Highlight Results, Don't Hide Them: Improving Reproducibility, with Applications to NARPS

POSTER #731

Paul A. Taylor[1], Richard C. Reynolds[1], Vince D. Calhoun[2], Javier Gonzalez-Castillo[3], Daniel A. Handwerker[3], Peter A. Bandettini[3,4], Amanda F. Mejia[5], Gang Chen[1]

[1]SSCC (NIMH, NIH, USA), [2]TReNDS, Georgia State, Georgia Tech, and Emory U (USA), [3]SFIM (NIMH, NIH, USA), [4]FMRIF (NIMH, NIH, USA), [5]Dept of Statistics, Indiana U (USA)

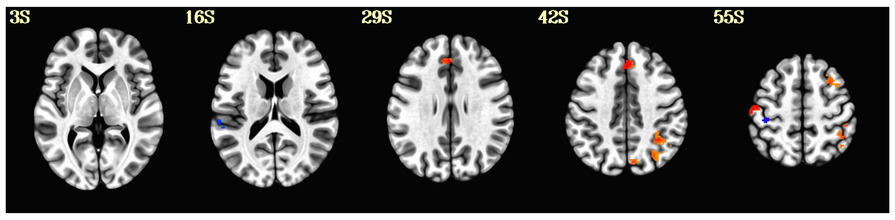
Contact: paul.taylor@nih.gov https://afni.nimh.nih.gov/pub/dist/doc/htmldoc/

Introduction

- Standard thresholded activity maps amplify differences, distort interpretations of results and harm reproducibility evaluation.[1]
- We show how the "highlighting" approach[2] (*focus* on the strongest results, but also show the rest) greatly improves reporting for interpretation and meta analysis.[3]
- We present examples using the NARPS[4] data, where >60 teams analyzed the same data with their own pipelines and compared results primarily using thresholded maps.
- Highlighting demystifies the kind of variability seen among the NARPS teams' results: *it is mainly a varied strength of agreement rather than disagreement.*

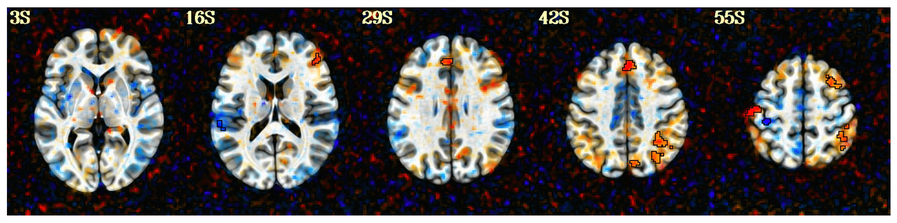
Example FMRI results reporting comparison

A) Standard "hiding": opaque threshold



- Hides information (good science presents evidence)
- Wastes data (modeling occurred everywhere!)
- Mathematically promotes selection bias
- ***Neither*** the brain ***nor*** BOLD are ON/OFF like this
- Results are inherently sensitive to threshold parameter: tiny differences are magnified
- Harms both understanding and reproducibility: similarities (or differences!) are hidden

