Smooth logistic mass univariate inference for MS lesion data using sign-flipping

Samuel Davenport, Riccardo de Santis, Jesse Hemerik, Jelle Goeman, Livio Finos

University of California, San Diego

June 18, 2024

Example Lesion Images

We have Lesion data from 238 subjects with MS

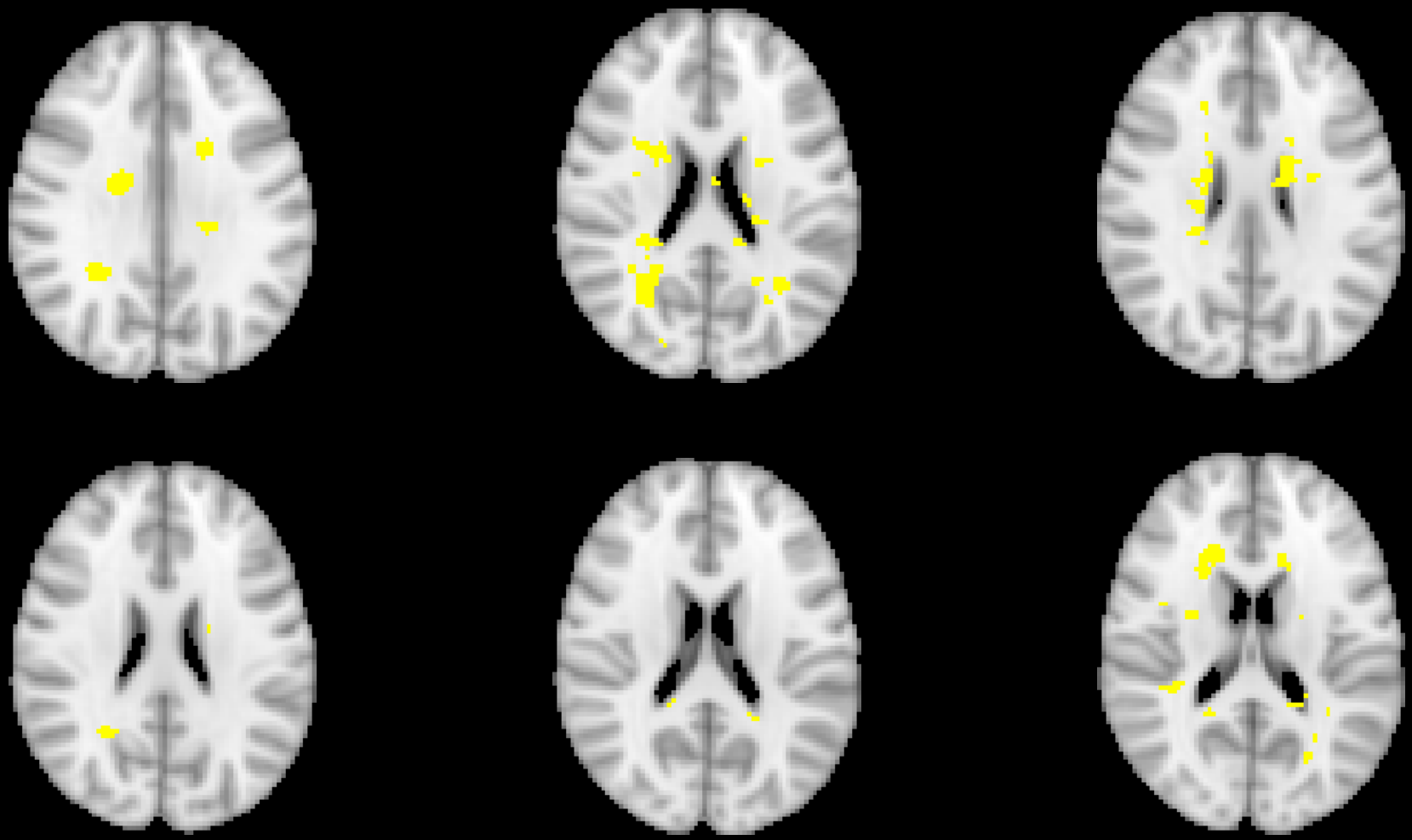
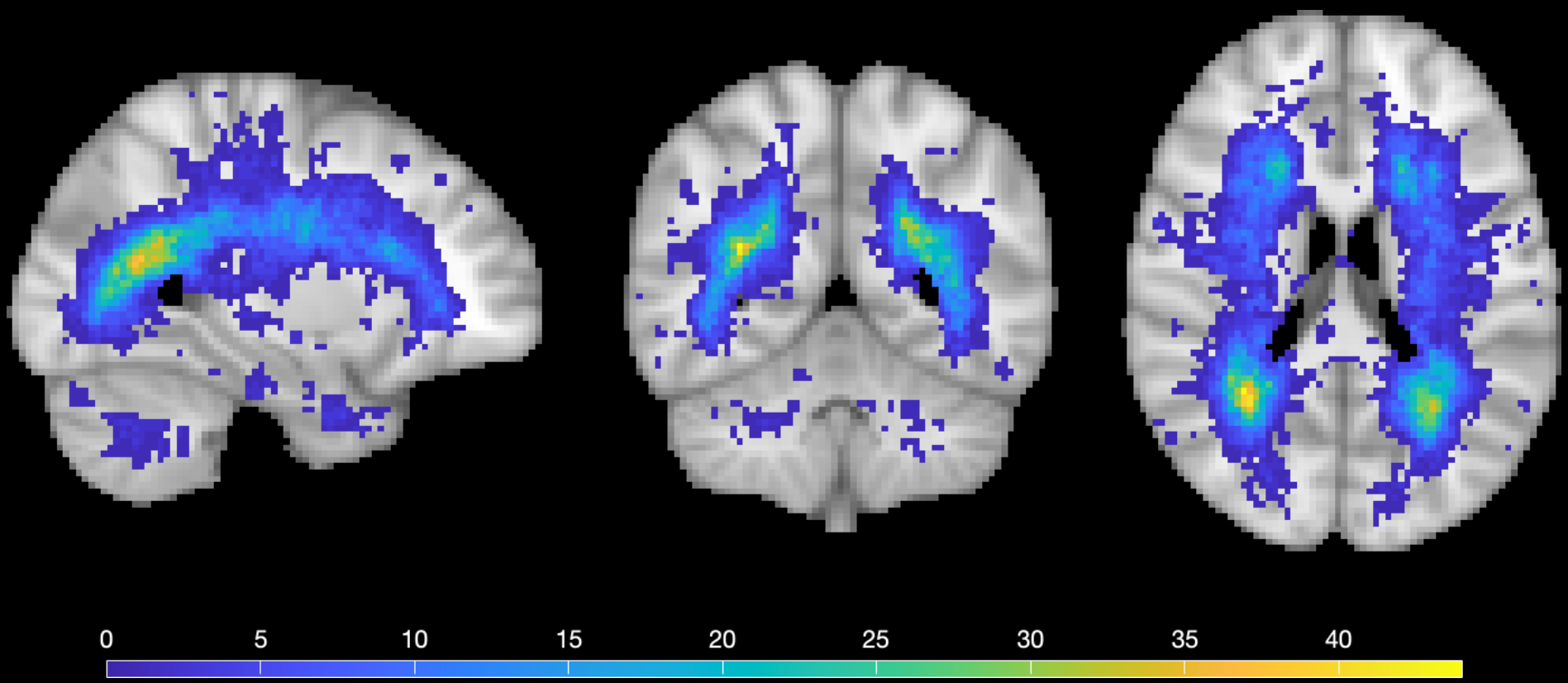


