Faster family-wise error control for neuroimaging with a parametric bootstrap

Vandekar et al (2018)

University of Oxford

February 17, 2021

Subset Pivotality

Hypothesis testing framework

Define test-statistics $\{T_v\}_{v \in \mathcal{V}}$ (one for each voxel).
For $v \in \mathcal{V}$ let H_v denote the **null hypothesis** at voxel v. For $\mathcal{V}_0 \subset \mathcal{V}$, let

$$H_{\mathcal{V}_0} = \bigcap_{v \in \mathcal{V}_0} H_v$$

denote the **intersection null hypothesis**.

Strong versus weak control of the FWER

Definition

Given a voxelwise method $\mathcal{M}$ which rejects the null at v if $T_v > u$ for some threshold u (which is the same at each voxel), we say that $\mathcal{M}$ **weakly controls the FWER** (to a level $0 < \alpha < 1$) if

$$\mathbb{P}\left(\bigcup_{v \in \mathcal{V}} \{T_v > u\} \middle| H_{\mathcal{V}}\right) \leq \alpha$$

and **strongly controls the FWER** if for all $\mathcal{V}_0 \subset \mathcal{V}$,

$$\mathbb{P}\left(\bigcup_{v \in \mathcal{V}_0} \{T_v > u\} \middle| H_{\mathcal{V}_0}\right) \leq \alpha$$

Note that controlling the FDR weakly controls the FWER.

Subset Pivotality

Definition

We say that **subset pivotality** holds (for the test-statistics) if for all subsets $\mathcal{V}_0 \subset \mathcal{V}$ the distribution of

$$\max_{v \in \mathcal{V}_0} T_v | H_{\mathcal{V}_0}$$

and

$$\max_{v \in \mathcal{V}_0} T_v | H_{\mathcal{V}}$$

are the same.

Subset pivotality $\implies$ strong control

Suppose that u_α is the $100(1-\alpha)\%$ quantile of the maximum test statistic, (e.g. obtained using voxelwise permutation/RFT or the bootstrap approach) then

$$\mathbb{P}\left(\bigcup_{v\in\mathcal{V}}\{T_v > u_\alpha\}\middle| H_{\mathcal{V}}\right) = \mathbb{P}\left(\max_{v\in\mathcal{V}} T_v \geq u_\alpha | H_{\mathcal{V}}\right) = \alpha$$

so rejecting using the voxelwise threshold u_α weakly controls the FWER. Subset pivotality is useful as it implies that rejecting at u_α provides strong control of the FWER. This holds because

$$\begin{aligned}\mathbb{P}\left(\bigcup_{v\in\mathcal{V}_0}\{T_v > u_\alpha\}\middle| H_{\mathcal{V}_0}\right) &= \mathbb{P}\left(\bigcup_{v\in\mathcal{V}_0}\{T_v > u_\alpha\}\middle| H_{\mathcal{V}}\right) \\ &\leq \mathbb{P}\left(\bigcup_{v\in\mathcal{V}}\{T_v > u_\alpha\}\middle| H_{\mathcal{V}}\right) = \alpha.\end{aligned}$$

See e.g. Hayasaka and Nichols (2003) for further details.

Does subset pivotality hold in fMRI?

Suppose that we observe a zero mean Gaussian random image $\epsilon : \mathcal{V} \to \mathbb{R}$ and that $\mathcal{V}$ is a finite set of points embedded in $\mathbb{R}^3$. Given some function $\mu : \mathcal{V} \to \mathbb{R}$, let $Z = \mu + \epsilon$. Given some kernel K, for $v \in \mathcal{V}$ let

$$\begin{aligned} Y(v) &= \sum_{v' \in \mathcal{V}} K(v - v') Z(v') \\ &= \sum_{v' \in \mathcal{V}} K(v - v') \mu(v') + \sum_{v' \in \mathcal{V}} K(v - v') \epsilon(v') = \mu^*(v) + \epsilon^*(v). \end{aligned}$$

Then we can consider two different sets of null hypotheses:

$$H_v^o = \{\mu(v) = 0\}$$

and

$$H_v^s = \{\mu^*(v) = 0\}$$

(Can similarly define the intersection hypotheses.) Note that if K has finite support A then for $v \in \mathcal{V}$ if $\mu^*(v') = 0$ for all $v' \in \{v + A\}$ then $\mu(v) = 0$.

Subset pivotality holds with respect to H^s

Suppose that $T_v = \epsilon^*(v)$. Then given $\mathcal{V}_0 = \{v_1, \ldots, v_n\} \subset \mathcal{V}$ suppose that $H^s_{\mathcal{V}_0}$ holds, i.e. $\mu^*(v) = 0$ for all $v \in \mathcal{V}_0$. Then

$$\begin{pmatrix} T_{v_1} \\ \vdots \\ T_{v_n} \end{pmatrix} = \begin{pmatrix} \epsilon^*(v_1) \\ \vdots \\ \epsilon^*(v_n) \end{pmatrix} \sim N(0, \Lambda)$$

for some covariance matrix Λ. If $\mu^*(v) = 0$ for all $v \in \mathcal{V}$ then the above distribution still holds. As such the distribution of the maximum test statistics is the same, i.e. subset pivotality holds.

- Note that subset pivotality does not hold w.r.t. H^o and the smoothed test-statistics. That's because there is leakage of the signal. However may be able to make strong statements up to the support of the kernel.
- This argument relies on the assumption of Gaussianity (which is not reasonable for fMRI data). I'm not sure how well it generalizes (something to discuss).

Model

Set up

Given N i.i.d. subjects, and for each voxel data $Y_v \in \mathbb{R}^N$ (one entry per subject), for each $v \in \mathcal{V} = \{1, \ldots, V\}$, they consider the regression:

$$Y_v = X_0\alpha_v + X_1\beta_v + \epsilon_v = X\zeta_v + \epsilon_v.$$

- Here $\alpha_v \in \mathbb{R}^{m_0}$ and $\beta_v \in \mathbb{R}^{m_1}$(some $m_0, m_1 \in \mathbb{N}$).
- $X_0 \in \mathbb{R}^{N\times m_0}$ and $X_1 \in \mathbb{R}^{N\times m_1}$
- And $X = [X_0, X_1], \zeta_v = (\alpha_v^T, \beta_v^T)^T$.
- Error $\epsilon_v \in \mathbb{R}^N$.
- Let $Y = (Y_1, \ldots, Y_V) \in \mathbb{R}^{N\times V}$.

