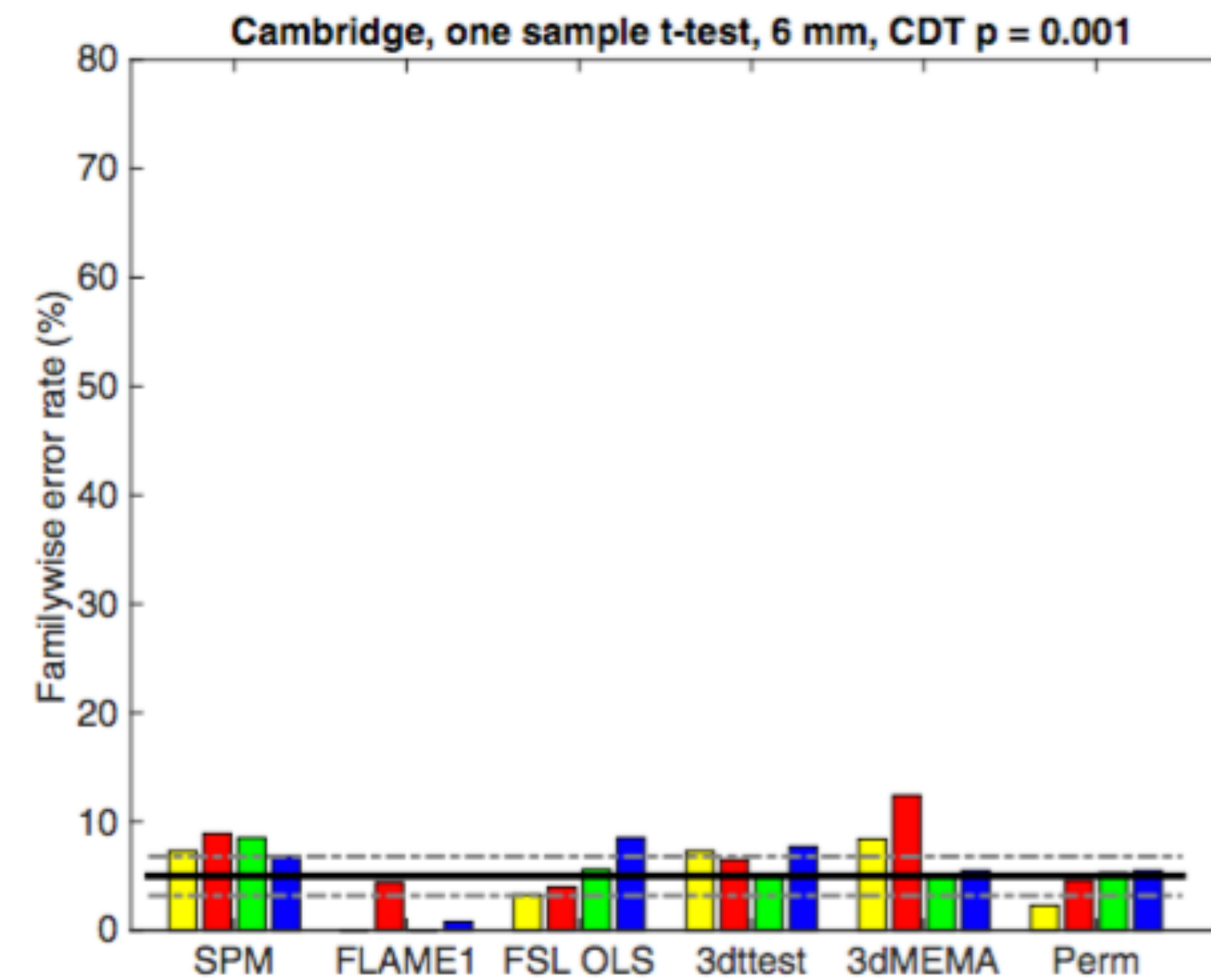
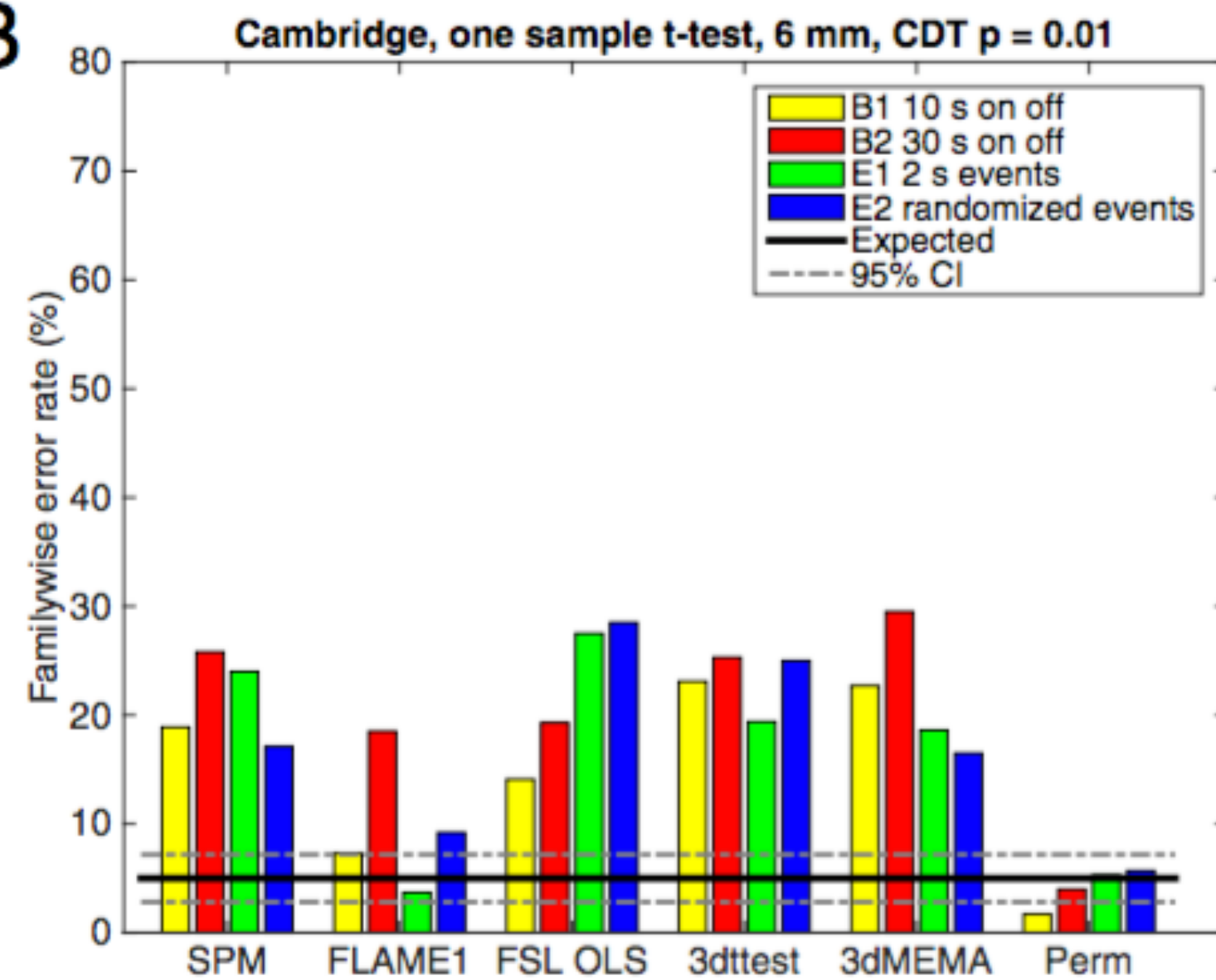
CDT $p = 0.01$

CDT $p = 0.001$

A



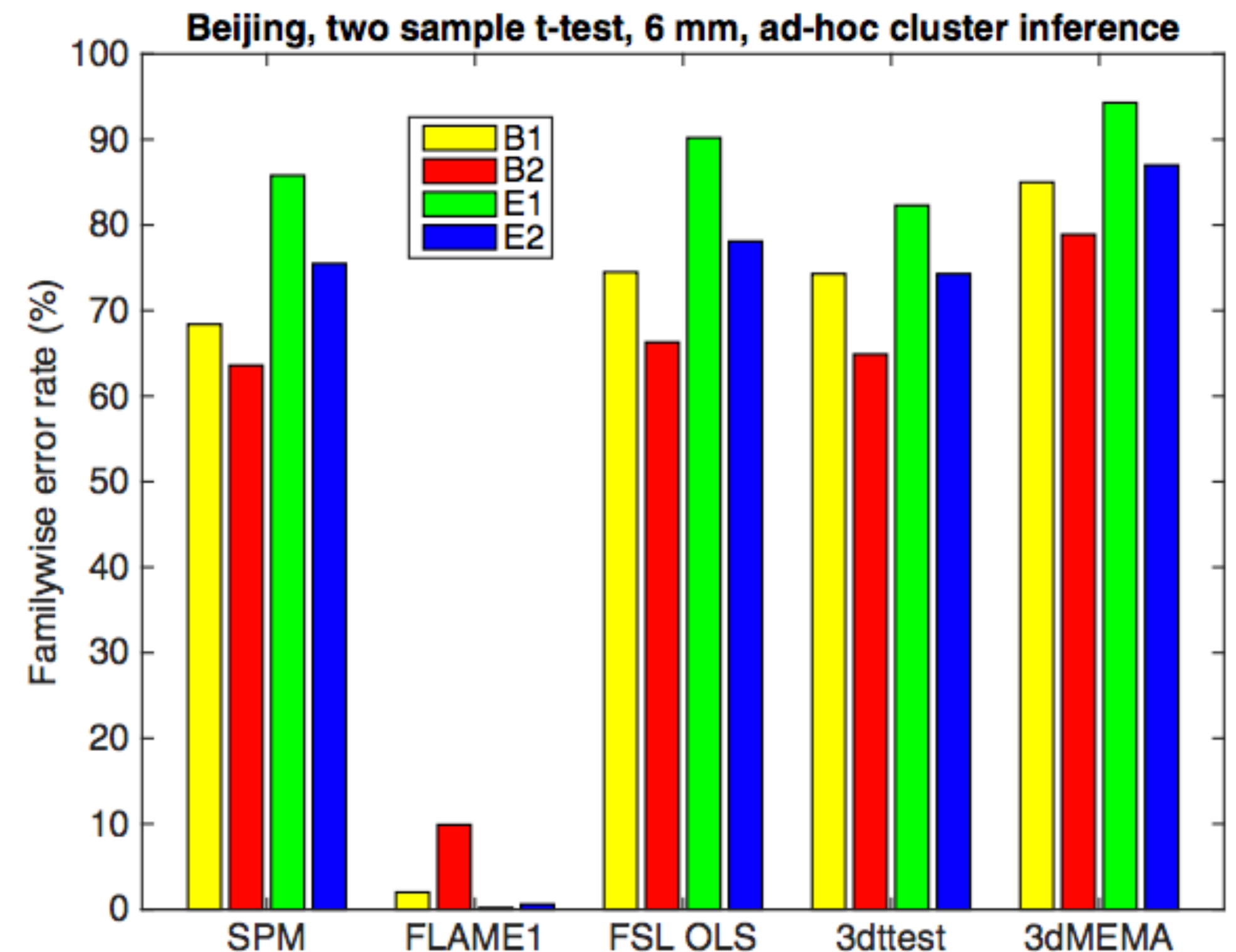
B



In two sample t-test

We also examined an ad hoc but commonly used thresholding approach, where a CDT of $P = 0.001$ (uncorrected for multiple comparisons) is used together with an arbitrary cluster extent threshold of 10 8-mm³ voxels (26, 28). We conducted an additional 1,000 analyses repeated for four assumed activity paradigms and the five different analysis tools (Fig. 2). Although no precise control of false positives is assured, we found this makeshift inference method had FWE ranging 60–90% for all functions except FLAME1 in FSL. Put another way, this “ $P = 0.001 + 10$ voxels” method has a FWE-corrected P value of 0.6–0.9. We now seek to understand the sources of these inaccuracies.

