fMRI Quality Assurance

Daniel Handwerker
Catherine Walsh

Section on Functional Imaging Methods
National Institute of Mental Health

July 28, 2026

National Institute of Mental Health

Quality Assurance = QA
Quality Control = QC

Main sources for this talk

The art and science of using quality control to understand and improve fMRI data

Teves, Gonzalez-Castillo, Holness, Spurney, Bandettini, Handwerker

Front. Neurosci. (2023) https://doi.org/10.3389/fnins.2023.1100544

Preliminary results presented at OHBM 2026, poster #1442

Hippocampus involvement in a dual-retrocue working memory task

Walsh, Smith, Gonzalez-Castillo, Handwerker, Morgan, Bandettini

https://fim.nimh.nih.gov/assets/presentations/walsh_ohbm_2026.pdf

Take home messages

Part 1

- Reproducible & relevant science requires knowing your data
  - Pipelines & sample size don't matter if your data is not good
  - "What is good data" varies with context & application
- Quality assurance **protocols** are central to knowing your data
- Automation is great, but insufficient
- Training on how to do QA needs more attention
- QA is an active area of research & should be more active

Why focus on quality control?

Significant transitions in the field

- **Good**: Large N studies and shared data!

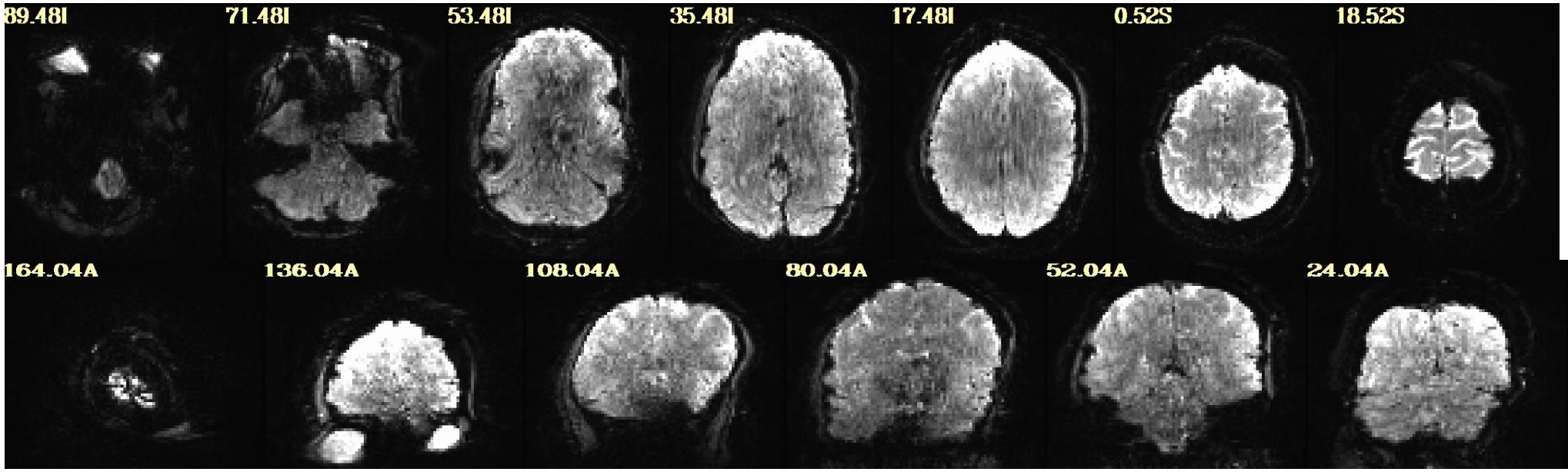
- **Neutral**: Different people design, collect, preprocess, and analyze data

- **Less good**:
- People make assumptions about data quality from previous phases
- Teaching assessment of data quality is often hands-on
  - With more people using data from others, fewer people get this hands-on training