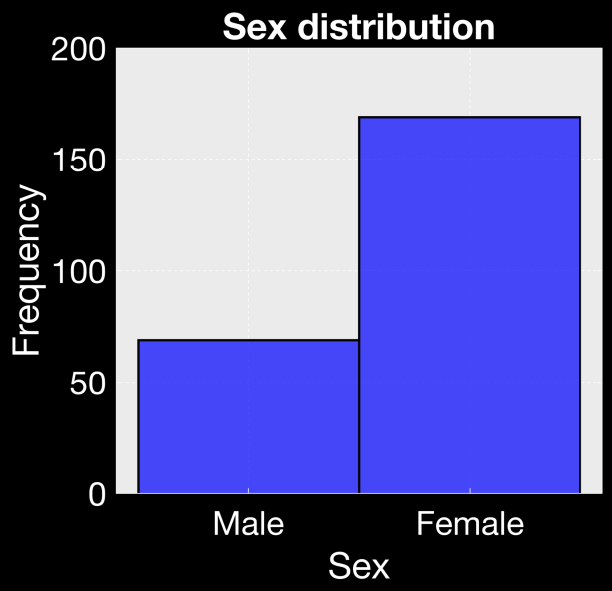
Figure 1: Brain lesions from 6 example subjects

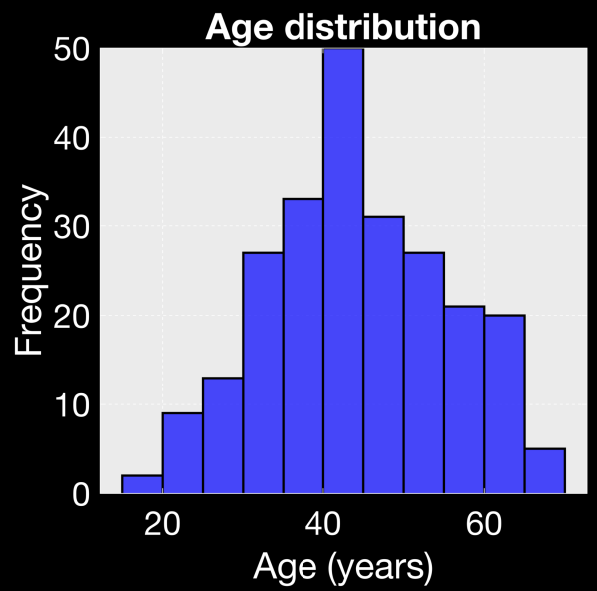
Lesion distribution over all 238 subjects