Anders Eklund et al, PNAS, 2016.



Discussion in PNAS paper

- SPM, FSL, and AFNI can produce **FWE-corrected cluster P -value that are erroneous**, being spuriously low and inflating statistical significance.
- The most common cluster extent threshold used is 80 mm^3 (10 voxels of size $2 \times 2 \times 2 \text{ mm}$), for which the FWE was estimated to be 60-90%.
- a 15-year-old **bug** was found in 3dClustSim (the bug was fixed by the AFNI group as of May 2015). The bug essentially reduced the size of the image searched for clusters, producing underestimate smoothness.
- FWE rates are lower with the bug-fixed 3dClustSim function.
- Parametric methods work well for a high CDT (resulting in small clusters) and not as well for a low CDT. This is well known and consistent with RFT.

Analysis of family-wise error rates in statistical parametric mapping using random field theory

Guillaume Flandin and Karl J. Friston

*Wellcome Trust Centre for Neuroimaging
University College London, UK*

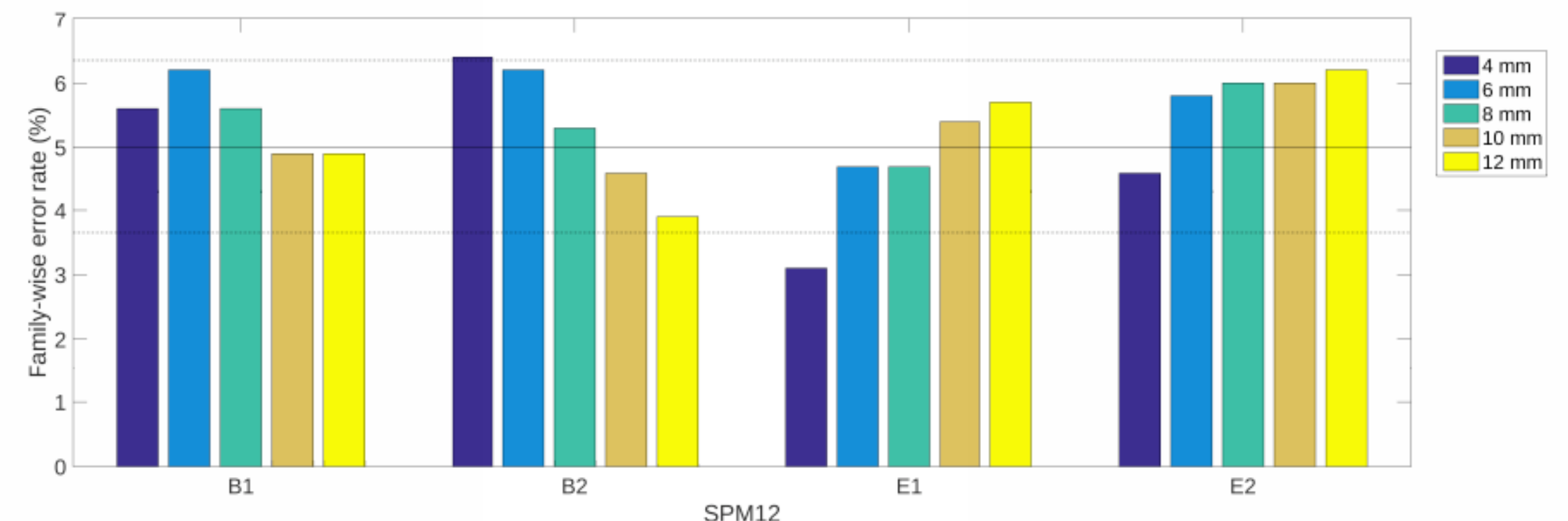
27th June 2016

Abstract

This technical report revisits the analysis of family-wise error rates in statistical parametric mapping – using random field theory – reported in (Eklund et al., 2015). Contrary to the understandable spin that these sorts of analyses attract, a review of their results suggests that they endorse the use of parametric assumptions – and random field theory – in the analysis of functional neuroimaging data. We briefly rehearse the advantages parametric analyses offer over nonparametric alternatives and then unpack the implications of (Eklund et al., 2015) for parametric procedures.

Review of Edlund et al. results

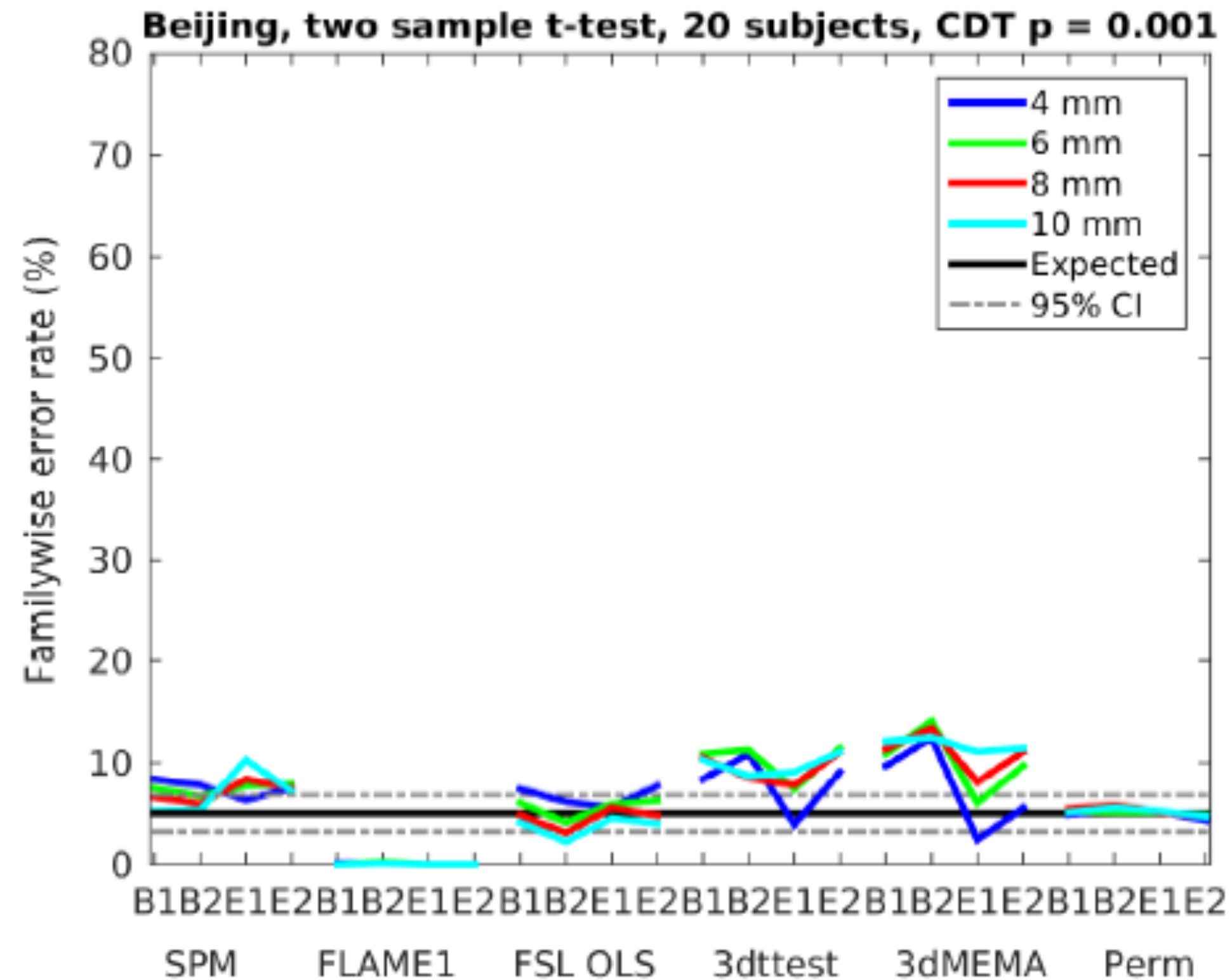
- increasing the smoothness will lead to more exact inference.
- In Eklund (2015), authors used **the same regressor for all subjects**.
This was a mistake because any systematic fluctuation in resting state time series will lead to significant one sample t-test against the null hypothesis of zero.
- The results provided by (Eklund et al, 2015) endorse the use of RFT and reveal behaviour that is consistent with the underlying distributional assumptions.



Parametric vs Non-parametric

- According to Neyman-Pearson lemma,
Any nonparametric test can only be as efficient as a parametric test or less efficient.
- Parametric approaches are reproducible.
- Parametric approaches eschew the problem of complying with the exchangeability criteria of nonparametric procedures.
- Parametric approaches are computationally more efficient because they use distributional assumptions.

Eklund et al. (2016)



Flandin et al. (2016)

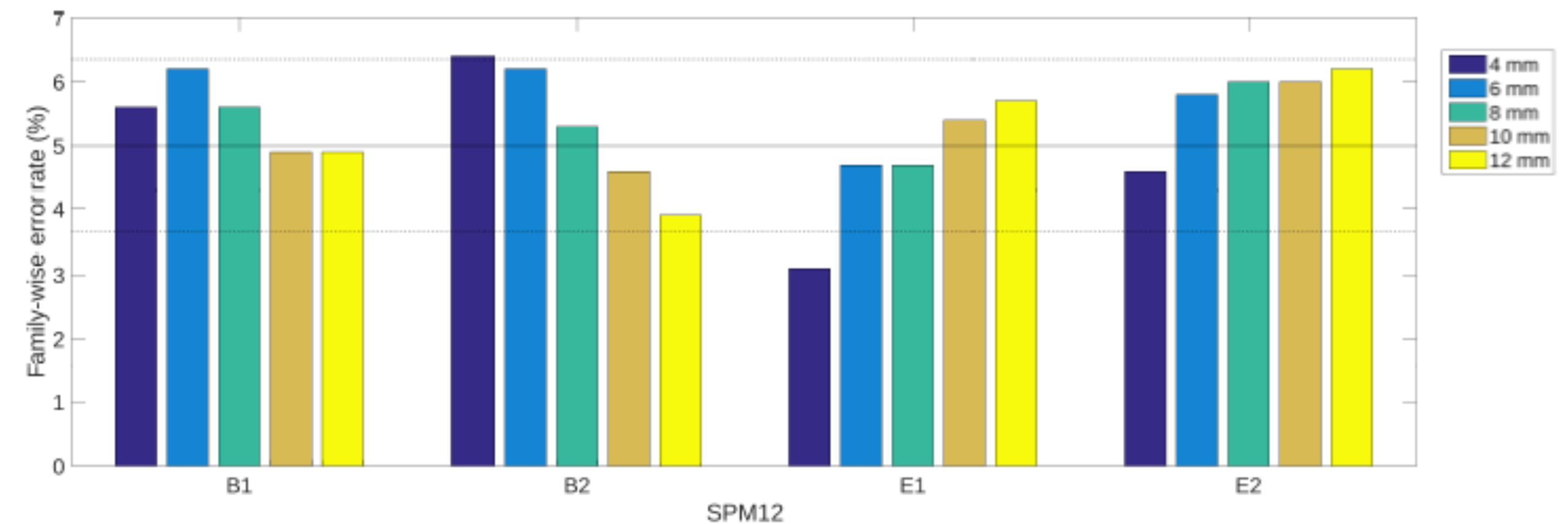


Figure 1: Cluster-level inference results for a two-sample t -test (two groups of ten random subjects, repeated a thousand times) with the Beijing dataset using a cluster-forming threshold of $p < 0.001$ and the SPM12 software (r6685). Five levels of spatial smoothing were evaluated (4, 6, 8, 10 and 12 mm isotropic Gaussian kernels) with four different regressors (see (Eklund et al., 2015) for details).

