Which Findings from the Functional Neuroimaging Literature Can We Trust?

Daniel Kessler[a,1,2], **Michael Angstadt**[a,1], **and Chandra Sripada**[a,1]

[a]Department of Psychiatry, University of Michigan, Ann Arbor

In their recent "Cluster Failure" paper, Eklund and colleagues cast doubt on the accuracy of a widely used statistical test in functional neuroimaging. Here, we leverage nonparametric methods that control the false discovery rate to offer more nuanced, quantitative guidance about which findings in the existing literature can be trusted. We show that, in the task studies examined by Eklund et al., most clusters originally reported to be significant are indeed trustworthy by the false discovery rate benchmark.

In a substantial contribution to the functional magnetic resonance imaging (fMRI) field, Eklund et al. [1] use nonparametric methods to demonstrate that random field theory (RFT)-based family-wise error (FWE) correction techniques for cluster-level inference do not control errors as they are supposed to, and this discrepancy is particularly pronounced for lenient cluster defining thresholds (CDT). Moreover, they point to violations of RFT assumptions as the culprit for this discrepancy.

Given these results, what advice can we offer to the reader exploring the existing fMRI literature when faced with a table of cluster-wise RFT-based FWE corrected p-values ($p_{\text{RFT-FWE}}$)? To suggest caution is reasonable but incomplete; we require concrete, quantitative guidelines to enable appropriate calibration of skepticism.

Here, we undertake an initial attempt to construct such guidance. We heed Eklund et al.'s warning and prefer null distributions obtained through nonparametric methods rather than through RFT. However, we focus on the False Discovery Rate (FDR; [2]), which is a more natural target for multiple testing control (a point well recognized by Nichols in previous work; [3]). A researcher determining which clusters are significant is more concerned with the proportion that are false positives (FDR) than whether *any* are false positives (FWE). Given these considerations, a reader faced with a table of clusters significant under RFT-FWE correction might naturally ask: Which of these results would have survived had the study instead employed a nonparametric FDR-based method?

We address this question using the same task fMRI data [4, 5] analyzed by Eklund et al., which is available from openfMRI [6]) (code, data, and Extended Methods for present analysis are at http://github.com/mangstad/FDR_permutations).

In brief, for each contrast, we generate 5,000 realizations of the data through sign-flipping. To obtain a null distribution of cluster extents (for an arbitrarily chosen cluster), we combine normalized frequency counts of cluster extents calculated at each realization (realizations with no clusters assign all their mass to 0). This distribution is used to assign uncorrected p-values to each observed cluster. We next submit the vector of uncorrected p-values for each contrast to Benjamini and Hochberg's [2] FDR procedure with $\alpha_{\text{FDR}} = .05$ (cf. [7] for a parametric implementation of cluster-wise FDR).

We compare $p_{\text{RFT-FWE}}$-values to q_{FDR}-values and note whether they survive FDR-correction under $\alpha_{\text{FDR}} = .05$. We generate separate plots for this analysis conducted at CDT={.001, .01}.

Results (see Figure 1) show that for CDT=.001, only one cluster significant at $\alpha_{\text{RFT-FWE}} \leq .05$ failed to survive at $\alpha_{\text{FDR}} \leq .05$, thus suggesting RFT-based FWE closely approximates effective FDR control. This finding has promising implications for the existing body of fMRI studies using RFT-based cluster-level inference that used this stricter CDT, estimated to be upwards of 8,500 reports [8, 9]. For CDT=.01, used in approximately 3,500 studies [8, 9], Eklund et al. and others [10] have urged caution and our results agree. We found that $\alpha_{\text{RFT-FWE}}$ must be very strict (at least .00001) for effective FDR control to be achieved (i.e., many $.05 \geq p_{\text{RFT-FWE}} \geq .00001$ fail to survive correction at $\alpha_{\text{FDR}} = .05$).

These results offer initial quantitative guidance on interpreting the past literature that employed RFT-based FWE, providing a more granular appreciation of the relationship between $p_{\text{RFT-FWE}}$ and trustworthiness of the result. A more comprehensive examination of fMRI task data sets that used RFT-based FWE may further refine this guidance.

1. Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences* p. 201602413.
2. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1):289–300.
3. Genovese CR, Lazar NA, Nichols T (2002) Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage* 15(4):870–878.
4. Duncan KJ, Pattamadilok C, Knierim I, Devlin JT (2009) Consistency and variability in functional localisers. *NeuroImage* 46(4):1018–1026.
5. Tom SM, Fox CR, Trepel C, Poldrack RA (2007) The Neural Basis of Loss Aversion in Decision-Making Under Risk. *Science* 315(5811):515–518.
6. Poldrack RA et al. (2013) Toward open sharing of task-based fMRI data: the OpenfMRI project. *Frontiers in Neuroinformatics* 7:12.
7. Chumbley JR, Friston KJ (2009) False discovery rate revisited: FDR and topological inference using Gaussian random fields. *NeuroImage* 44(1):62–70.
8. Nichols TE (2016) Bibliometrics of Cluster Inference, 06/07/16, Neuroimaging Statistics Tips & Tools. http://blogs.warwick.ac.uk/nichols/entry/bibliometrics_of_cluster/.
9. Woo CW, Krishnan A, Wager TD (2014) Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *NeuroImage* 91:412–419.
10. Flandin G, Friston KJ (2016) Analysis of family-wise error rates in statistical parametric mapping using random field theory. *arXiv:1606.08199 [stat]*. arXiv: 1606.08199.