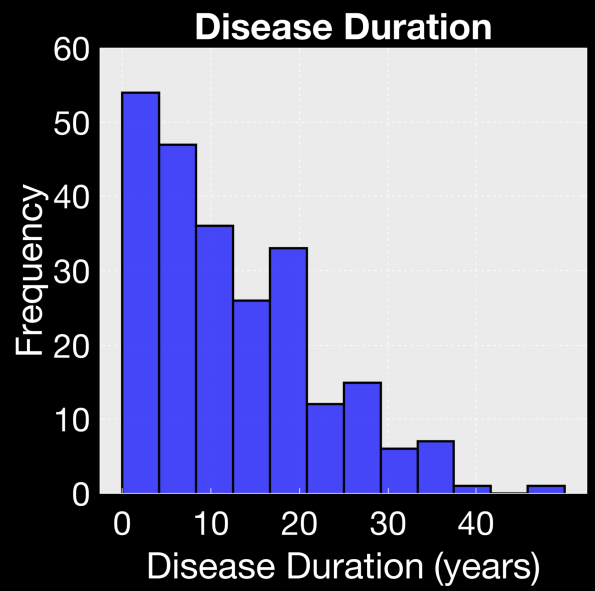


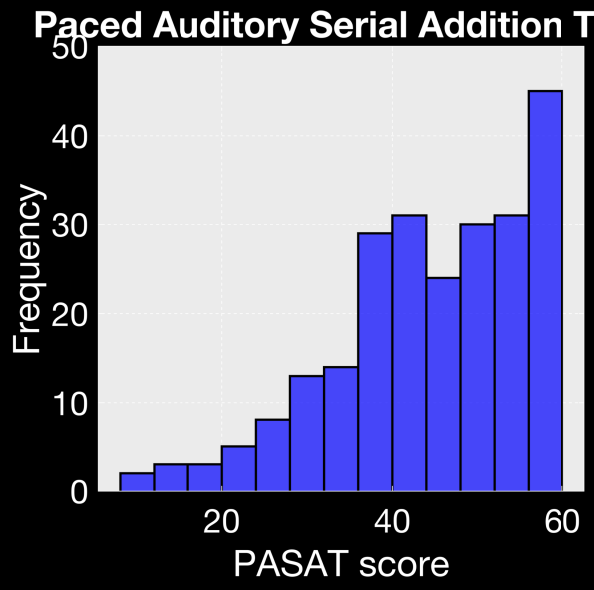
We shall fit lesion count against some covariates of interest.

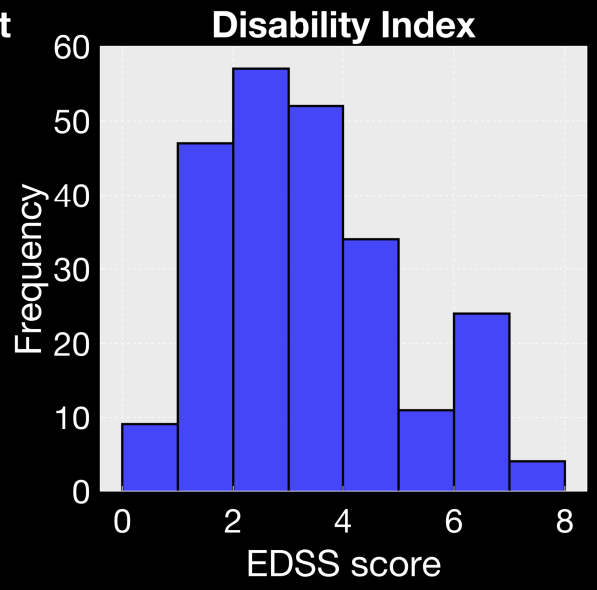
Covariates

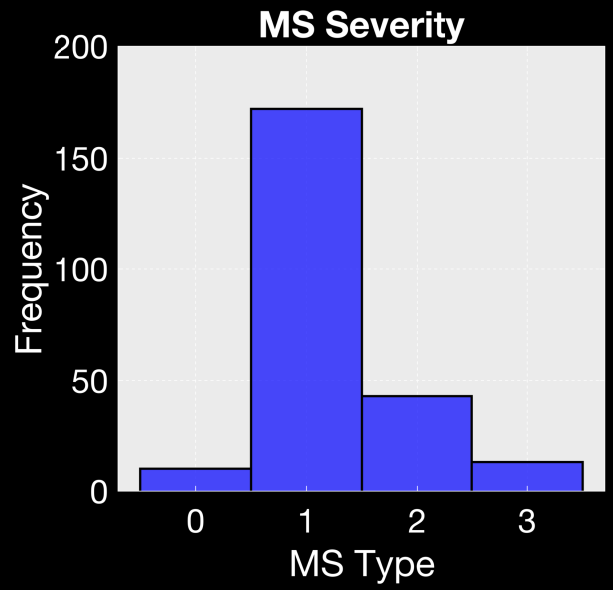

Sex distribution
Frequency
200
150
100
50
0
Male
Female
Sex


Age distribution
Frequency
50
40
30
20
10
0
20
40
60
Age (years)


Disease Duration
Frequency
60
50
40
30
20
10
0
0
10
20
30
40
Disease Duration (years)


Paced Auditory Serial Addition Test
Frequency
50
40
30
20
10
0
20
40
60
PASAT score


Disability Index
Frequency
60
50
40
30
20
10
0
0
2
4
6
8
EDSS score


MS Severity
Frequency
200
150
100
50
0
0
1
2
3
MS Type

Model set up

Let $\mathcal{L}$ be the set of voxels and assume $Y_i(l) \sim \text{Binomial}(q_i(l))$ where $q_i : \mathcal{L} \to \mathbb{R}$ and

$$\log\left(\frac{q_i(l)}{1-q_i(l)}\right) = x_i^T\beta(l) + z_i^T\gamma(l) \tag{1}$$

At each voxel $l \in \mathcal{L}$, we will want to test the null hypothesis

$$H_0(l) : \beta(l) = 0.$$

This results in a very large multiple testing problem and so we shall seek to control the FWER over voxels.