Two differences for these analyses: First, Flandin et al. used voxel size of 3mm, but Eklund et al. employed voxel size of 2mm; Second, Flandin et al. used SPM12 but Eklund et al. utilized the SPM8.

In summary, Parametric vs. non-parametric

using distribution
ex) FWE in SPM

using Monte Carlo simulation
or permutation testing

Parametric vs. non-parametric

- 뇌영상 연구에서 활성화 정도를 통계기법으로 찾을때 사용하는 방법을 크게 두가지로 나눈다면 ‘모수적 방법(parametric)과 비모수적 방법(nonparametric)’ 이 될것임.
- 둘의 차이는 데이터에 대해 얼마나 많은 가정을 하는지에 달려 있는데, 모수적 방법이 비모수적 방법에 비해 더 많은 가정이 필요함.
- 그 가정이 옳으면 모수적 방법법이 비모수적 방법보다 더 효율적이며 정확하지만, 그렇지 않으면 비모수적 방법이 더 나은 결과를 산출함.
- 특히 뇌영상 분석에서 많이 사용되는 SPM, FSL, AFNI 등의 프로그램은 모두 데이터의 특정 측면에 대한 ‘정규분포’ 가정을 갖고 있는데, 이것이 틀렸을 수도 있다는 것.

In voxel-wise neuroimaging study,

There are two thresholds to be controlled

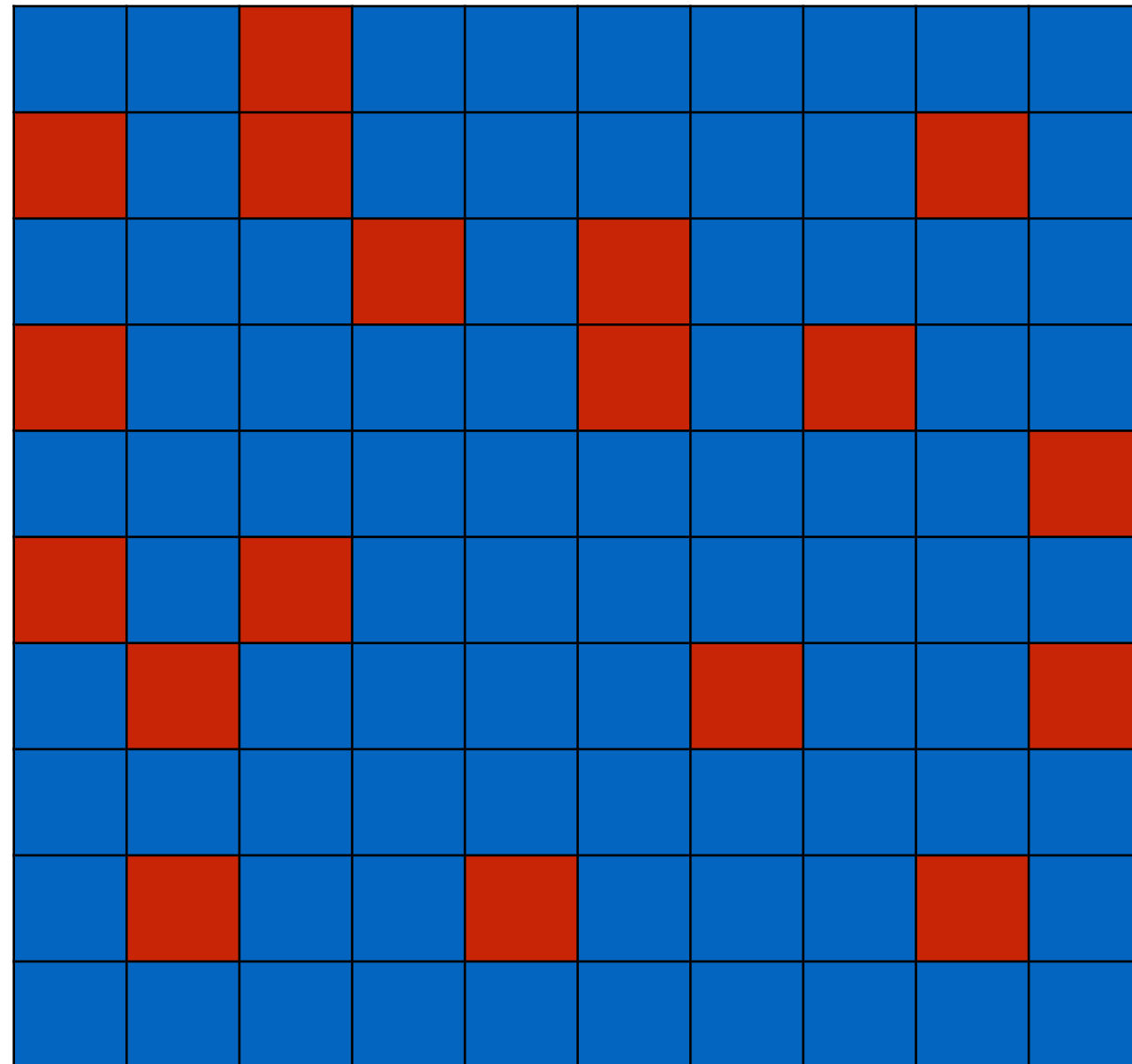
a significant activation at each voxel and

a significance of a cluster size

Significant results at each voxel

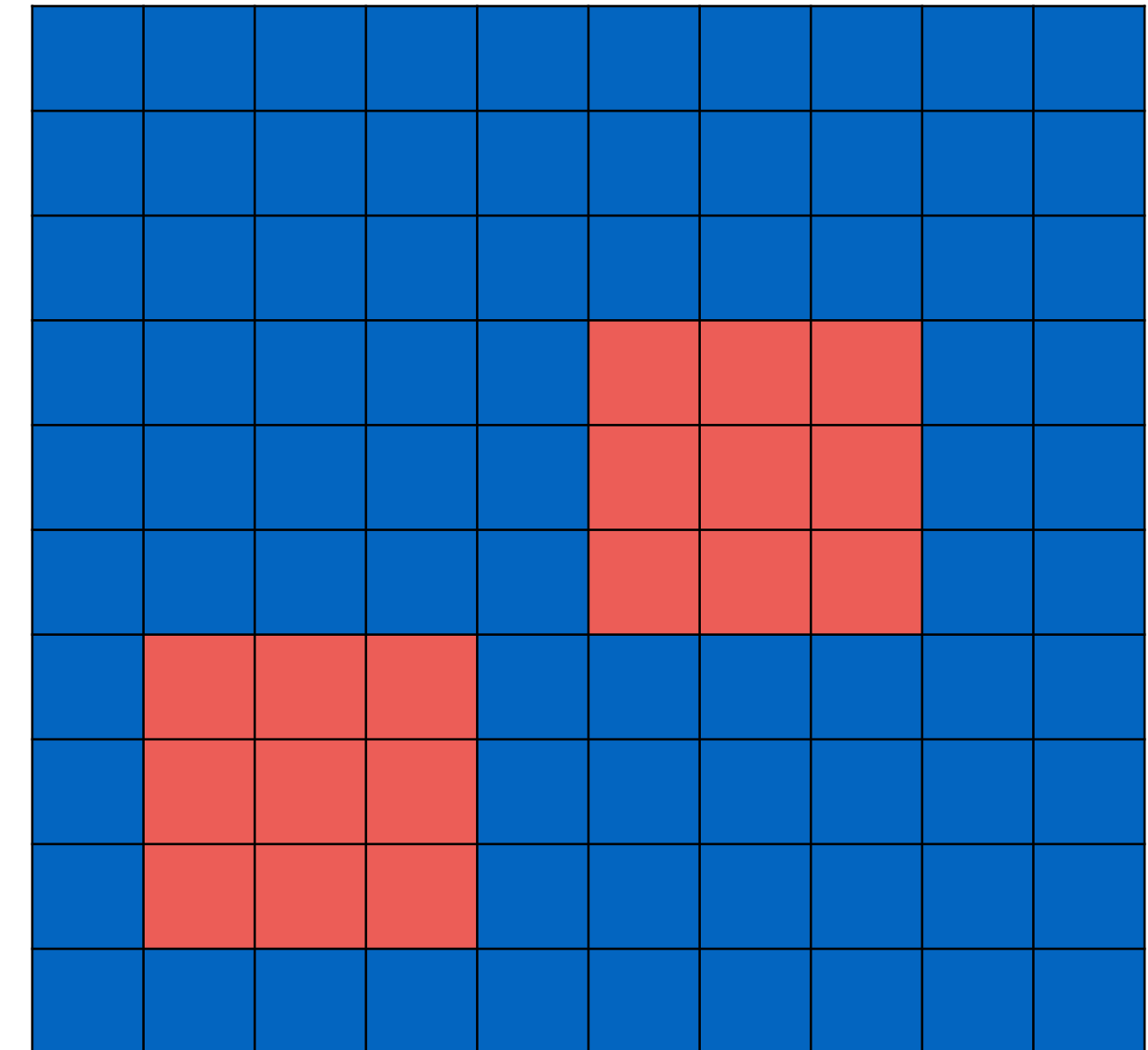
which is more strongly activated?

