Inflated false positive rates in fMRI depend on the voxel size of normalized images

Karsten Mueller[1] , Jöran Lepsien[1] , Harald E Möller[1] , Gabriele Lohmann[2,3]

[1]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [2]Max Planck Institute for Biological Cybernetics, Tuebingen, Germany, [3]Department of Biomedical Magnetic Resonance, University Hospital Tübingen, Tübingen, Germany

karstenm@cbs.mpg.de

Introduction

- In a recent manuscript, Eklund et al. [1] reported inflated false positive rates in functional magnetic resonance imaging (fMRI) using common software packages including SPM, FSL, and AFNI.
- Briefly, a nominal family-wise error rate of 5% in the parametric statistical evaluation was shown to be conservative for voxel-wise inference but not for cluster-wise inference.
- As a cause of the observed invalid cluster inferences, the authors suggested that the spatial autocorrelation functions do not follow the assumed Gaussian shape.
- We would like to draw attention to an important aspect that was not addressed in this publication: ***Statistical inferences obtained using the Gauss random field approach depend heavily on a pre-processing parameter that was not included in the analysis performed by Eklund et al. [1], namely the spatial resolution to which the data are resampled and interpolated during pre-processsing.***
- Eklund et al. [1] used the common default setting of 2×2×2 mm³.
- In response to the paper by Eklund et al., Flandin and Friston [2] used a different setting of this parameter, namely 3×3×3 mm³. Together with a more stringent initial cluster-forming threshold, they did not observe inflated false positive rates.
- However, a spatial resolution of 2×2×2 mm³ is the default value in two major software packages (SPM, FSL) and, hence, it is likely to be used for processing fMRI data by these packages.
- In previous work, Friston and colleagues [3] stated that resampling to 2×2×2 mm³ renders the analysis “more sensitive”.
- It is, thus, completely unclear what a valid setting for this parameter should be. Therefore, it is of substantial relevance to systematically assess its influence on statistical inference.

Methods

- We analyzed 47 resting-state fMRI data sets, each acquired at a nominal spatial resolution of **3×3×4 mm³** with 300 volumes.
- Using a strategy analogous to that of Eklund et al. [1] we imposed various fake designs including block- and event-related types.
- Analysis was performed using SPM12.
- Normalization included a resampling with three different voxel sizes: 3×3×3 mm³, 2×2×2 mm³, and 1×1×1 mm³.
- Two normalization pipelines: (a) normalizing the raw data and performing the individual statistics with normalized data, and (b) normalizing the contrast images.
- Using a family-wise error (FWE) correction for multiple comparisons based on the Gauss Random Field approach, we first evaluated each data set separately. Thereafter, we performed a group-level inference in which all 47 data sets were pooled.