Calculating the effective scores

At each $l \in \mathcal{L}$ let $S_n(l)$ be the effective score at voxel l. Then it turns out that we can write

$$S_n(l) = n^{-1/2} \sum_{i=1}^{n} \nu_i(l).$$

as the sum of score contributions for each subject. Importantly, under the null hypothesis that $\beta(l) = 0$,

$$\{S_n(l)\}_{l \in \mathcal{L}}$$

converges in distribution.

Smoothing the effective scores

In order to increase SNR, we can apply smoothing to the effective scores. Given a smoothing kernel K, let

$$\tilde{\nu}_i(l) = \sum_{l' \in \mathcal{L}} K(l - l')\nu_i(l).$$

We then consider the test-statistic:

$$T_n(l) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\nu}_i(l).$$

Sign-flipping the effective scores

It is possible to show that $\{T_n(l)\}_{l\in\mathcal{L}} \overset{d}{\Longrightarrow} N(0, G)$, some unknown G. In order to infer on the limiting distribution we use sign-flipping. In particular let

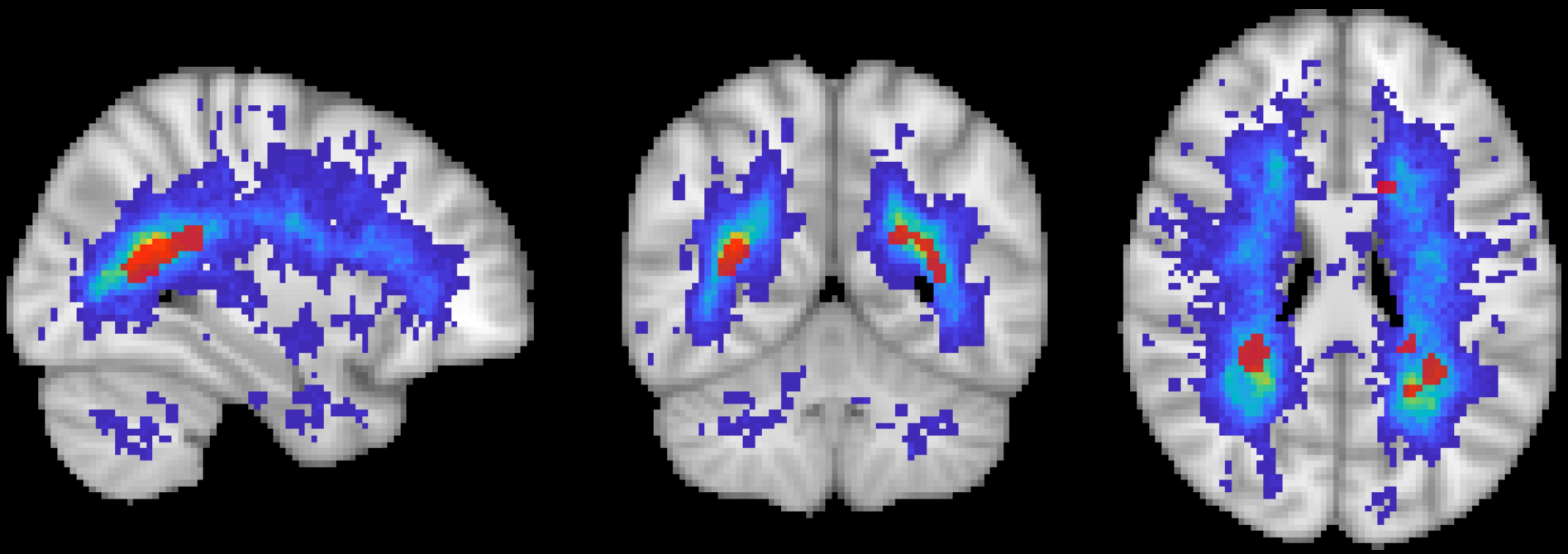
$$T_n^b(l) = n^{-1/2} \sum_{i=1}^{n} g_{bi} \tilde{\nu}_i(l),$$

Where g_{bi}, $1 \leq b \leq B, 1 \leq i \leq n$ are i.i.d. from $\{-1, 1\}$. We show that

Theorem: $\{T_n^b(l)\}_{l\in\mathcal{L}} \overset{d}{\Longrightarrow} N(0, G)$.

Application to the MS lesion dataset - FWHM 4 voxels

Let Q be the 95% quantile of the sign-flipped distribution of $\max_{l \in \mathcal{L}} T_n$ to control the FWER, rejecting $H_0(l)$ if $T_n(l) > Q$.



Conclusions

- Our approach allows resampling in the context of multiple generalized linear models. Further methodological details are available in the SIS submission and in our other paper *Permutation-based multiple testing when fitting many generalized linear models* available on arxiv and at `sjdavenport.github.io/research/`.