www.pnas.org/cgi/doi/10.1073/pnas.XXXXXXXXXX

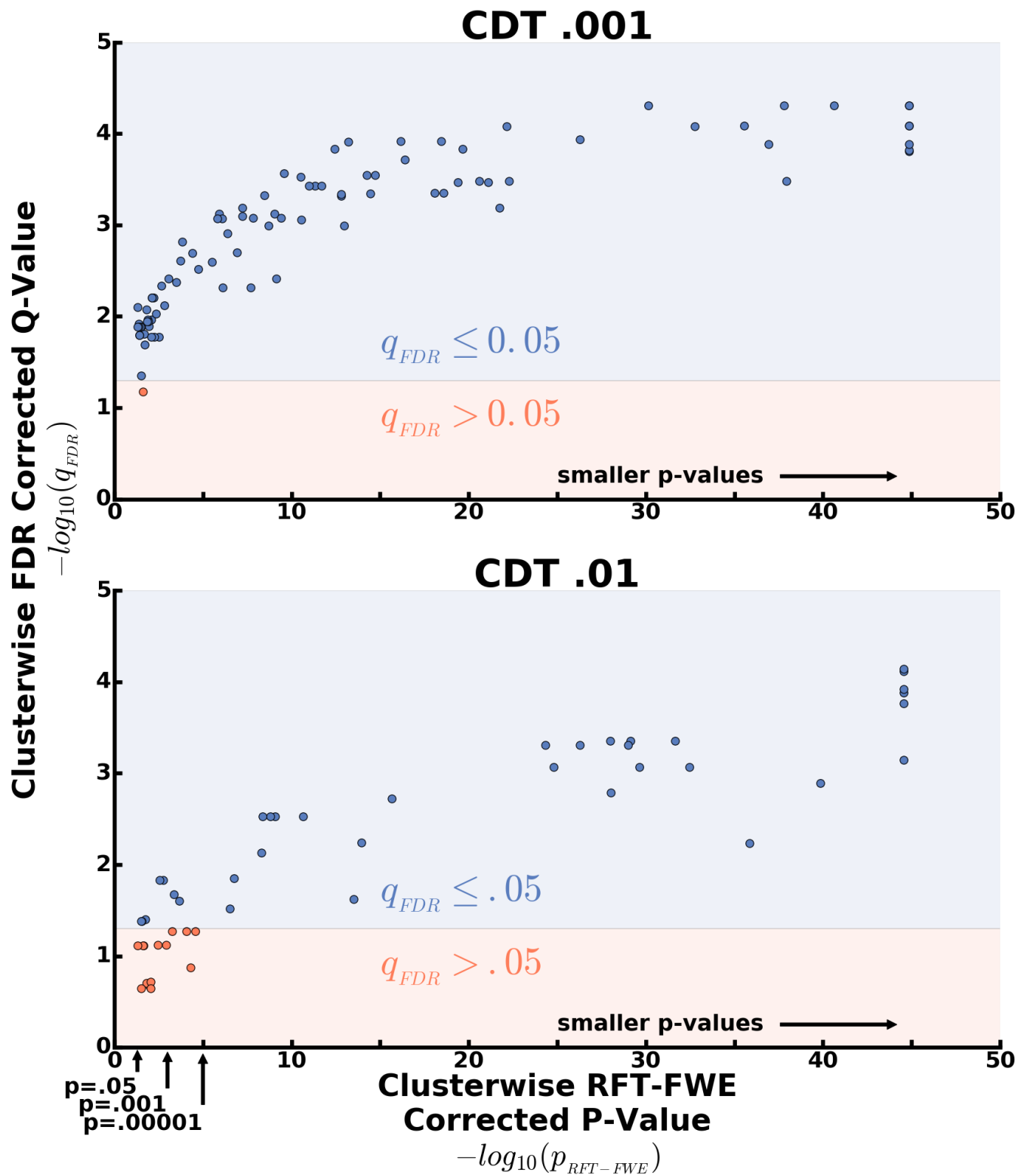PNAS | August 29, 2016 | vol. XXX | no. XX | 1–2

Fig. 1. Assessing RFT-Based FWE Using an FDR Benchmark. We submitted the same task data analyzed by Eklund et al. to nonparametric clusterwise FDR analysis. For CDT = .001 (top), RFT-based FWE approximates effective FDR control with $\alpha_{\text{FDR}} = .05$. For CDT = .01 (bottom), only clusters with $p_{\text{RFT-FWE}} \leq .00001$ reliably survived correction at $\alpha_{\text{FDR}} = .05$.