Results & Discussion

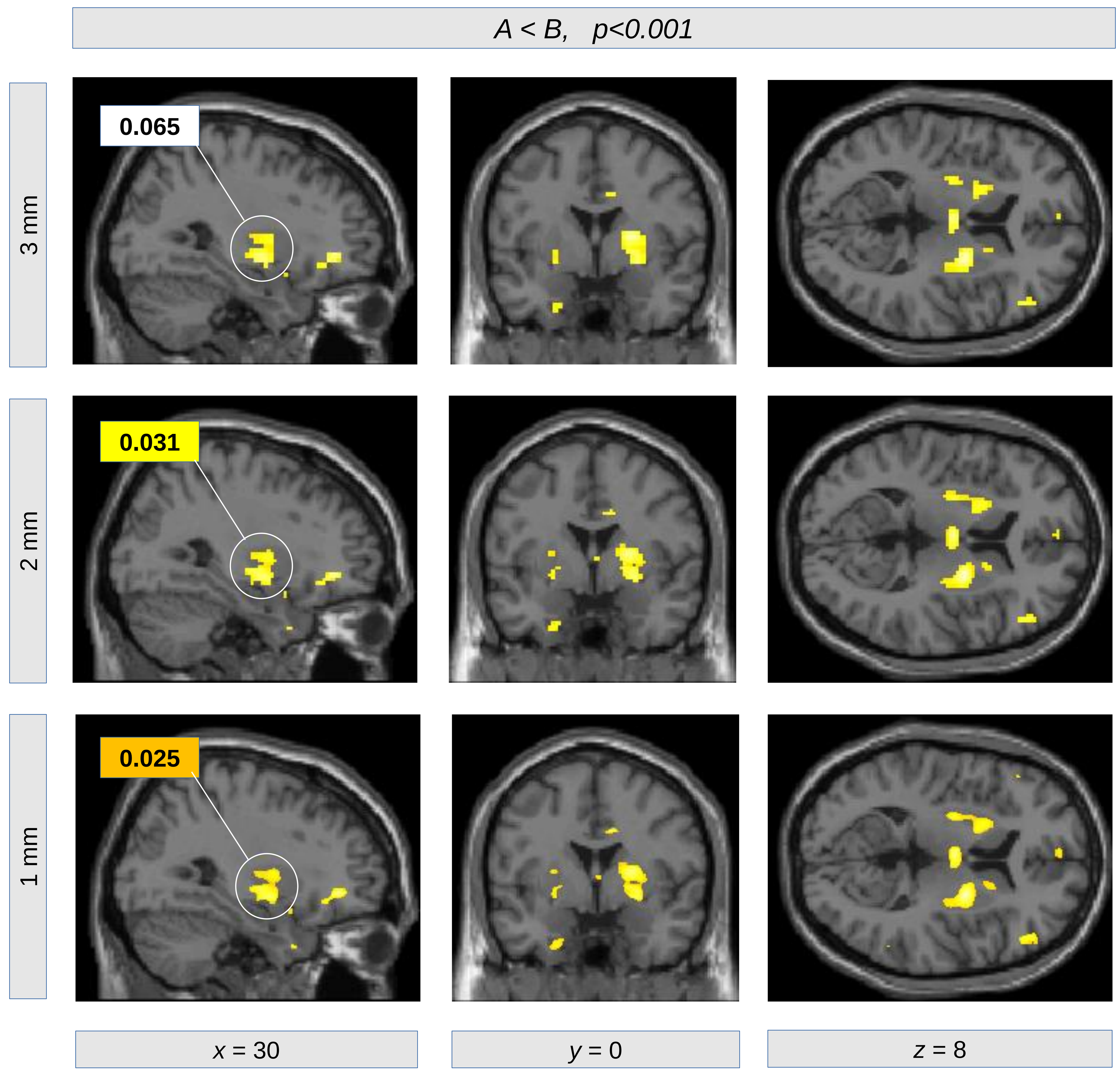

Figure 1. Orthogonal brain sections showing brain activity differences between two experimental conditions A and B using a fake design with resting-state fMRI data. During normalization into the standard space, data were scaled to 3x3x3 mm³ , 2x2x2 mm³ , and 1x1x1 mm³ . Family-wise error (FWE) corrected *p*-values become smaller when using a higher upscaling. The white box shows the FWE-corrected *p*-value of a selected cluster that reaches significance that is a false positive result.

- On the individual level, we found that with higher resampling resolutions, the FWE-corrected *p*-values decrease systematically so that more and more false positives occur.
- As expected by our null hypothesis, we did not obtain any positive clusters with 3×3×3 mm³ resolution. However, with 2×2×2 mm³, the *p*-values are already smaller leading to a significant cluster. With a resolution of 1×1×1 mm³, the *p*-value decreased again.
- Figure 1 shows the effect using a fake event-related design with two experimental conditions.
- We obtained the same effect when using an arbitrary on/off-design with a block length of 20 s.
- It appears that there is a ***systematic*** dependence of the false positive rate on the resampling parameter with smaller voxel sizes leading to smaller FWE-corrected *p*-values and hence more false positives.
- While some dependence on pre-processing parameters may be inevitable, a systematic dependence of this type is clearly worrisome, because researchers may be tempted to interpolate their data until the desired statistical significance level is reached.
- Statistical inference should certainly not depend in such a systematic way on a pre-processing parameter that can be set *ad libitum*.
- Clearly, this issue requires further in-depth analysis.

References

[1] Eklund A, Nichols TE, Knutsson H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates, Proc Natl Acad Sci U S A. 2016;113(28):7900-7905. doi:10.1073/pnas.1602413113

[2] Flandin G, Friston KJ. Analysis of family-wise error rates in statistical parametric mapping using random field theory, arXiv 2016;1606.08199v1 [stat.AP], 27 Jun 2016, http://arxiv.org/pdf/1606.08199.pdf

[3] Hopfinger JB, Büchel C, Holmes AP, Friston KJ. A study of analysis parameters that influence the sensitivity of event-related fMRI analyses, Neuroimage 2000;11(4):326-333. doi:10.1006/nimg.2000.0549