B) Improved "highlighting": translucent threshold



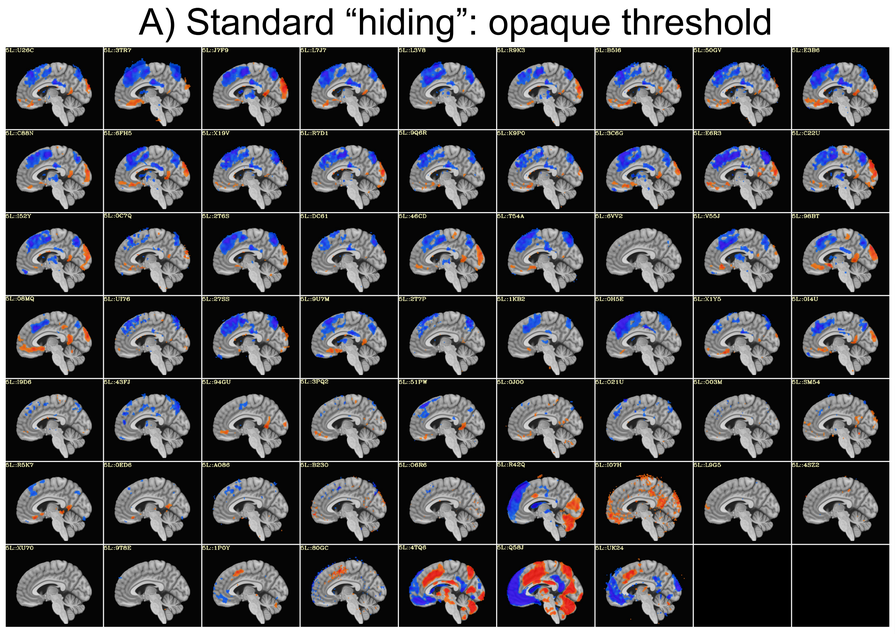
- Presents more scientifically: show ***evidence*** instead of threshold-sensitive decisions (let reader see more)
- Emphasize focal regions—but place in context
- Improve quality control (QC): reduce lurking artifacts (even check outside brain!) and reveal poor modeling
- Reduce artificial and arbitrary threshold dependency
- Allow for more meaningful comparisons across studies: better for evaluating reproducibility

Methods and Results

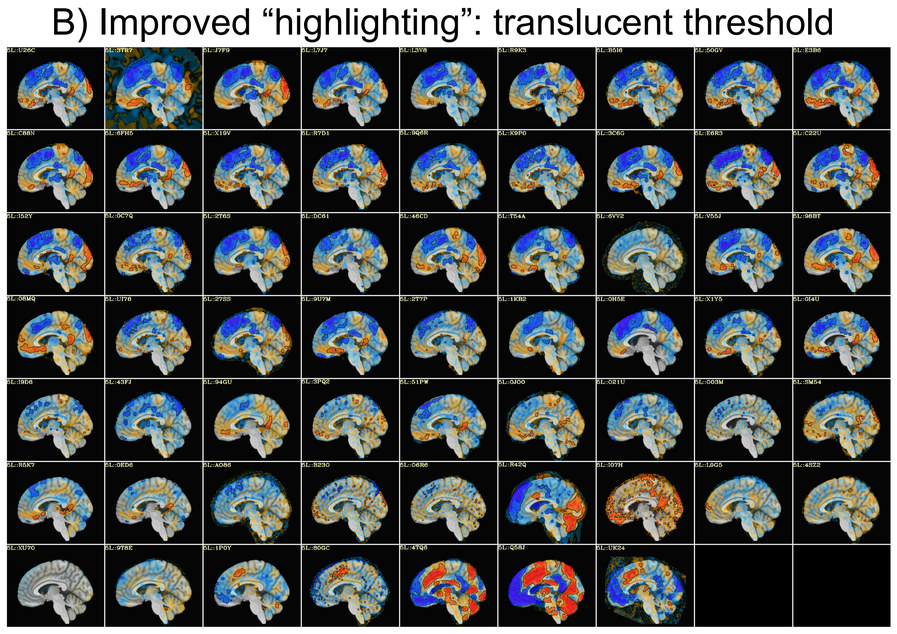
- We performed cross-study comparisons of the NARPS teams' results; outcomes were similar for each NARPS hypotheses.
- Compare seeing each NARPS teams' stats result for Hyp #1&3 in: hiding mode (A) vs in highlighting mode (B).
  - "A" appears to show little agreement, which might be interpretable as wide variability or disagreement
  - "B" shows more information and that *teams' results actually have high agreement, simply with varied strength.*
- This can be quantified with similarity matrices: degree of agreement in A is unsurprisingly lower, while B shows mostly excellent agreement, with a few results that are different (and a few with inverted signs).
- The results are analogous across all Hyps[3]: transparent thresholding shows that most NARPS teams produce similar activation patterns, just with varied strength.