Spatial Correlation Matrix

Let $\Psi \in \mathbb{R}^{V \times V}$ be the spatial covariance such that

$$\Psi_{v,w} = \text{cov}(Y_{vn}, Y_{wn})$$

where Y_{vn} denotes the nth entry of the vector Y_v. Define the spatial correlation $\Sigma \in \mathbb{R}^{V \times V}$:

$$\Sigma_{v,w} = \frac{\Psi_{v,w}}{(\Psi_{v,v}\Psi_{w,w})^{1/2}}$$

for $v, w \in \mathcal{V}$. Define the variance at each voxel $v \in \mathcal{V}$ to be

$$\sigma_v^2 = \Psi_{v,v}.$$

Estimating the correlation

Let $\hat{\epsilon}_v = (I - P)Y_v \in \mathbb{R}^N$ be the residuals (here $P = X(X^TX)^{-1}X^T$). We can estimate the variance at a voxel v as

$$\hat{\sigma}_v^2 = \frac{1}{N-m}\|\hat{\epsilon}(v)\|^2.$$

and estimate the spatial correlation between two voxels: v, w as

$$\hat{\rho}_{v,w} = \frac{\hat{\epsilon}_v^T \hat{\epsilon}_w}{\hat{\sigma}_v \hat{\sigma}_w} = \frac{1}{N-m}\left(\frac{\hat{\epsilon}_v}{\|\hat{\epsilon}_v\|}\right)^T \left(\frac{\hat{\epsilon}_w}{\|\hat{\epsilon}_w\|}\right).$$

We can estimate Σ using the sample correlation matrix $\hat{\Sigma} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ such that

$$\hat{\Sigma}_{v,w} = \begin{cases} 1 & v = w \\ \hat{\rho}_{v,w} & \text{otherwise} \end{cases}.$$

This is a consistent estimator for Σ.

Testing

At each voxel they are interested in testing

$$H_v : \beta_v = 0.$$

This can be tested using an F-statistic:

$$F_{vN} = \frac{Y_v^T (P - P_0) Y_v / m_1}{Y_v^T (I - P) Y_v / (N - m)} \sim F_{m_1, N-m}$$

where $P = X(X^T X)^{-1} X^T$ and $P_0 = X_0 (X_0^T X_0)^{-1} X_0^T$.

Single Step FWER

In fMRI the standard way to test this is using the Freedman Lane permutation algorithm (e.g. Winkler 2014) which permutes under the assumption that $\beta_v = 0$.
Given B permutations and permuted test-statistics: $T_{v1}, \ldots, T_{vB}$ at each voxel v (note typically take $T_{v1} = T_v$ to be the observed test-statistic).
Single step permutation calculates a corrected p-value at each voxel of

$$\frac{1}{B} \sum_{b=1}^{B} 1[\max_{v \in \mathcal{V}} T_{vb} \geq T_v].$$

Rejecting at a level α, this procedure ensures FWER $\leq \alpha$.
Could also do bootstrapping here - which is the idea of the paper.

PBJ Algorithm

Wishart Distribution

Definition

Given a number of degrees of freedom d and number of parameters V, suppose that we have i.i.d column vectors $g_1, \ldots, g_N$ such

$$g_i \sim N_V(0, \Sigma).$$

Let $G = [g_1, \ldots, g_d] \in \mathbb{R}^{V \times d}$. Then the **Wishart distribution** is the distribution of the $V \times V$ matrix

$$S = GG^T = \sum_{i=1}^{d} g_i g_i^T.$$

We write this as

$$S \sim W_V(d, \Sigma).$$

Wishart Distribution (continued)

Given a $V \times V$ matrix A let diag(A) denote the vector in $\mathbb{R}^V$ corresponding to the diagonal of A.

Definition

We say that a vector $Z \in \mathbb{R}^V$ is **diagonal Wishart** and write

$$Z \sim \text{diag}(W_V(d, \Sigma))$$

if it has the same distribution of the diagonal of the corresponding Wishart.
We talk about **singular** (diagonal) Wishart distributions if $d < V$.

Theorem 1

Define the diagonal matrix $\Phi \in \mathbb{R}^{V \times V}$ such that for $v = 1, \ldots, V$

$$\Phi_{v,v} = \frac{1}{\Psi_{v,v}^{1/2}} = \frac{1}{\sigma_v}$$

Theorem

Assume that $\epsilon_v \sim N(0, \sigma_v^2 I_N)$, *then under the null,*

1. $$\Phi Y^T (P_0 - P) Y \Phi \sim \mathcal{W}_V(m_1, \Sigma)$$

   and

   $$\Phi Y^T P Y \Phi \sim \mathcal{W}_V(m_1, \Sigma)$$

2. $$m_1[F_{1N}, \ldots, F_{VN}] \xrightarrow{d} diag(\mathcal{W}_V(m_1, \Sigma))$$

PBJ algorithm

1. Regress Y onto X to obtain the test-statistics T_v to test H_v (dropping dependence on N). Order these as $T_{(1)}, \ldots, T_{(V)}$. Let

$$E = \left[\frac{\hat{\epsilon}_{(1)}}{\left\| \hat{\epsilon}_{(1)} \right\|}, \ldots, \frac{\hat{\epsilon}_{(1)}}{\left\| \hat{\epsilon}_{(1)} \right\|} \right] \in \mathbb{R}^{N \times V}$$

   be a matrix of the standardized residuals and let $r = \min\{N - m, V\}$.

2. Perform the singular value decomposition:

$$E = U D \tilde{M}^T$$

   where $D \in \mathbb{R}^{r \times r}$ is diagonal and where $M \in V \times r$ and $U \in \mathbb{R}^{N \times r}$ have orthogonal columns. Let $M = \tilde{M} D$.