Threshold at each voxel is $p < 0.001$



or

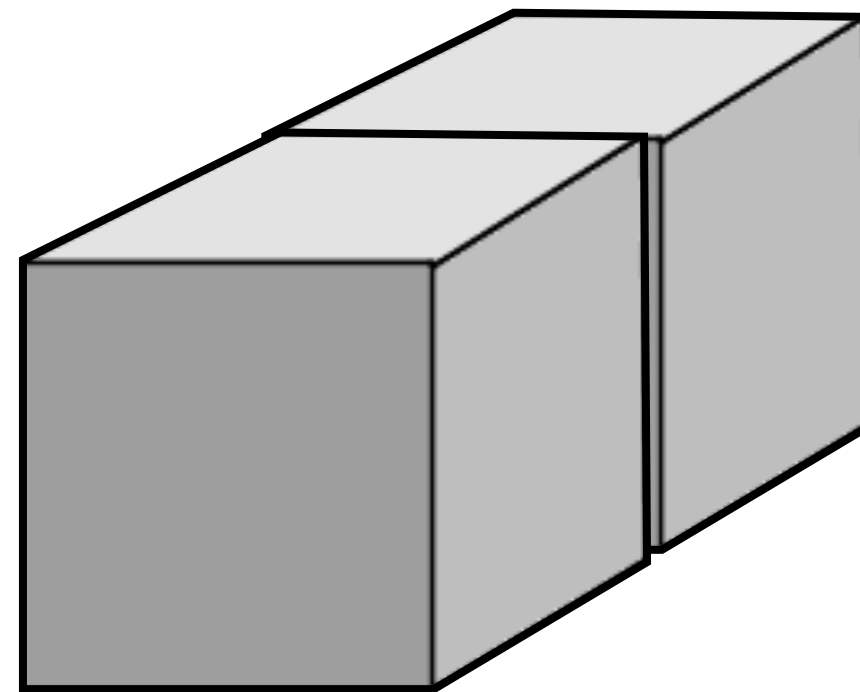
Threshold at each voxel is $p < 0.005$ & $k \geq 9$



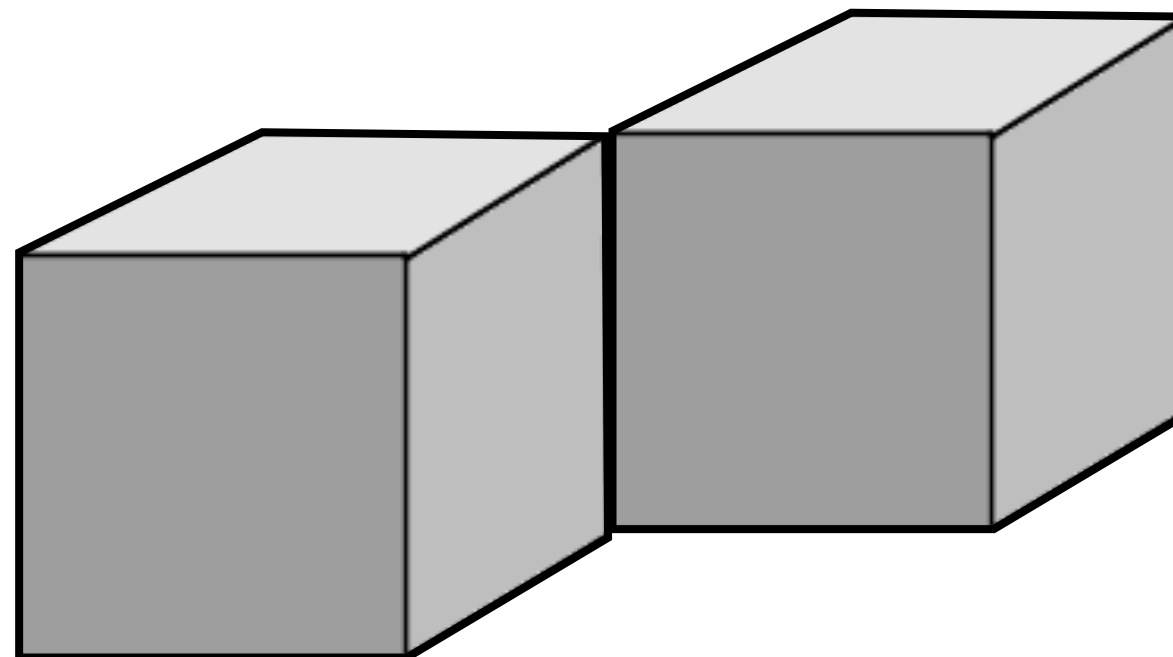
What is the meaning of a cluster?

Contiguous voxels?

face-connected
(or surface)

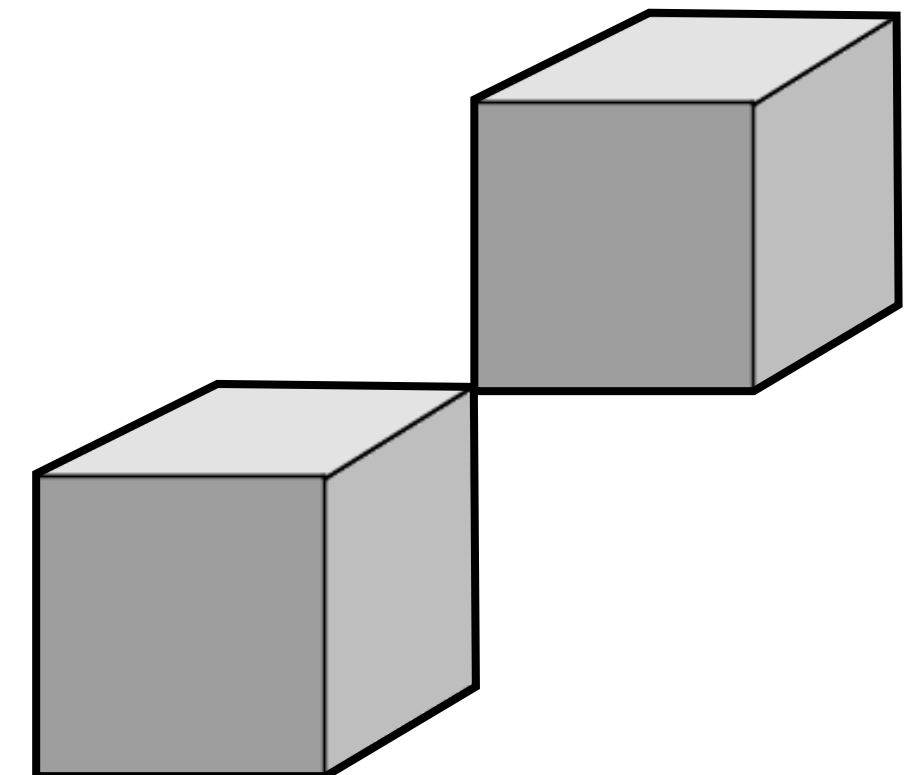


edge-connected

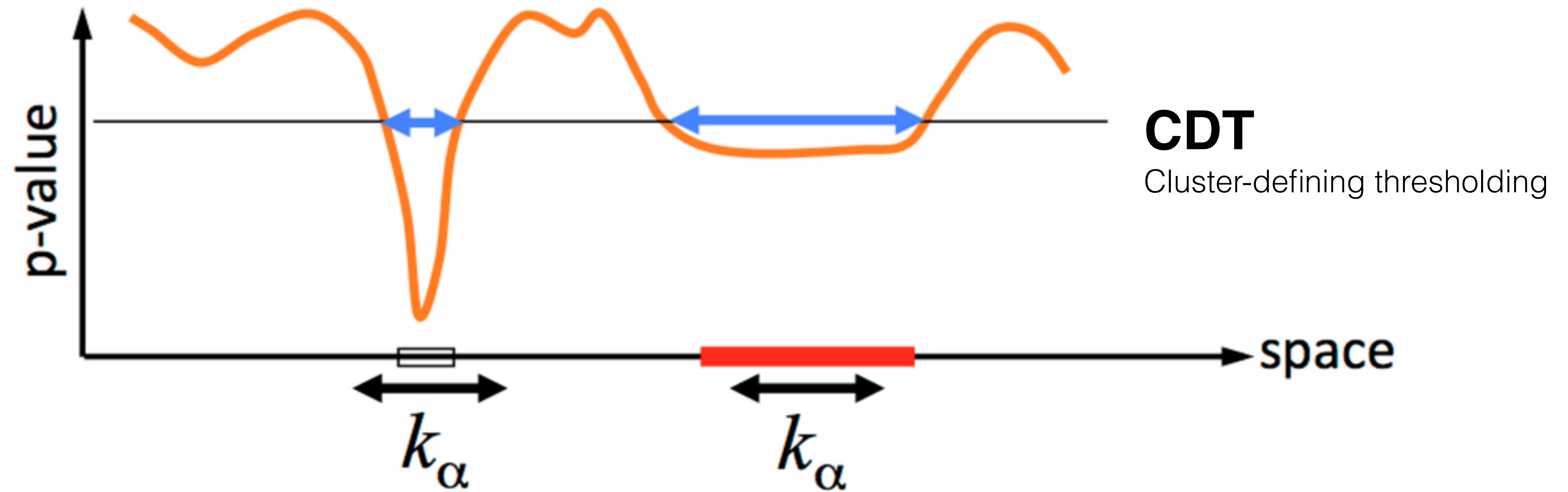


(SPM default)

corner-connected



How cluster-based thresholding works



Trade-off between a significance of CDT and a size of cluster extent

Family-wise error correction

- a. using SPM built-in functionality
- b. using SnPM external toolbox
- c. using 3dClustSim & 3dFWHMx (AFNI software)