-In particular allows smoothing to be combined into the framework which helps to increase detection power.

- Slides for this talk are available on my website: `sjdavenport.github.io/talks`

- Code to implement these methods are available in the flipscores R package, the pyperm python package and the matperm matlab package.

Further theory

Theorem: Let $\mathcal{N}(K)$ be the null set up to the support of the kernel K. Then

$$\lim_{n \to \infty} \mathbb{P}\left(|\mathcal{R}_n \cap \mathcal{N}(K)| > 0\right) \leq \alpha.$$

False Positive Rate Comparison

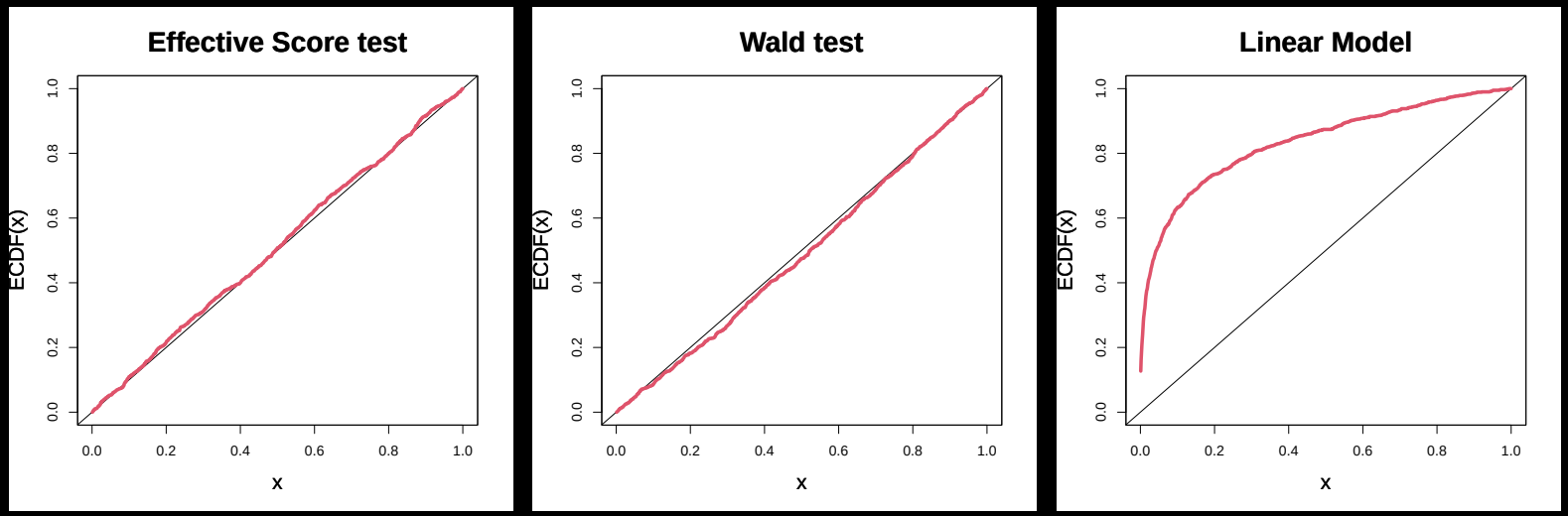
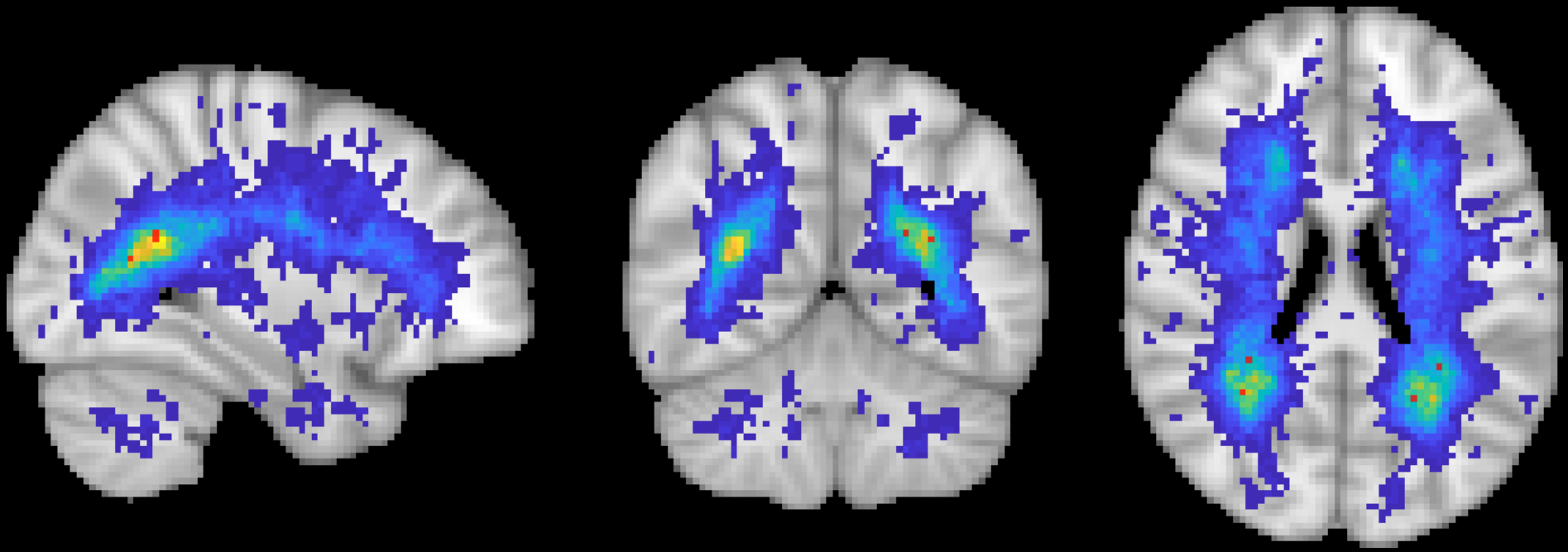