Explanation interlude

Why they we doing this? Well let $\hat{\epsilon}_{v,n}$ denote the nth entry of $\hat{\epsilon}_v$. Then for $v, w \in \mathcal{V}$

$$(E^T E)_{v,w} = \sum_{n=1}^{N} E^T_{vn} E_{nw} = \sum_{n=1}^{N} E_{nv} E_{nw}$$
$$= \frac{1}{\|\hat{\epsilon}_v\| \|\hat{\epsilon}_w\|} \sum_{n=1}^{N} \hat{\epsilon}_{v,n} \hat{\epsilon}_{w,n} = \frac{\hat{\epsilon}_v^T \hat{\epsilon}_w}{\|\hat{\epsilon}_v\| \|\hat{\epsilon}_w\|} = (N-m)\hat{\rho}_{v,w}.$$

Now since $E = UD\tilde{M}$, we can write

$$E^T E = \tilde{M} D U^T U D \tilde{M}^T = \tilde{M} D^2 \tilde{M}^T = M M^T$$

where $M = \tilde{M} D$. As such if $S \sim N(0, I_r)$ then

$$\text{cov}(MS) = M M^T = E^T E = (N-m)\hat{\Sigma}$$

which has the right spatial covariance (up to $(N-m)$?).

PBJ algorithm continued

4. Choose a number of bootstraps B and for $b = 1, \ldots, B$ generate an $r \times m_1$ matrix S_b such that

$$(S_b)_{i,j} \sim_{iid} N(0, 1).$$

   Then the columns of MS_b are $N(0, \hat{\Sigma})$.

5. For each $b = 1, \ldots, B$, obtain the null test-statistic

$$T_b = \text{diag}(MS_bS_b^TM^T) \in \mathbb{R}^V$$

   $(T_b = (T_{1b}, \ldots, T_{Vb})^T)$.

Stepdown procedures

Single step procedures calculate adjusted p-values at each voxel of

$$\tilde{p}_v = \frac{1}{B}\sum_{b=1}^{B} 1[\max_{v\in\mathcal{V}} T_{vb} \geq T_v].$$

Step down procedures instead, for $v = 1, \ldots, V$ and $b = 1, \ldots, B$,

1. Compute $T_{vb}^{\max} = \max_{1\leq k\leq v} T_{(k)b}$
2. Compute

$$p^*_{(v)} = \frac{1}{B}\sum_{b=1}^{B} 1\left[T_{vb}^{\max} \geq T_{(v)}\right]$$

3. Calculate adjusted p-values of

$$\tilde{p}_{(v)} = \max_{k\leq v} p^*_{(v)}$$

Power increase

- It can be shown (see Westphal and Young (1993)) that step down procedures still control the FWER. (I think it's strong under subset pivotality - need to check though)
- The step down adjusted p-value is lower than the single step one so step down methods are always more powerful. (e.g. only $v = V$ is always the same as the single step procedure.)
- Vandekar's paper is the first to do so in fMRI, maybe as voxelwise is too computational?
- They claim that the more regions/voxels the greater the benefit from using a step-down procedure (if it is feasible that is)
- Holm's procedure is the step down version of using Bonferroni. It's also the closure of Bonferroni, are these equivalent?

Results

The Data

- 972 subjects, ages 8–21, from the Philadelphia Neurodevelopmental Cohort.
- For each subject have a $V \times V$ image of Cerebral Blood flow data (Tom's cue) for 127756 gray matter voxels.
- They smooth the data with 6mm FWHM.
- They consider voxelwise and regionwise analyses. For the regionwise they divide the brain into 112 regions and average the CBF within those regions (I think).

Simulations

They bootstrap resample the data to generate realistic data simulations (1000 for regionwise and 500 for voxelwise). They use this to generate samples of size 40,100,200,400.
At each voxel (or region - when doing regionwise) they fit the model:

$$Y_{vn} = \alpha_0 + \alpha_1 \text{age}_n + \alpha_2 \text{sex}_n + \alpha_3 \text{race}_n + \alpha_4 \text{MRD}_n + \sum_{j=1}^{3} \beta_{jv} g_j$$

where n denotes the nth subject where g_j are indicators that represent different clinical groups in equal proportions.
Interested in testing:

$$\text{DOF1} : H_0 : \beta_{1v} = 0$$

and

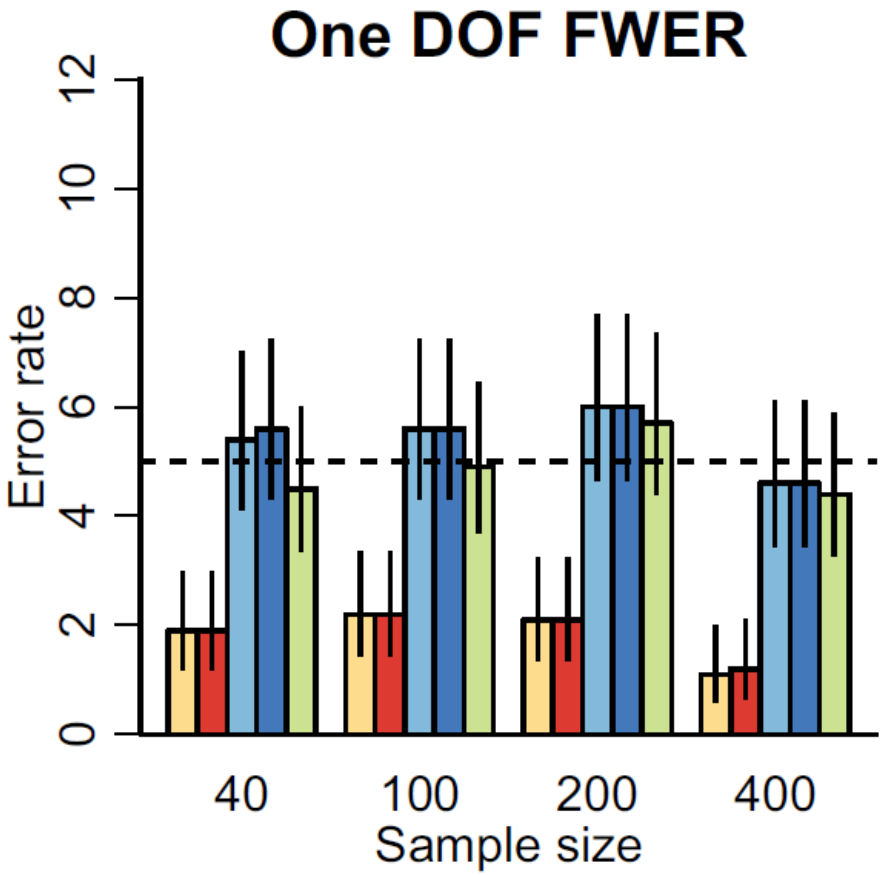
$$\text{DOF3} : H_0 : \beta_{jv} = 0 \text{ for all } j$$