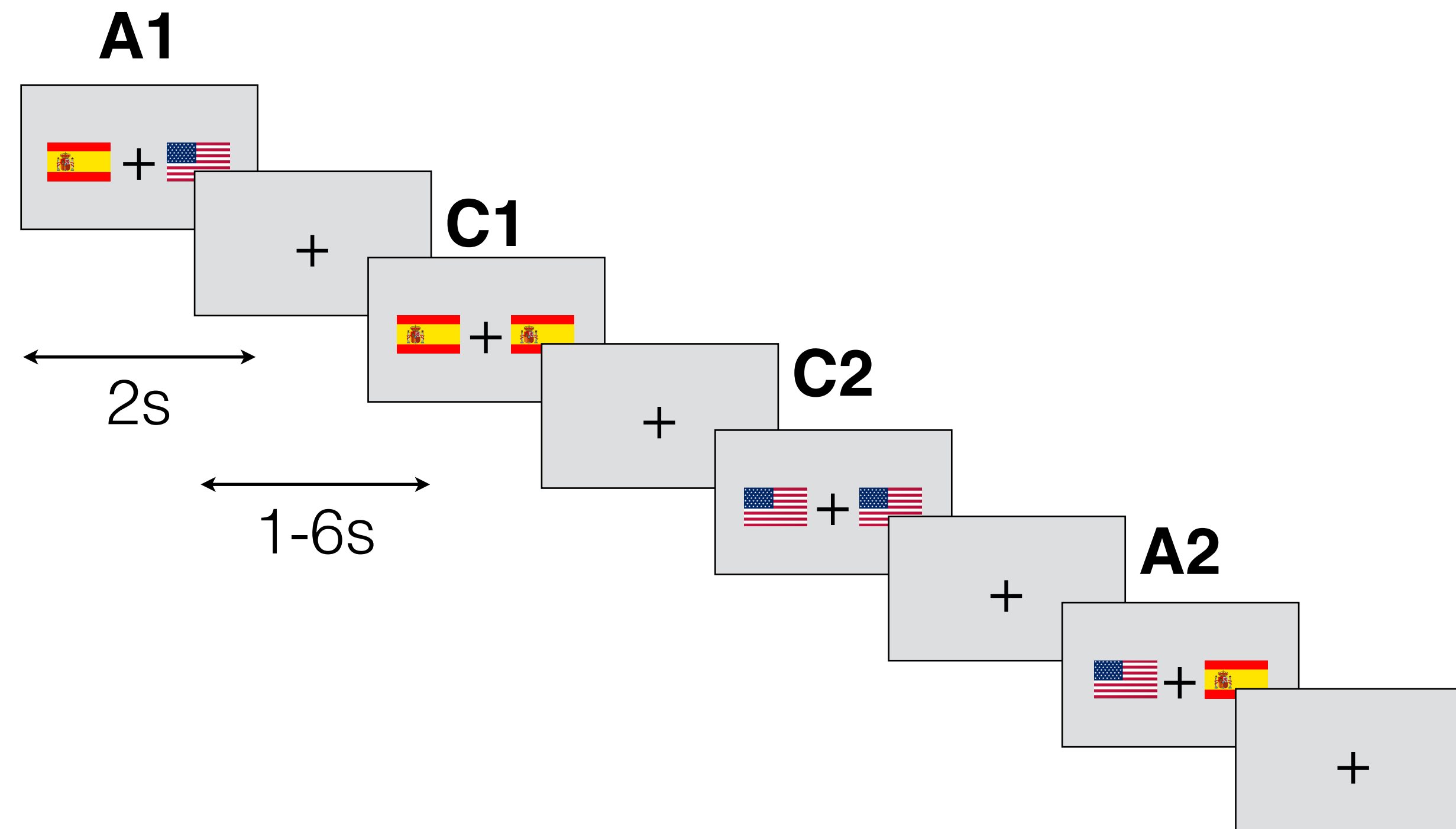
실습 데이터 다운로드

First-level 분석이 완료된 data:

https://drive.google.com/drive/folders/0B6863JbB_YV-bWFoUE5sbmRvOXc

어떤 데이터 일까?

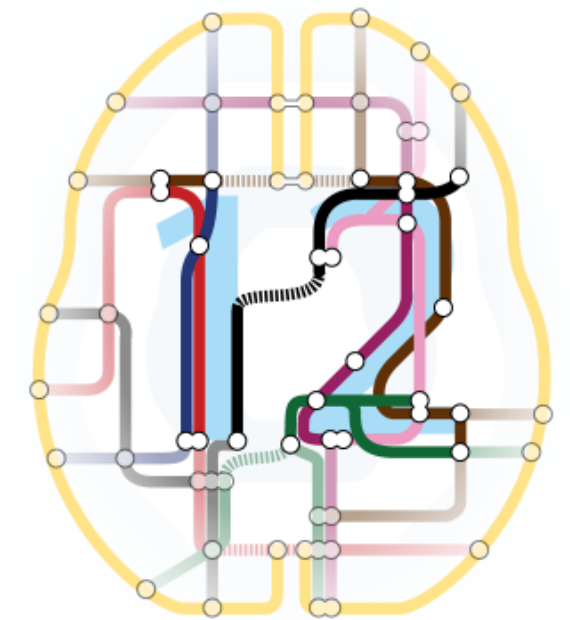
Cross-language repetition priming



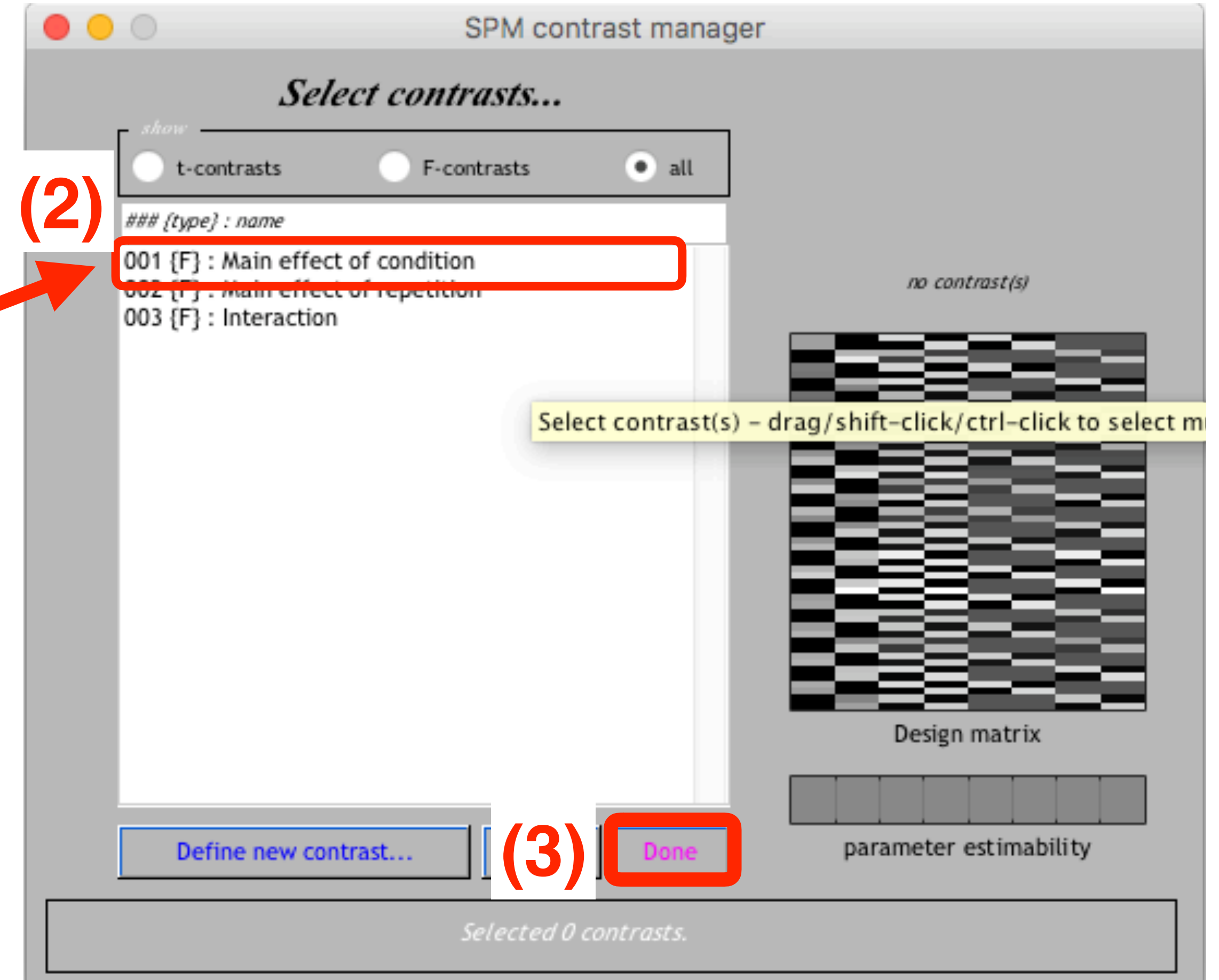
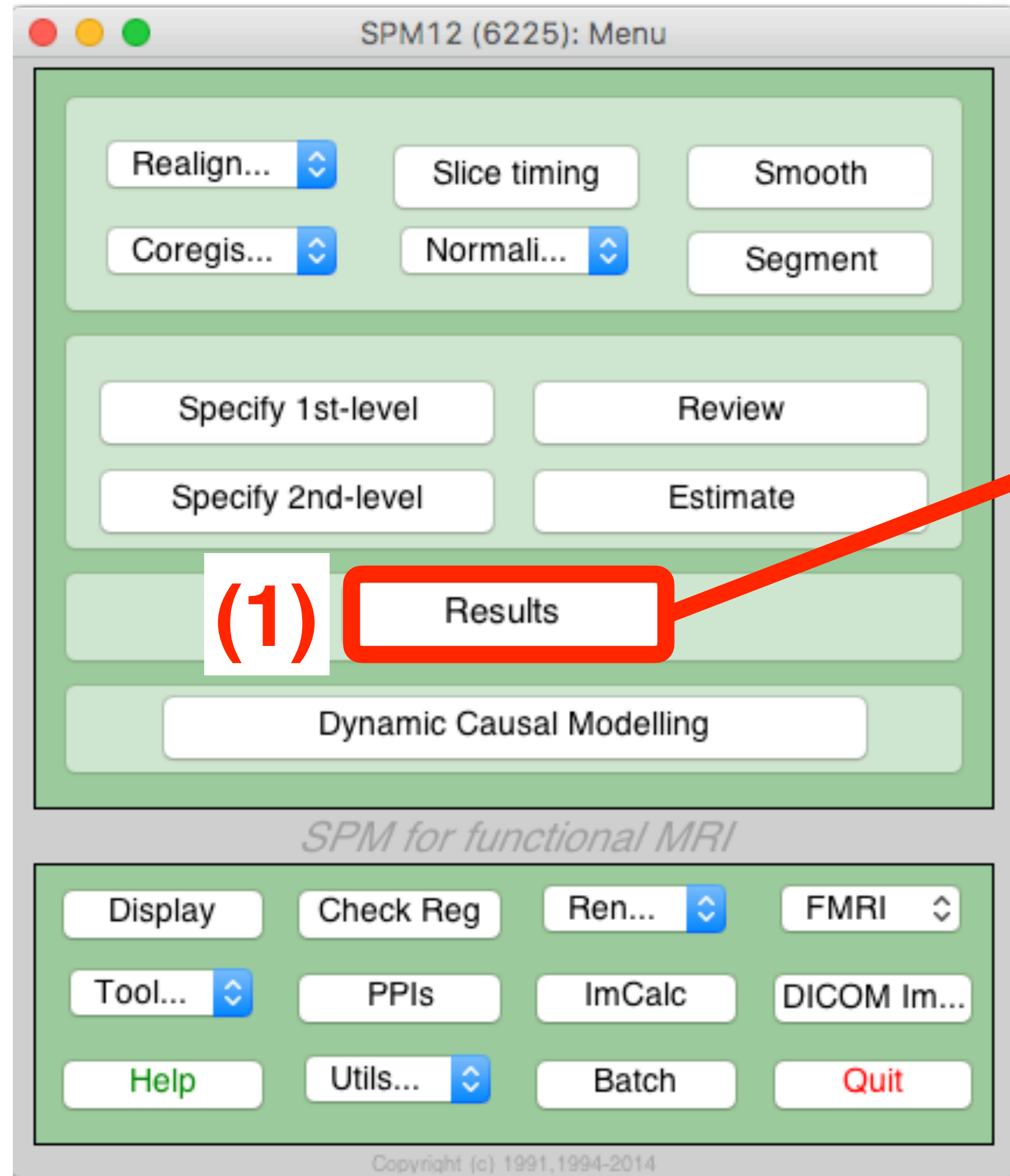
Concreteness (C/A)	
Repetition (1/2)	A1
	C1
Repetition (1/2)	A2
	C2