NARPS teams' results comparison (Hyp #1 and 3, +gain, indif grp)

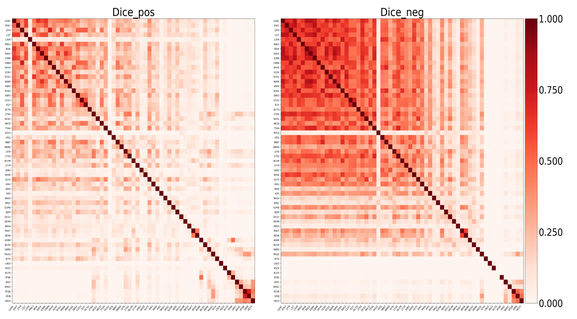
A) Standard "hiding": opaque threshold



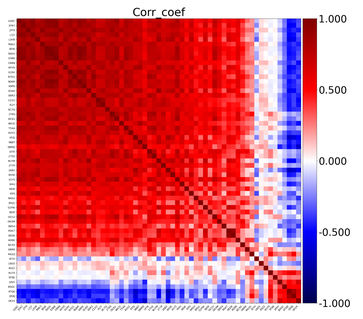
B) Improved "highlighting": translucent threshold



Dice coefficients between "hiding" thresholded positive and negative regions; low-similarity bias:



Correlation coefficients between "highlighting" unthresholded statistical maps; most data with very high correlation, and more details about other cases:



Conclusions

While data sharing is useful, the thresholding recommendations made here are distinct and important on their own. Results should be presented informatively from the start, without the need for extra work, downloading, re-analyzing, etc. A study's interpretation and a reader's initial impression should be as accurate and complete as possible. This work has major implications for the neuroimaging community in presenting results and performing meta-analyses. Individual studies and meta-analyses should adopt a "highlighting" approach, in order to accurately assess results.

Tutorial Pages and Download Scripts

- Download NARPS processing scripts, using AFNI[5]: https://github.com/afni/apaper_highlight_narps
- AFNI command examples for scripting images: https://afni.nimh.nih.gov/pub/dist/doc/htmldoc/tutorials/auto_image/auto_%40chauffeur_afni.html
- Transparent thresholding in AFNI GUI (AFNI Academy Bootcamp channel): https://www.youtube.com/watch?v=VT77zJ0zGnA&list=PL_CD549H9kgqwHr0EDtvAU8hylsOj30OK&index=3
- "Highlight" viewing in afni_proc.py quality control (APQC) HTML: https://afni.nimh.nih.gov/pub/dist/doc/htmldoc/tutorials/apqc_html/apqc_ex1.html#qc-block-vstat

Acknowledgment & References

PAT, RCR and GC were supported by the NIMH and NINDS Intramural Research Programs (ZICMH002888) of the NIH/HHS, USA. VDC was supported by NSF grant #2112455. JG-C, DH and PAB were supported by the NIMH Intramural Research Program (ZIAMH002783) of the NIH/HHS, USA. AM was supported by a NIBIB grant (R01EB027119) of the NIH/HHS, USA. This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov). We also thank the NARPS organizers and all of the team participants.

[1] Chen G, Taylor PA, Stoddard J, Cox RW, Bandettini PA, Pessoa L (2022). Aperture Neuro. 2.
[2] Allen EA, Erhardt EB, Calhoun VD (2012). Neuron 74:603-608.
[3] Taylor PA, Reynolds RC, Calhoun V, Gonzalez-Castillo J, Handwerker DA, Bandettini PA, Mejia AF, Chen G (2023). Neuroimage 274:120138.
[4] Botvinik-Nezer R, et al. (2020). Nature 582(7810):84-88.
[5] Cox RW. Computers and Biomedical Research, 29:162-173, 1996.