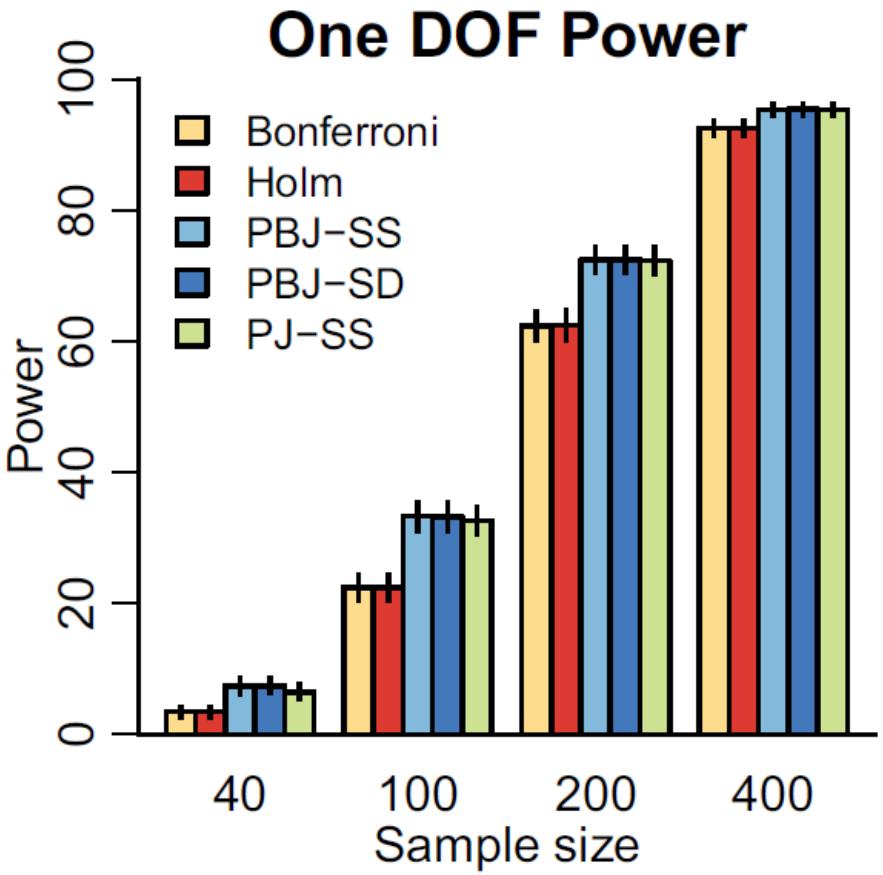
Testing Power

In addition to using the sims to test the FWER control, they also consider looking at power.

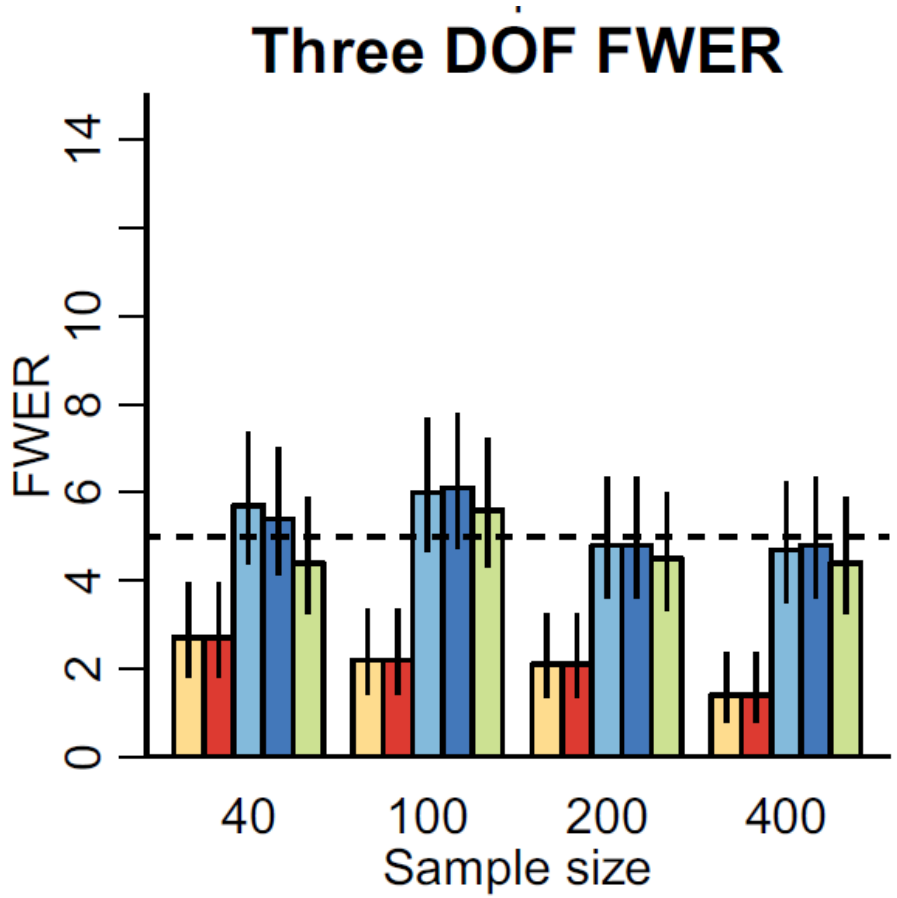
- For regionwise, they randomly selected 3 brain regions and for those set $\beta_{1v} = 10$ (and set the rest to 0) (for both DOF 1, 3)
- For voxelwise they selected a random gray matter voxel v_0, created a cube with a radius of 6 voxels centered at v_0 and set β_{1v} for the voxels within the cube. (They then smooth this image and add it to the data.)