Family-wise error correction

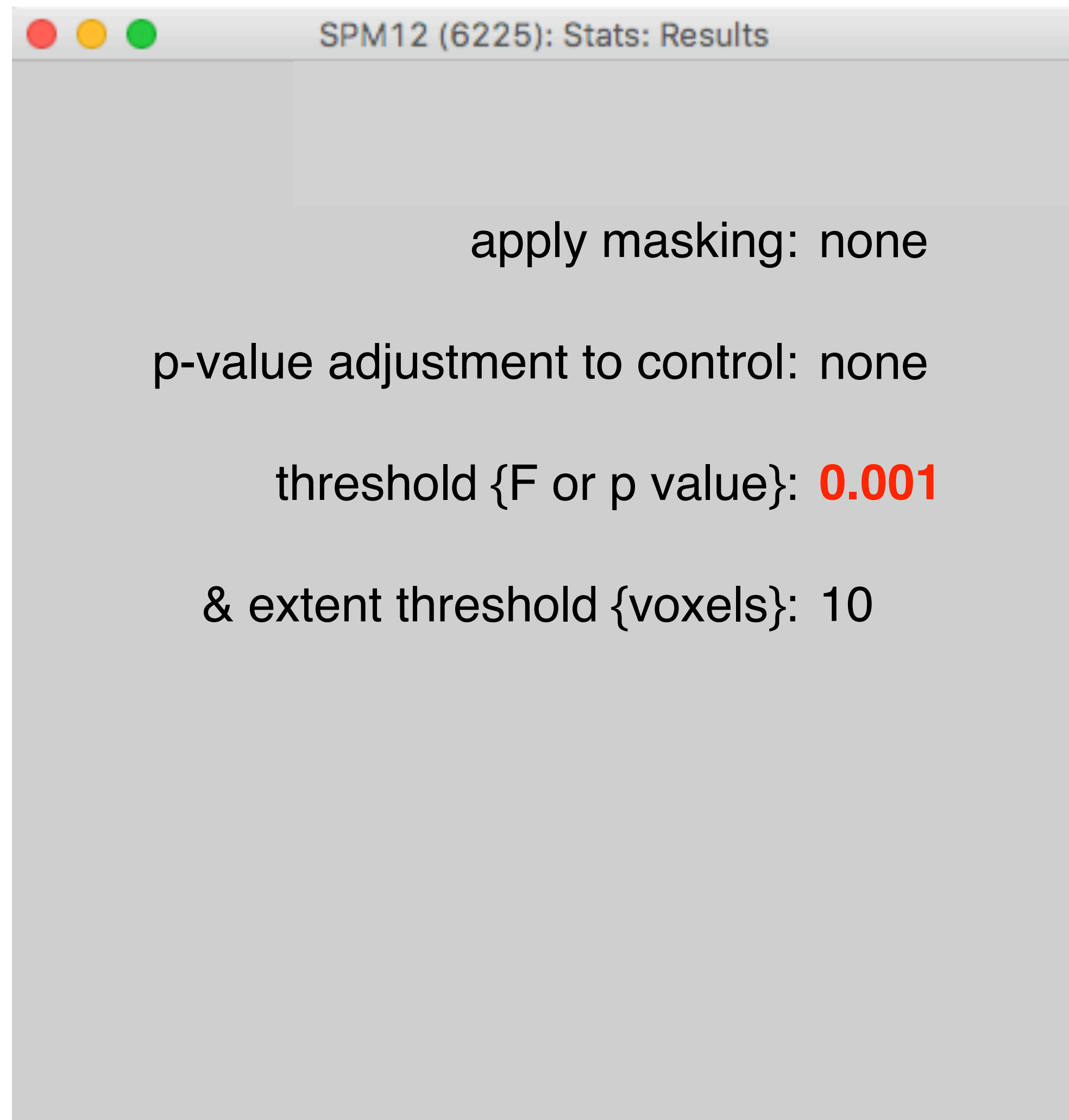
in Statistical Parametric Mapping (**SPM**) toolbox



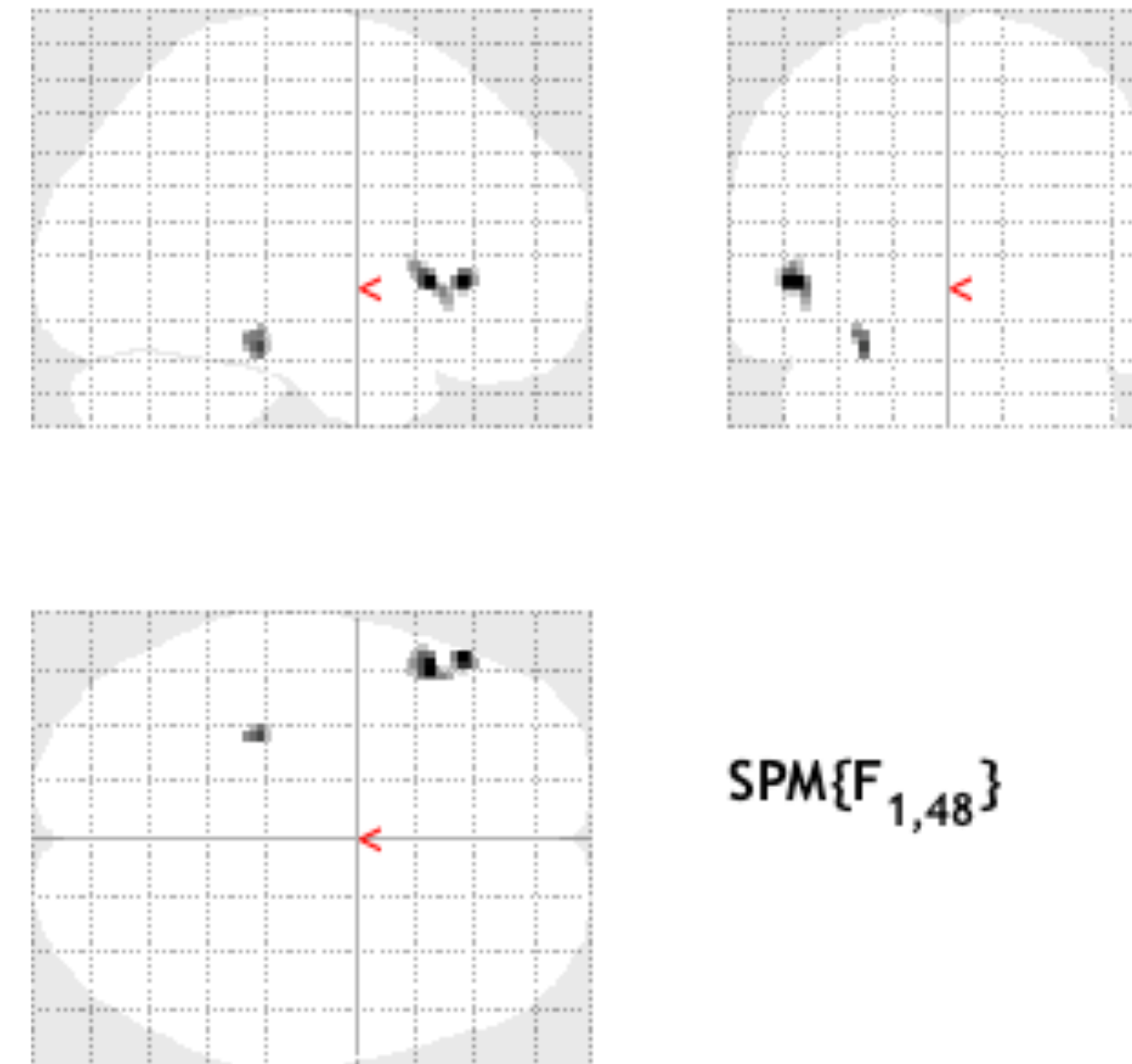
Results : Cluster-level correction (1/3)



Results : Cluster-level correction (2/3)