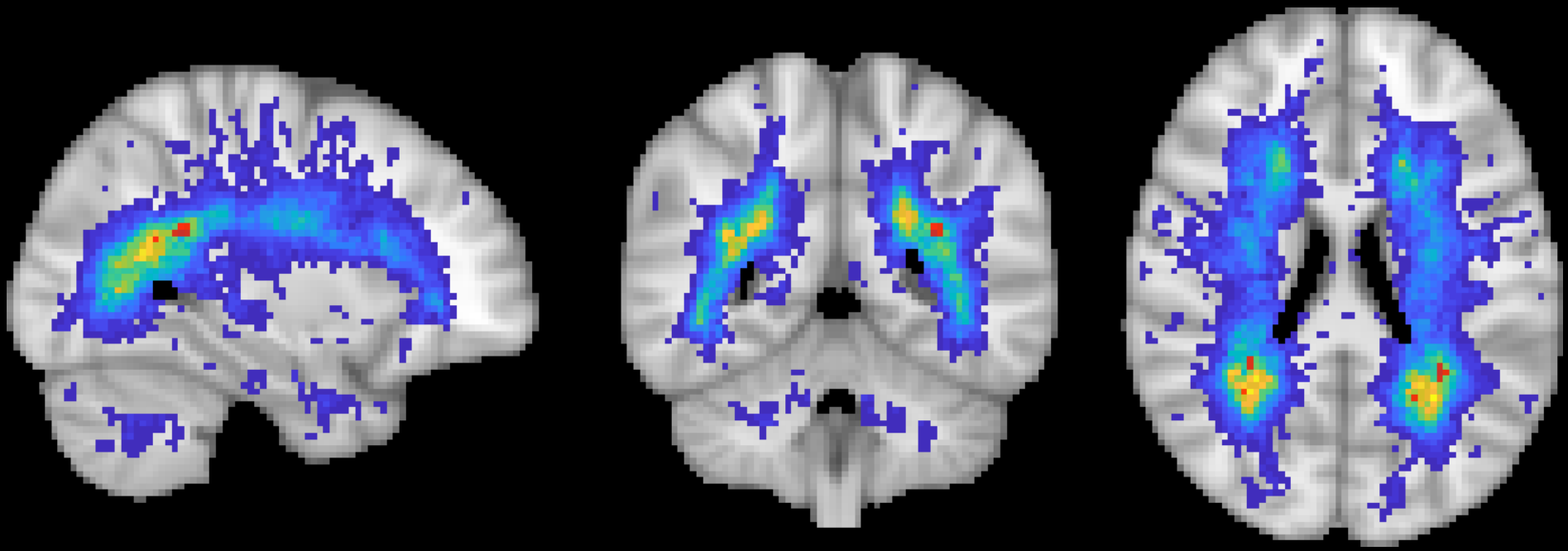
Figure 2: Empircal CDF of the simulated p-values. Fitting a linear model to the data results in high levels of false positives.

The model is: $(z_{i2}, z_{i3})_{i=1}^{n} \overset{i.i.d.}{\sim} N(0, I_2)$ and take $z_{i1} = 1$ and $x_{i1} = z_{i1} z_{i2}$. We take $\gamma = [-5, 1, 5]$ and let $\beta = 0$.