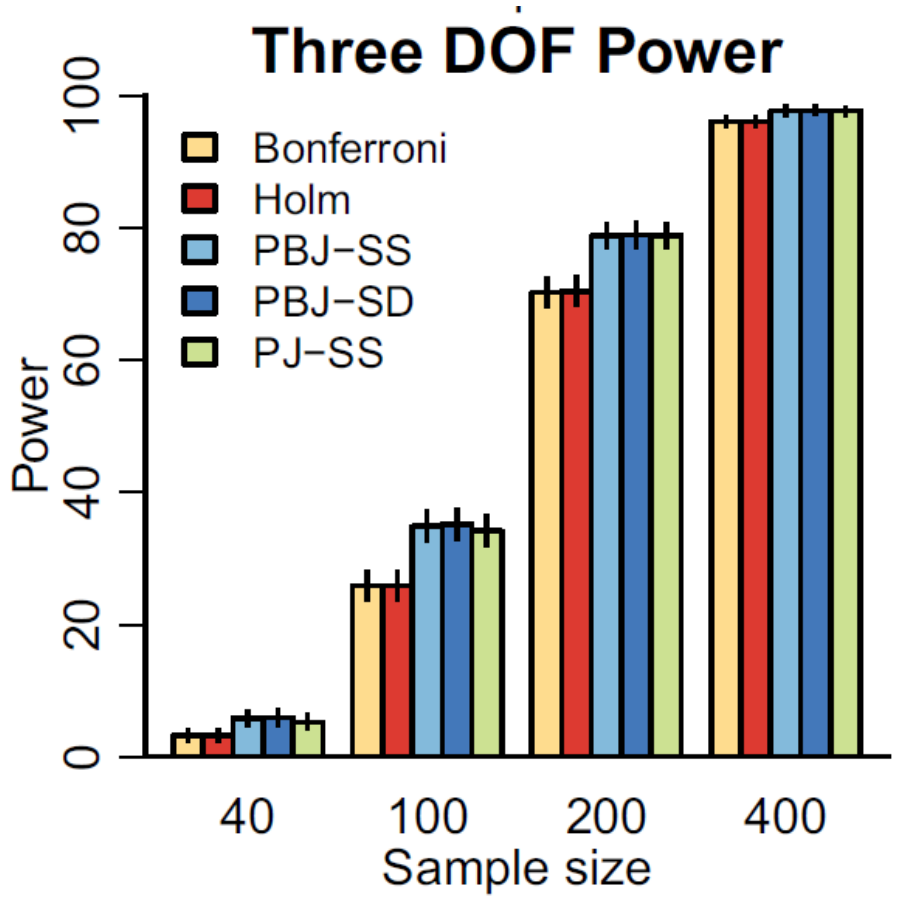
Regionwise results - DOF 1



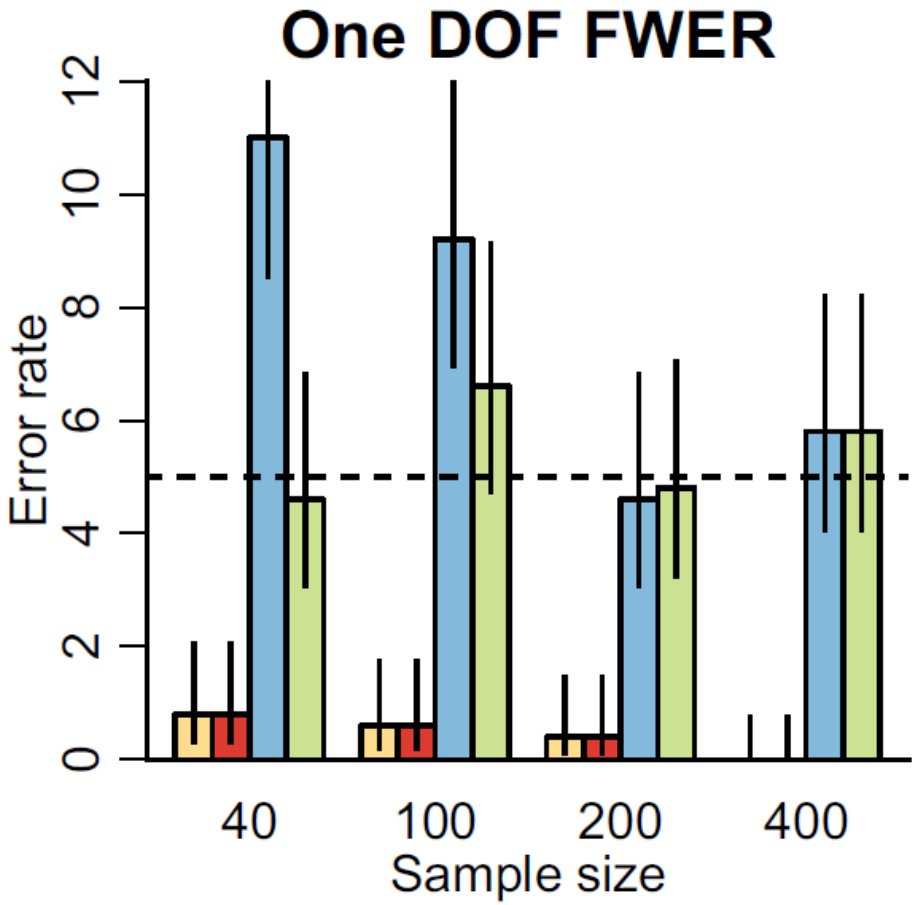


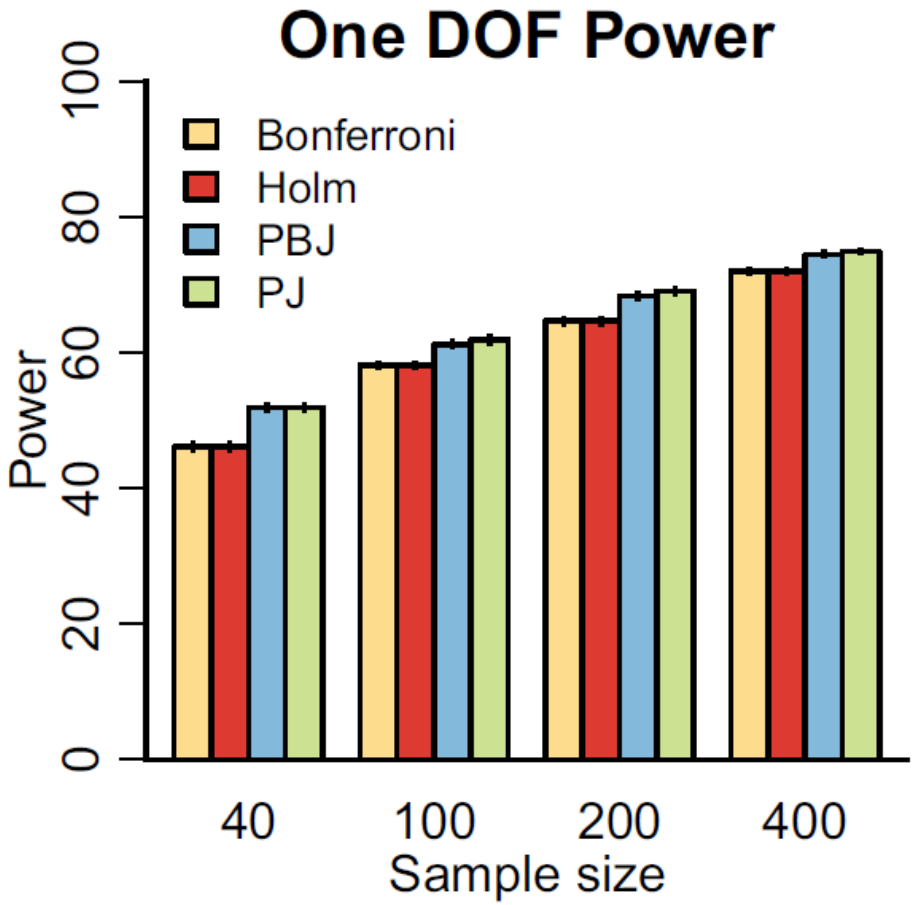
Regionwise results - DOF 3



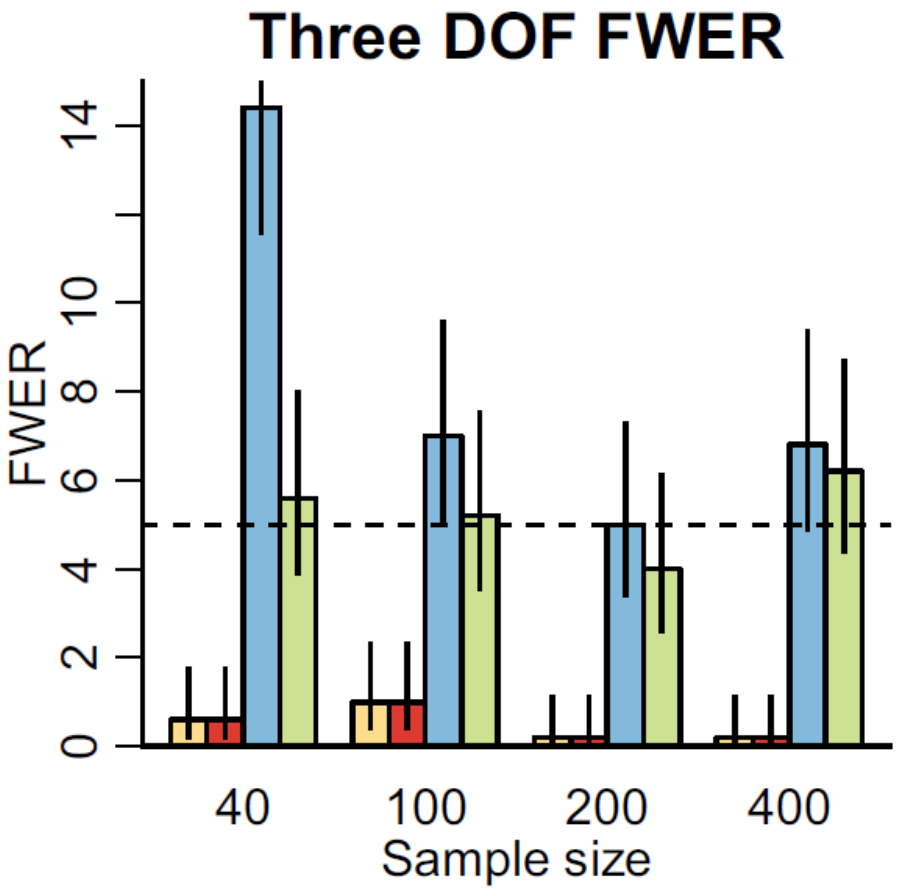


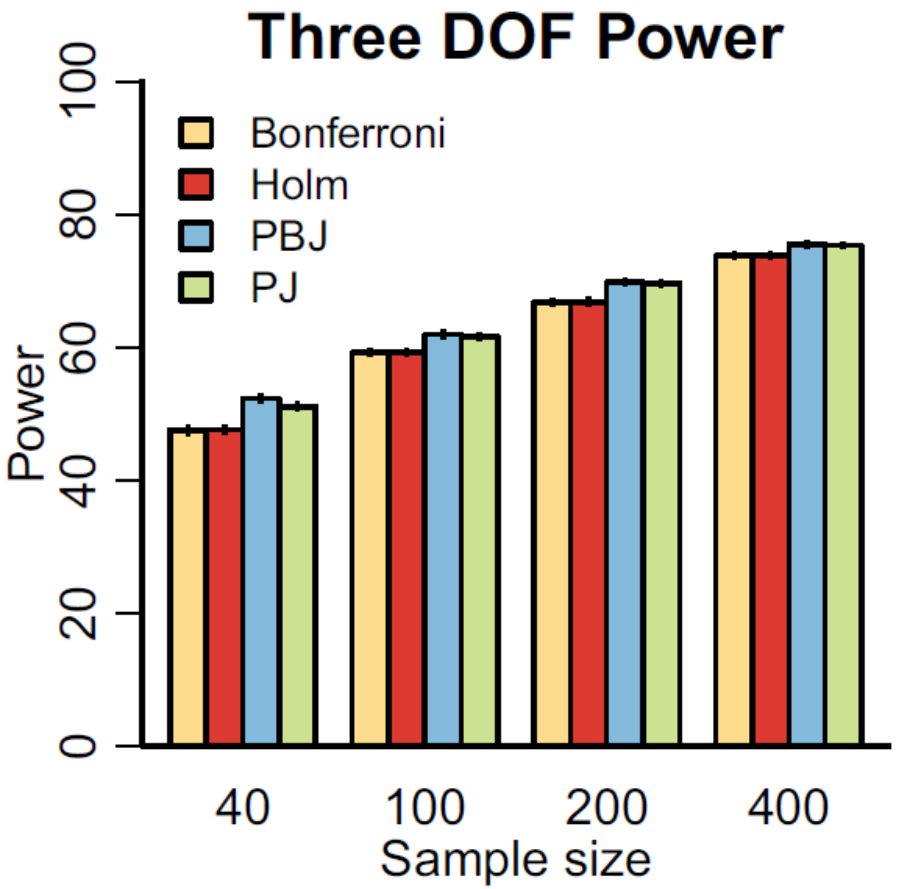
Voxelwise results - DOF 1



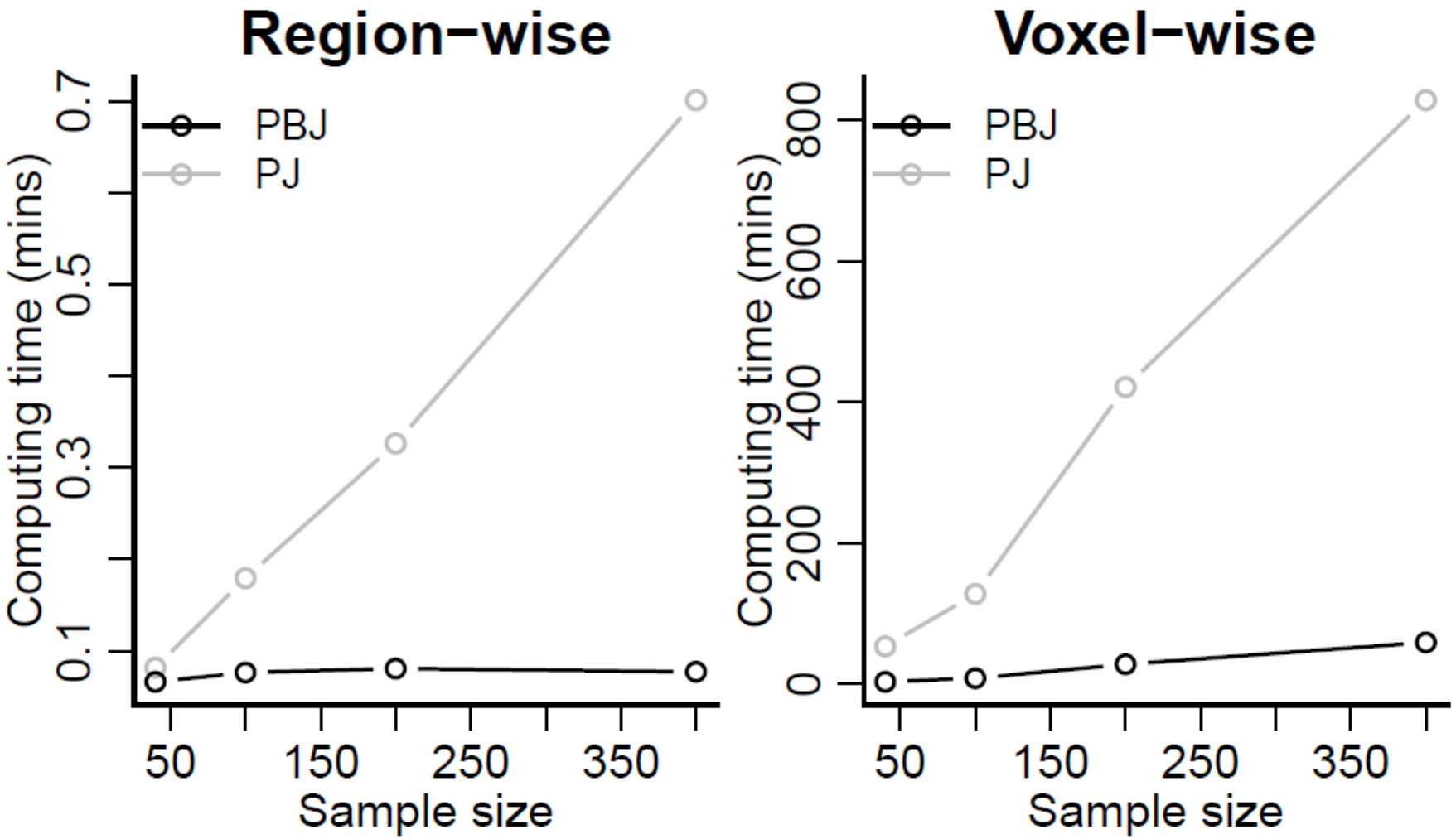