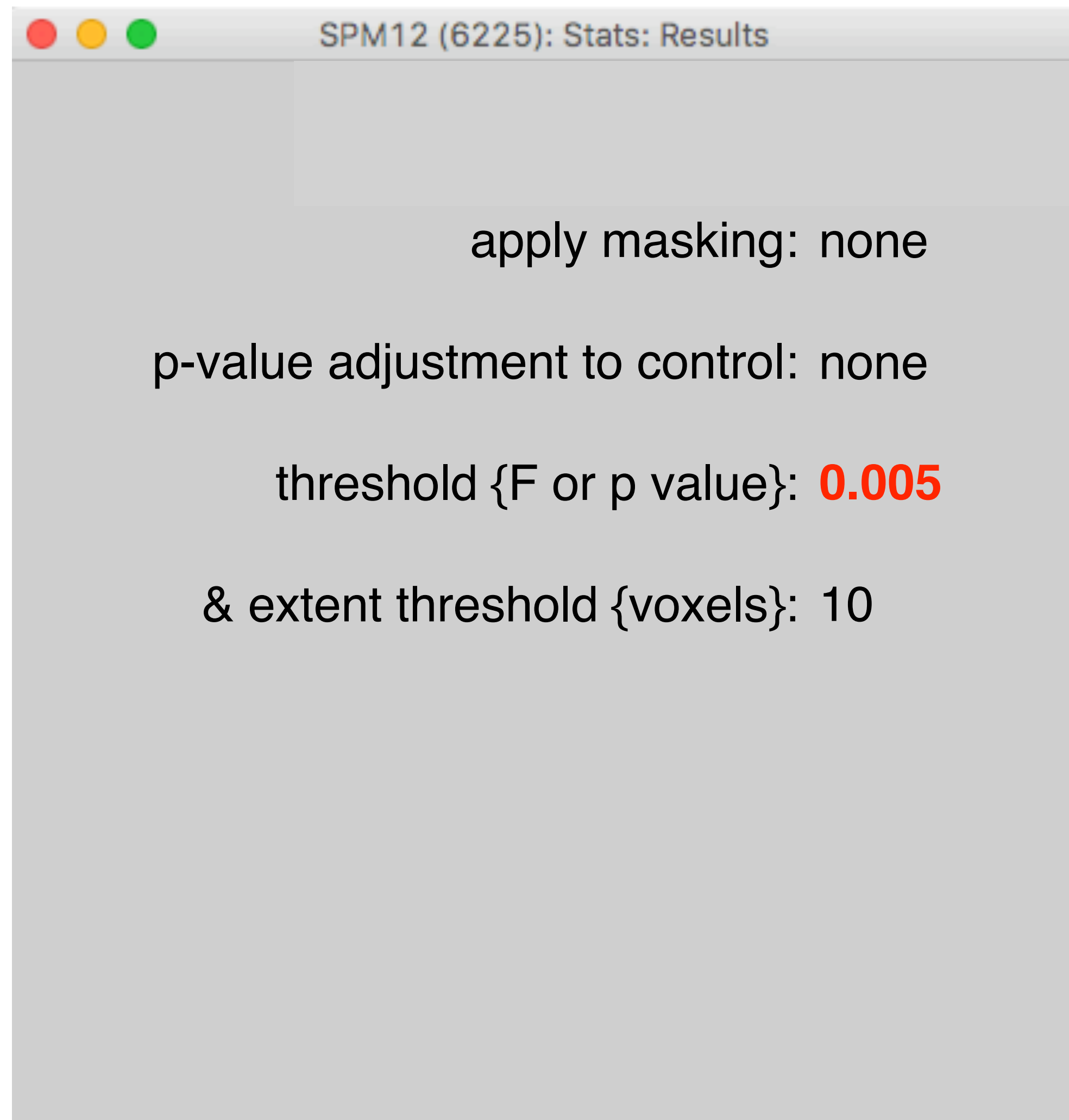
Main effect of condition



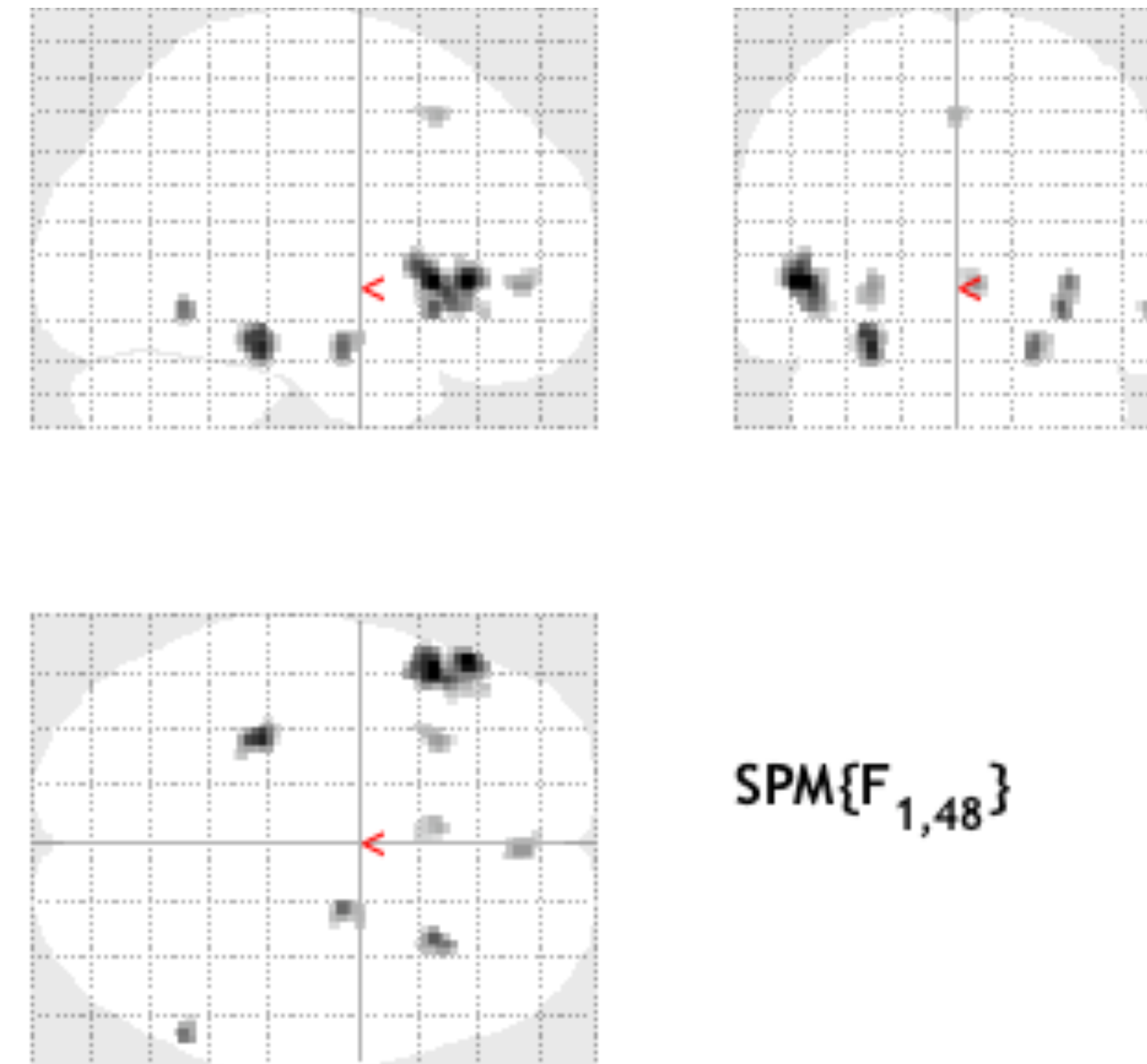
Statistics: *p-values adjusted for search volume*

set-level		cluster-level			
ρ	c	$\rho_{\text{FWE-corr}}$	$q_{\text{FDR-corr}}$	k_E	ρ_{uncorr}
0.897	3	0.896		21	0.114
		0.452		43	0.030
		0.951		17	0.152

Results : Cluster-level correction (3/3)



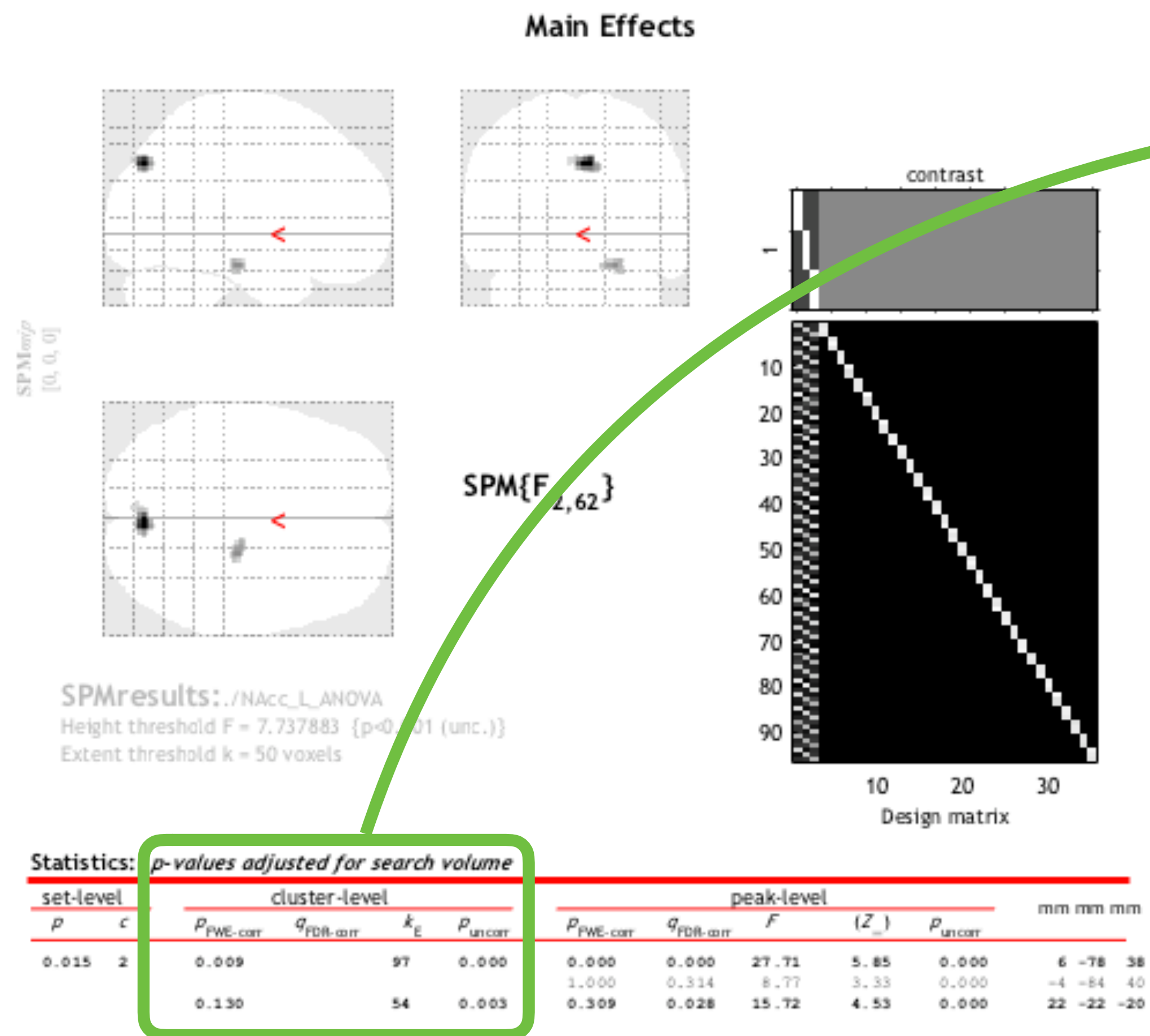
Main effect of condition



Statistics: *p-values adjusted for search volume*

set-level		cluster-level			
p	c	$p_{\text{FWE-corr}}$	$q_{\text{FDR-corr}}$	k_E	p_{uncorr}
1.000	8	0.014		274	0.000
		0.767		81	0.024
		0.943		60	0.048

Obtain corrected P in SPM



cluster-level			
$p_{FWE-corr}$	$q_{FDR-corr}$	k_E	p_{uncorr}
0.014		274	0.000
0.767		81	0.024

```
>> load SPM.mat
```

```
>> SPM.xVol.FWHM
```

```
ans =
```

```
4.5700 4.6808 4.9825 in voxel
```

```
>> SPM.xVol.FWHM*2
```

```
ans =
```

```
9.1400 9.3616 9.9651 in mm
```

How to describe ‘what we did’

SPM style

All reported regional clusters survived at a threshold of a corrected $P < 0.05$, which corresponds to the family-wise error corrected significance at the cluster level with a cluster-defining threshold of $P < 0.001$. Estimated smoothness (or spatial correlation of the results) was estimated as [FWHM in mm]=11.2, 11.3, 9.0.

AFNI style (if you use 3dClustSim)

Significant clusters ($P < 0.05$, corrected) were determined by AFNI’s 3dClustSim (or a Monte Carlo simulation) with 10,000 iterations, which in turn corresponded to the voxel-level threshold at uncorrected $P < 0.001$ and clusters involving $k > 47$ contiguous voxels.

Family-wise error correction

using 3dClustSim tool implemented in AFNI software

AFNI: <https://afni.nimh.nih.gov/download>)

XQuartz: <https://www.xquartz.org/>)



3dClustSim for determining cluster size

```
$ 3dClustSim -mask mask.nii -fwhmxyz sx sy sz -niter 10000 -NN 2 -prefix output
```

Options

- **mask** file is the a mask of the voxels you are going to test for significance in your data (you can use mask.nii if you are a SPMer)
- **fwhmxyz** is the smoothness of your data as determined by SPM (**SPM.xVol.FWHM**), FSL (smoothest), AFNI (3dFWHMx), FAST (fmri_fwhm)
[Note: if you are using SPM or FSL, you need to convert the smoothness to mm]
- **NN** option specifies the neighbor
[1=face-connected, 2=edge connected (SPM default), 3=corner connected]
- **prefix** option is the prefix of the output. the output file is a textfile named prefix.NN?.1D

Reference) <https://gate.nmr.mgh.harvard.edu/wiki/whynhow/index.php/3dClustSim>

Discussion in SPM Archives

Dear Andy,

I'm not familiar with 3dclustsim but if you do a search on this term on the mailing list, you should find advices from Donald and others regarding this tool. See also: <https://gate.nmr.mgh.harvard.edu/wiki/whynhow/index.php/3dClustSim> SPM smoothness estimate is available in:

```
>> load SPM.mat
>> SPM.xVol.FWHM % {voxels}
>> M = SPM.xVol.M(1:3,1:3);
>> SPM.xVol.FWHM .* sqrt(diag(M'*M))' % {mm}
```

That said, may I ask you what is your rationale for wanting to use 3dclustsim? As you are asking a question about AFNI on the SPM mailing list, I thought it would be appropriate to add a quote from someone working with FSL: <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=spm;2f091e87.1001> Anyway, this is a genuine question: queries about 3dclustsim/alphasim occur regularly and I'm therefore curious of what your answer is. Best regards, Guillaume.