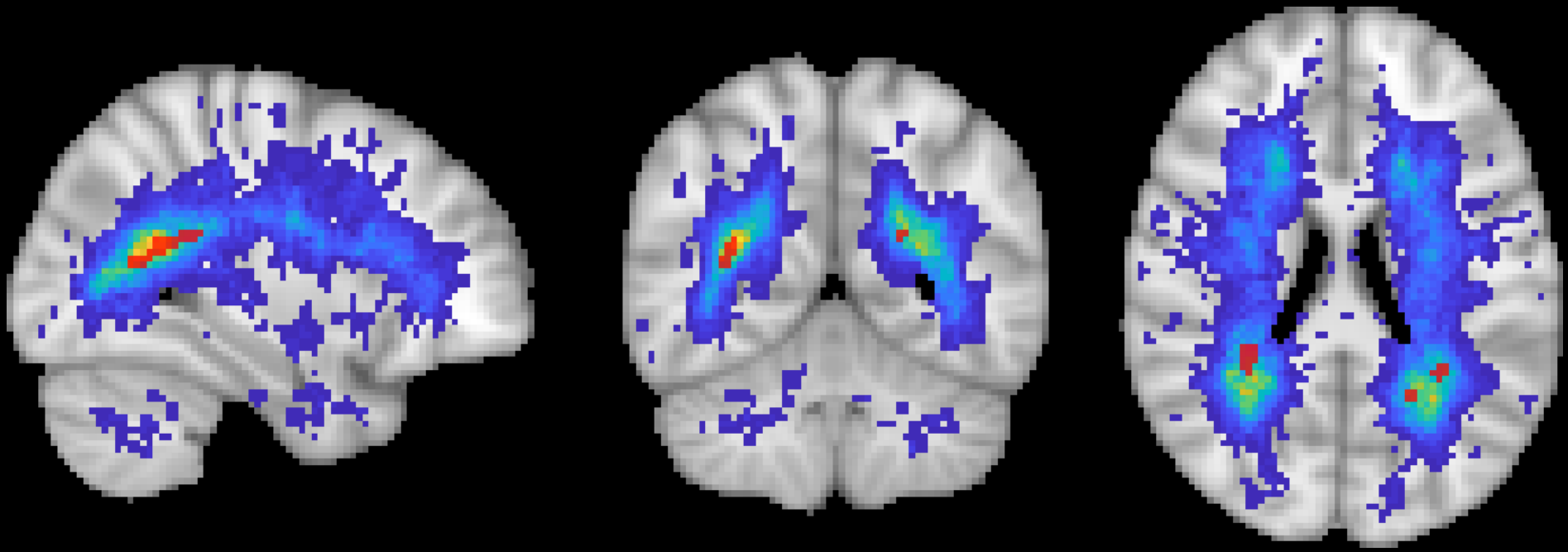
Application to the MS lesion dataset - FWHM 0 voxels



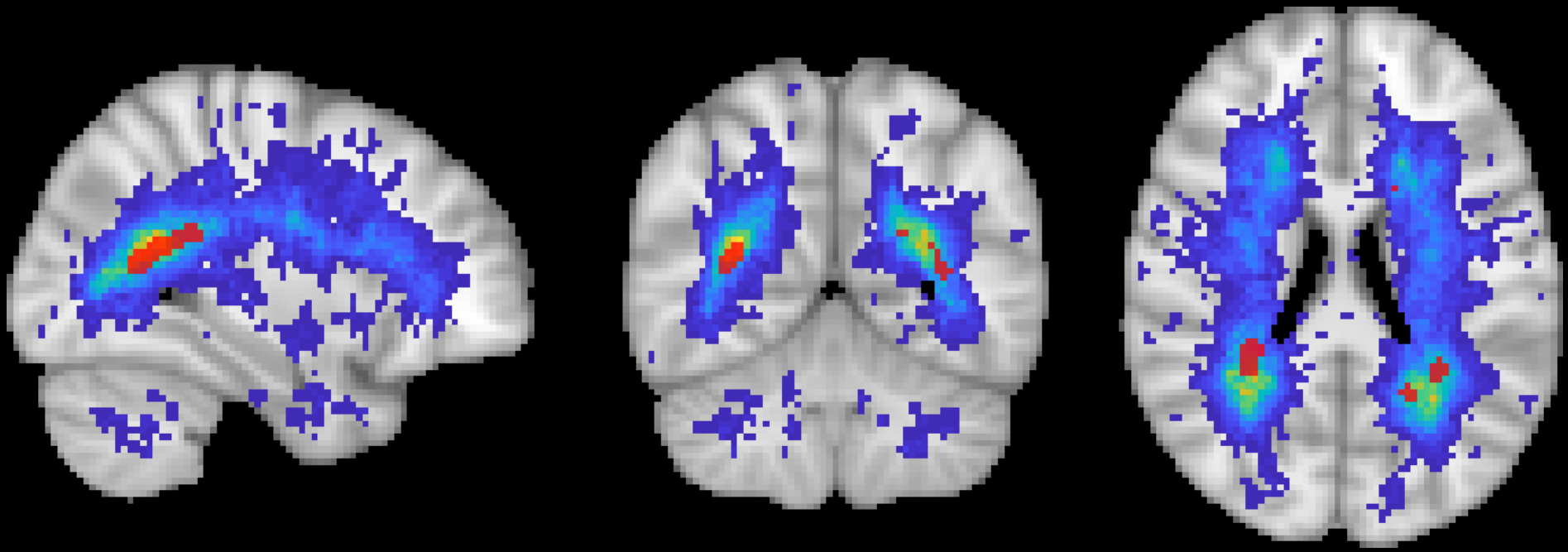
Application to the MS lesion dataset - FWHM 1 voxels