Voxelwise results - DOF 3





Time difference



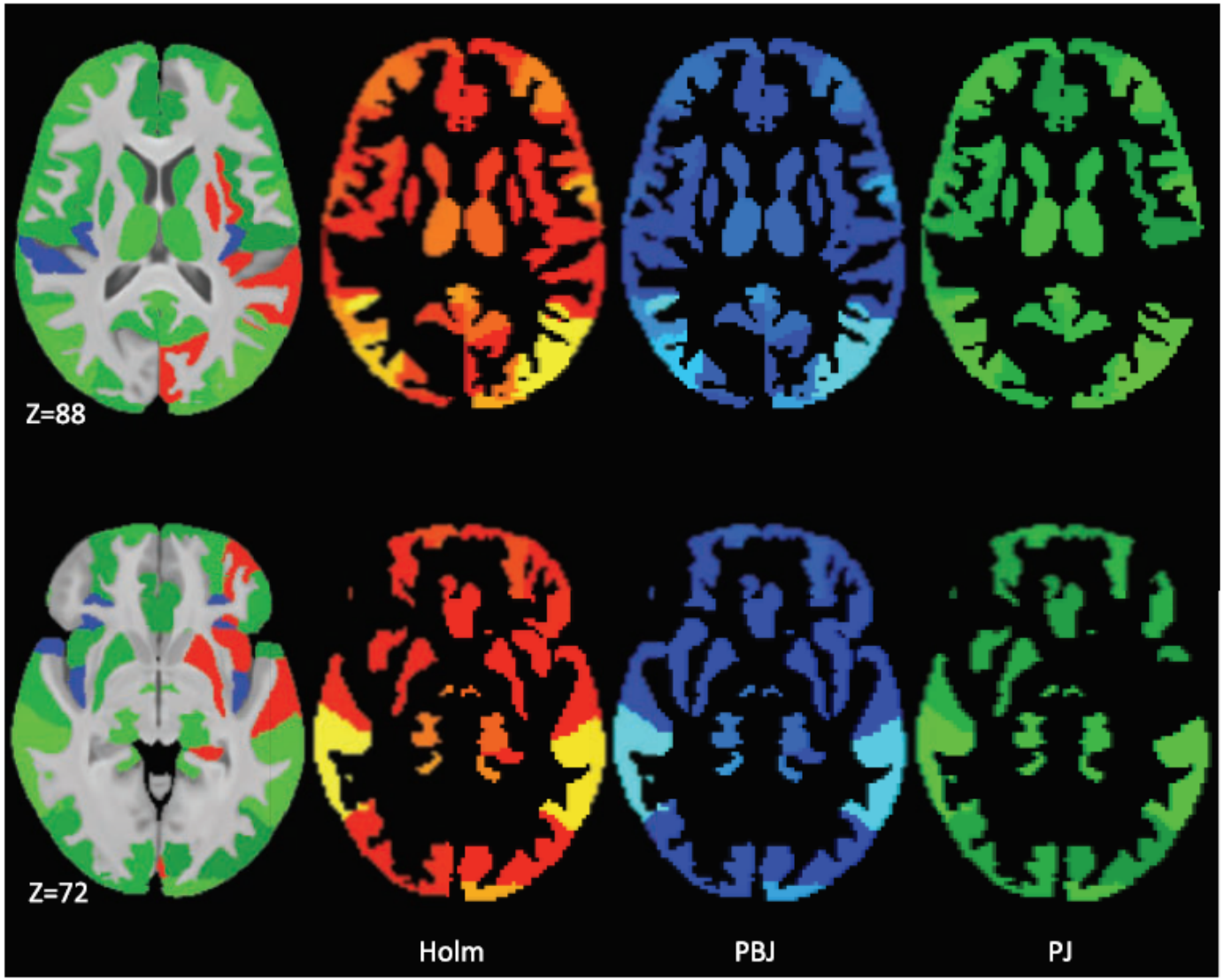


Fig. 3. FWER controlled results at $\alpha = 0.01$ for Holm, PBJ step-down, and PJ single-step for the region-wise analysis. Color scale is $-\log_{10}(p)$ and shows results greater than 2. The left-most images show the overlay of PBJ, Holm, and