On 08/05/15 08:25, Andy Yeung wrote:

```
> Dear all,
>
> I'd like to use 3dclustsim from AFNI.
> Documents say we need to input the estimated smoothness, and experts say
> estimate by using residual image. I wonder what values should I input
> for smoothness? Those from beneath the SPM GLM results table after
> clicking a T-contrast?
```

```
>
> Best,
> Andy -- Guillaume Flandin, PhD
Wellcome Trust Centre
for Neuroimaging University College
London 12 Queen Square London WC1N 3BG
```

SPM Mailing에 가입하면, 뇌영상 분석 방법에 대해 각종 궁금증에 대해 질문하면 전세계 유능한 과학자로부터 무료로 답변을 받을 수 있음. 또한 추후 Post-doc 구인/구직 등 좋은 정보도 얻을 수 있다.

<https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=spm>

Family-wise error correction

using Statistical non-Parametric Mapping (**SnPM**) toolbox



<http://warwick.ac.uk/snpm>

<http://www.fil.ion.ucl.ac.uk/spm/ext/#SnPM>

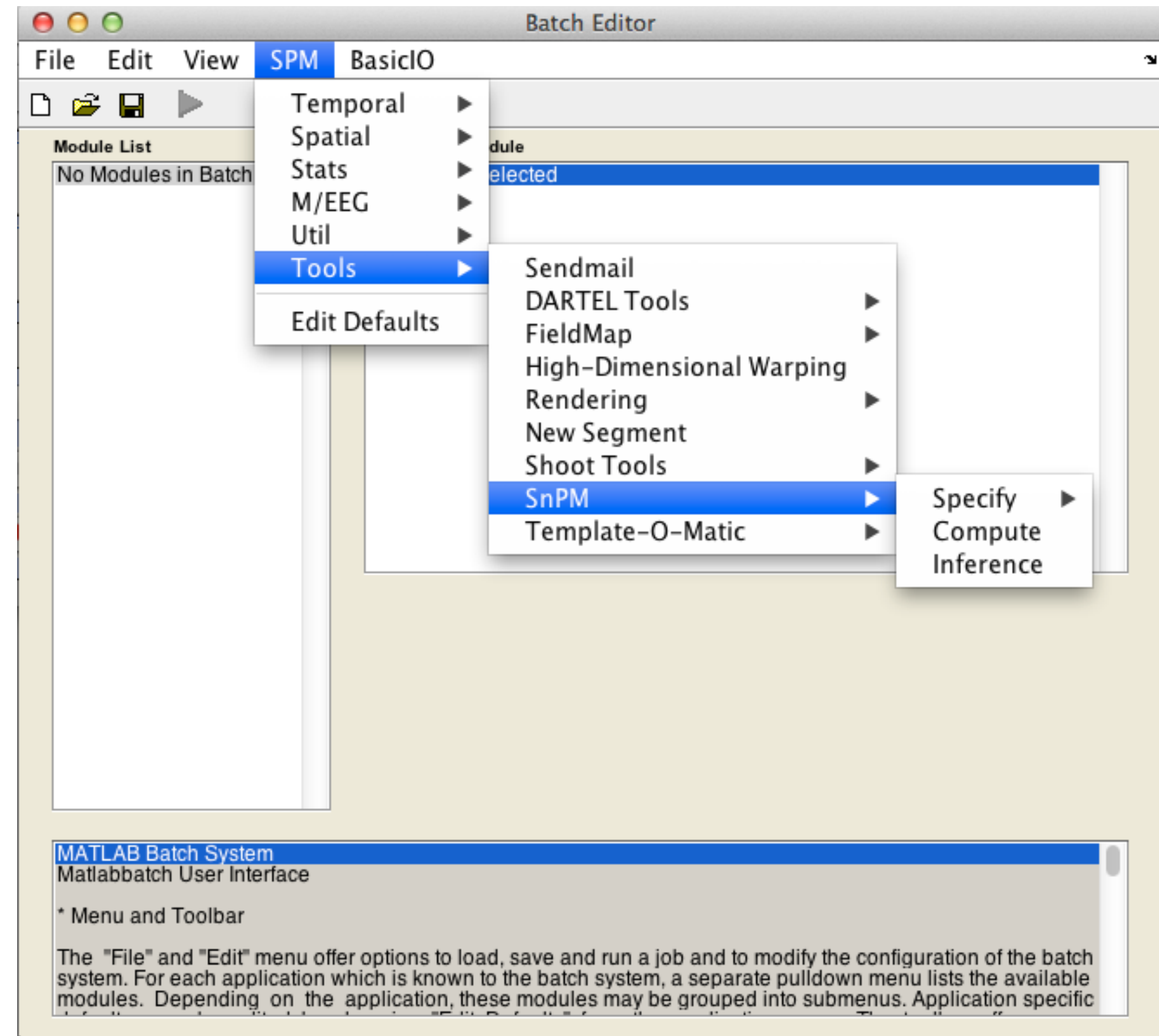
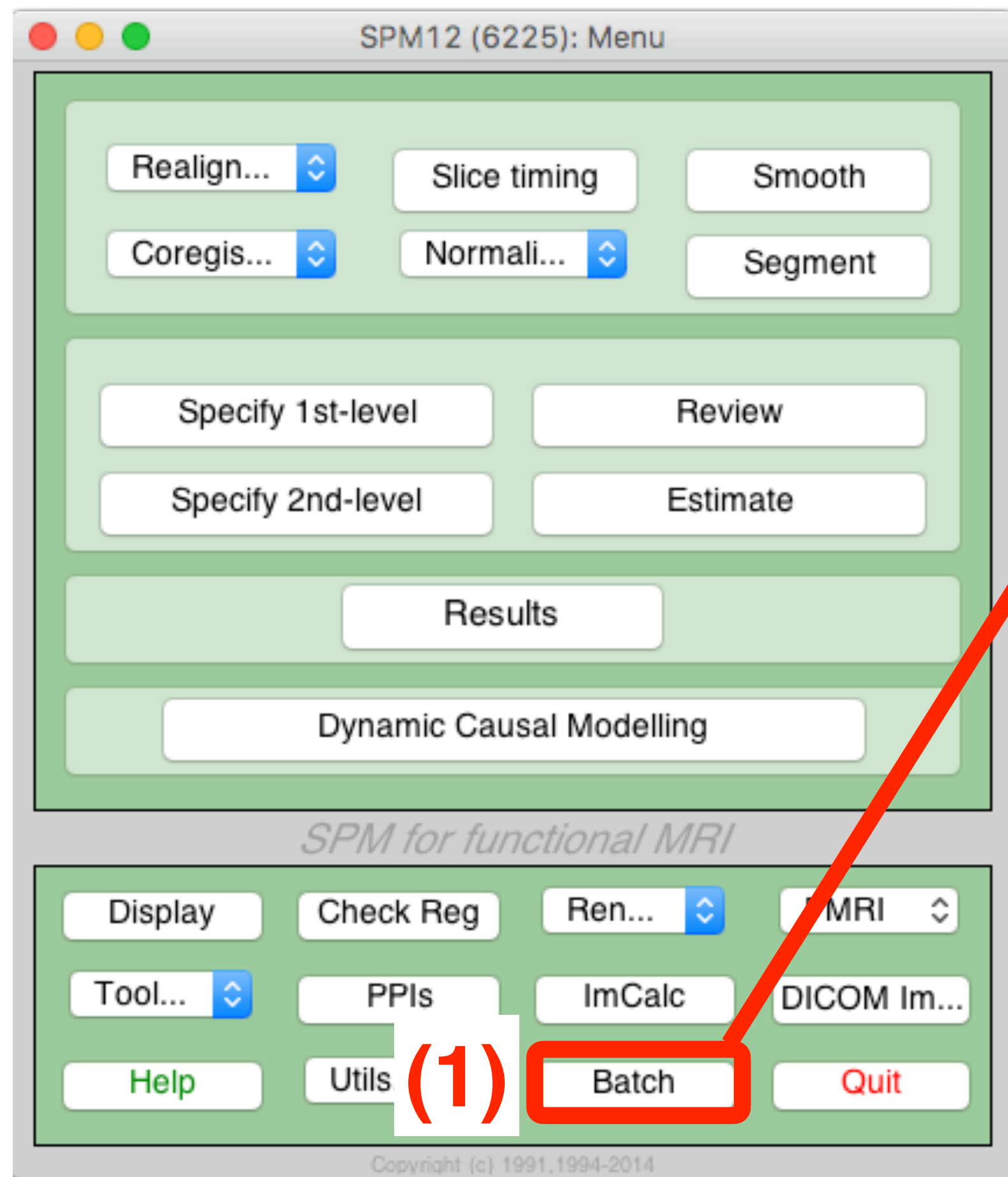
PlugIn modules in SnPM13

- SingleSub: Two Sample T test; 2 conditions, replications
- SingleSub: Simple Regression (correlation); single covariate of interest
- MultiSub: One Sample T test on diffs/contrasts; 1 condition, 1 scan per subject
- MultiSub: Simple Regression (correlation); single covariate of interest, 1 scan per subject
- MultiSub: **Paired T test**; 2 conditions, 1 scan per condition
- MultiSub: Within Subject ANOVA; multiple scans/subject
- **2 Groups**: Test diff of response; 2 conditions, 1 scan per condition
- **2 Groups**: Two Sample T test; 1 scan per subject
- **2 Groups**: Between Group ANOVA; 1 scan per subject

SnPM13 Installation

- SnPM13은 SPM의 plug-in toolbox로 설치 및 실행되는 프로그램입니다.
- 다음의 경로에서 툴박스를 다운로드 합니다.
<http://warwick.ac.uk/snpm>
<http://www.fil.ion.ucl.ac.uk/spm/ext/#SnPM>
- SnPM(SnPM13이 가장 최신버전)을 다운로드 한 후에 압축을 풉니다.
- SnPM13 폴더를 spm8 or spm12 가 설치되어 있는 폴더의 하위 폴더인 toolbox 로 이동시킵니다.
- 이제 SPM을 구동시켜 프로그램이 제대로 설치되었는지 확인합니다.

Non-parametric Analysis



Non-parametric Paired t-test: Specify

Batch editor → SPM → Tools → SnPM → Specify

SingleSub: Two Sample T test; 2 conditions, replications
SingleSub: Simple Regression (correlation); single covariate of interest
MultiSub: One Sample T test on diffs/contrasts
MultiSub: Simple Regression; 1 covariate of interest
MultiSub: Paired T test; 2 conditions, 1 scan per condition
Within Subject ANOVA, k diffs/contrasts per subject
2 Groups: Two Sample T test; 1 scan per subject
2 Groups: Test diff of response; 2 conditions, 1 scan per condition
Between group ANOVA; 1 scan per subject

Reporting Neuroimaging Results