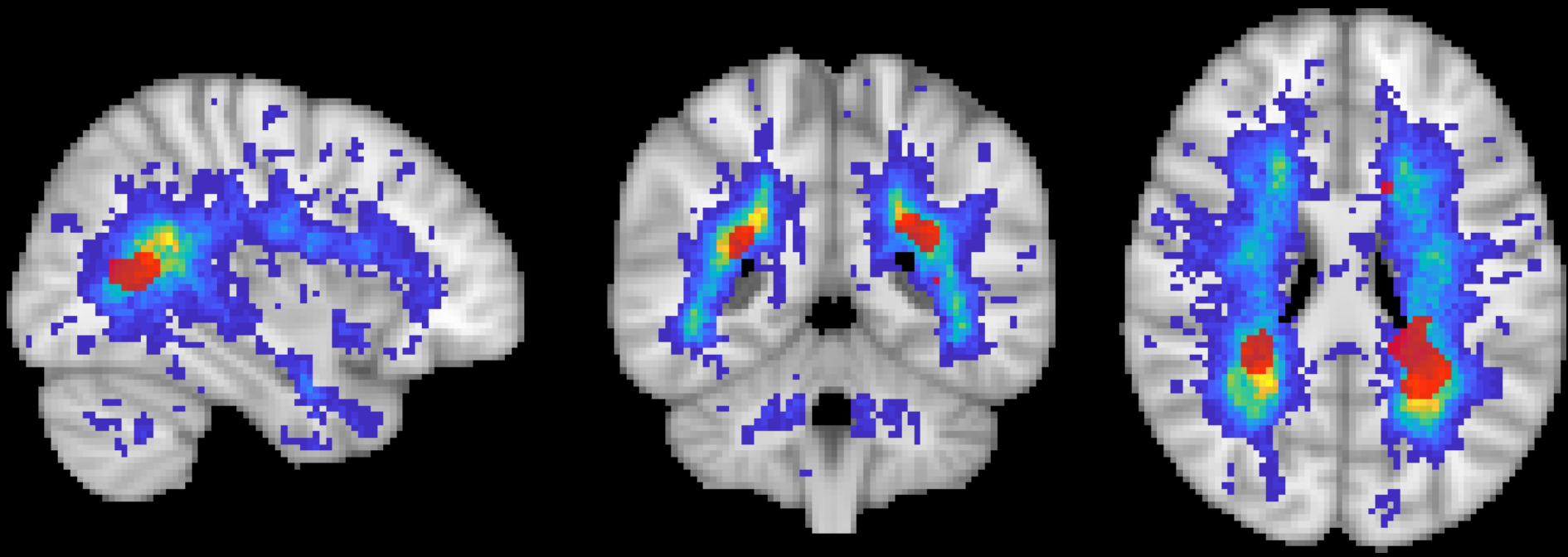
Application to the MS lesion dataset - FWHM 2 voxels



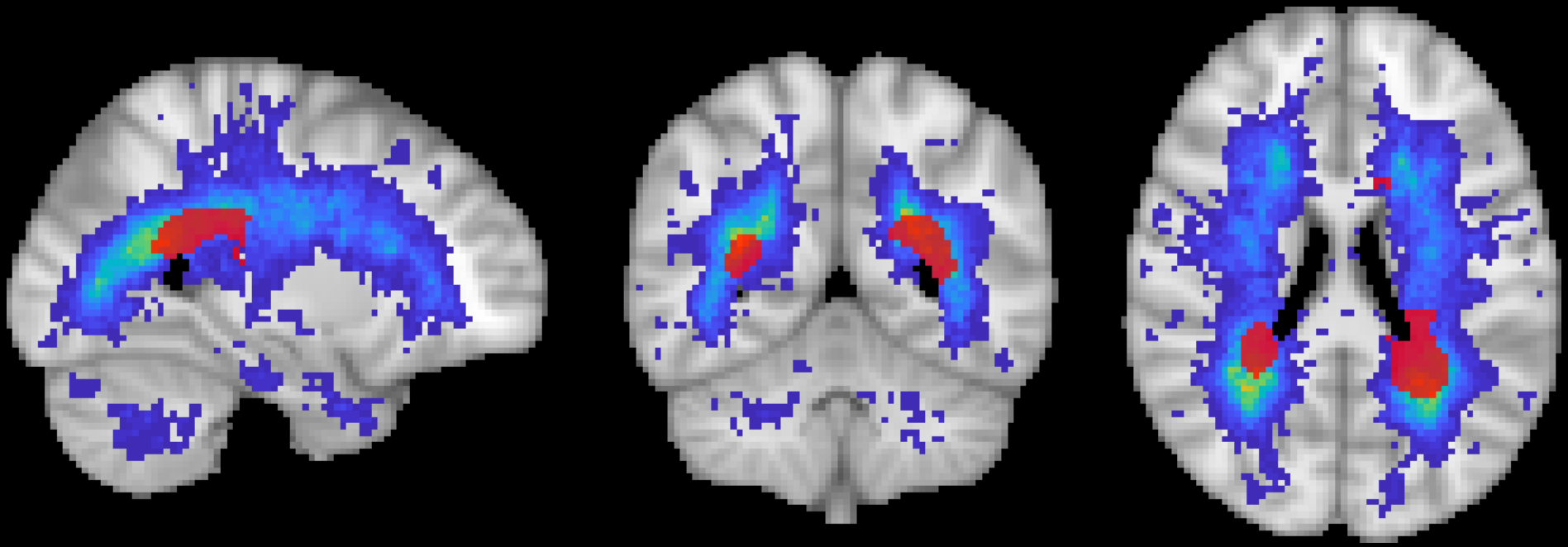
Application to the MS lesion dataset - FWHM 3 voxels



Application to the MS lesion dataset - FWHM 5 voxels



Application to the MS lesion dataset - FWHM 6 voxels



Application to the MS lesion dataset - FWHM 7 voxels