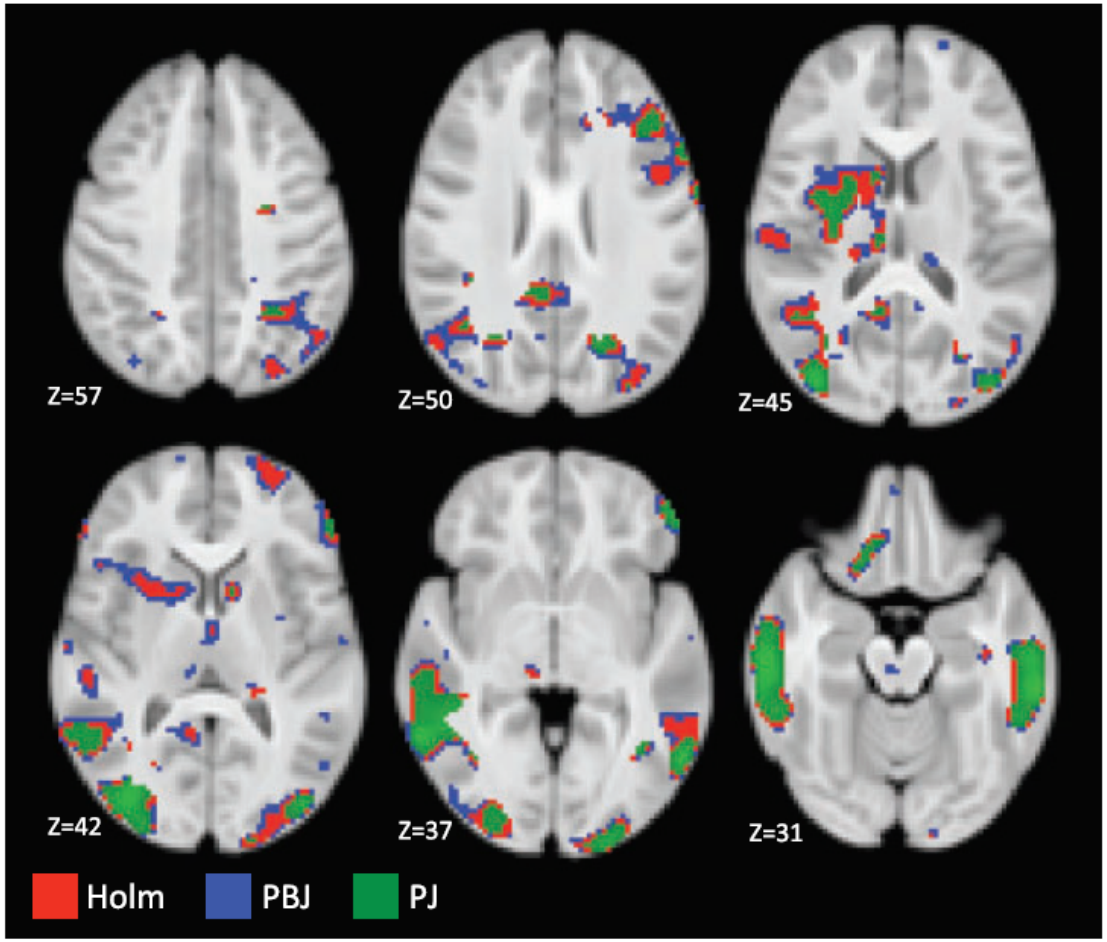


Fig. 4. FWER controlled results at $\alpha = 0.05$ for Holm, PBJ single-step, and PJ single-step for the voxel-wise analysis. Color scale is $-\log_{10}(p)$ for the adjusted p-values and shows results greater than 1.3. The overlay order is PBJ under

Overview

- Much faster - but requires largeish N to ensure false positive control
- Step down procedures should be used (when doing either bootstrap or permutation regionwise) (Not possible voxelwise?)
- Haven't discussed but they transformed the data using a Yeo-Johnson transformation to help improve Gaussianity.

Bibliography