A (mildly provocative) case study

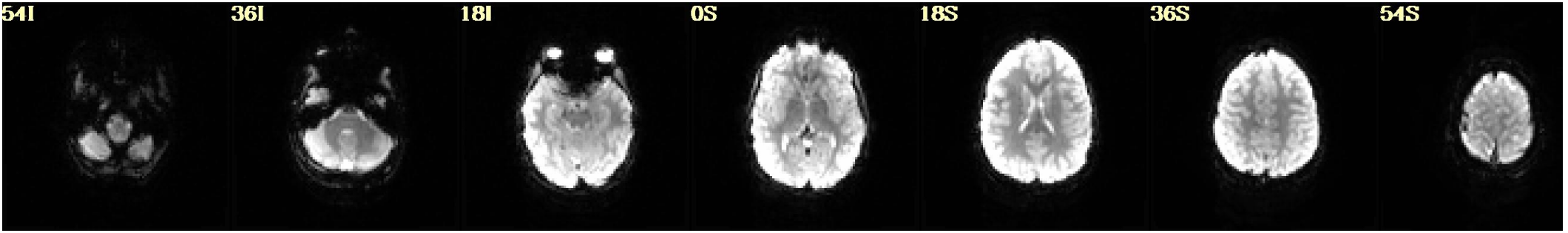
There is no such thing as ”gold standard data”

Arbitrary volunteer from the original Human Connectome Project



TR=0.72sec
Multi-slice accel=8
2mm^3 voxels

Arbitrary volunteer from ABIDE

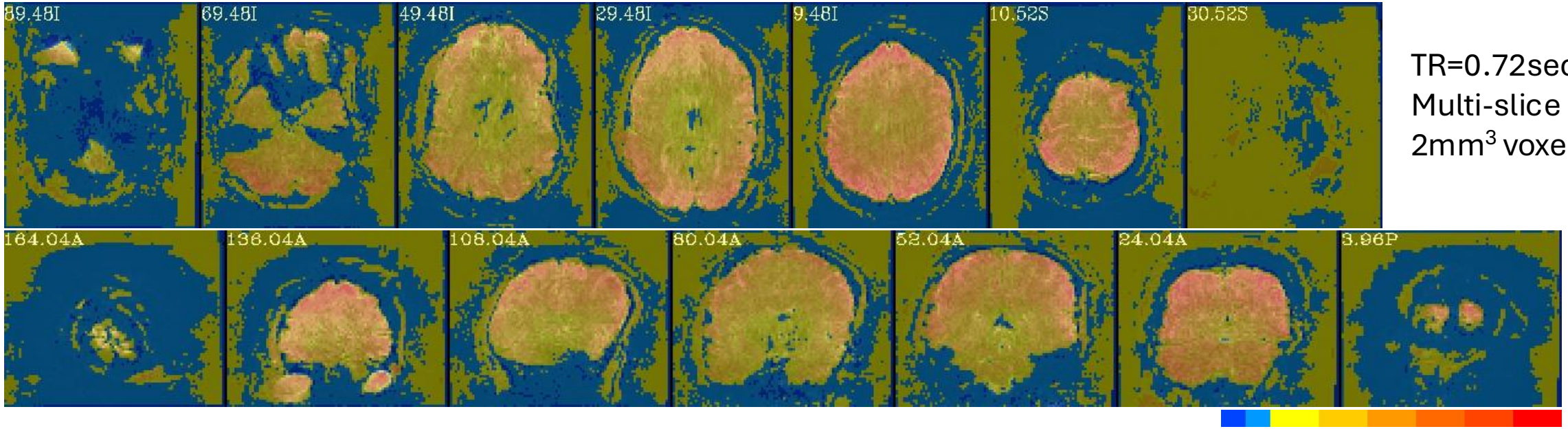


TR=2.5sec
In-slice accel=3
~3mm^3 voxels

A (mildly provocative) case study

There is no such thing as ”gold standard data”

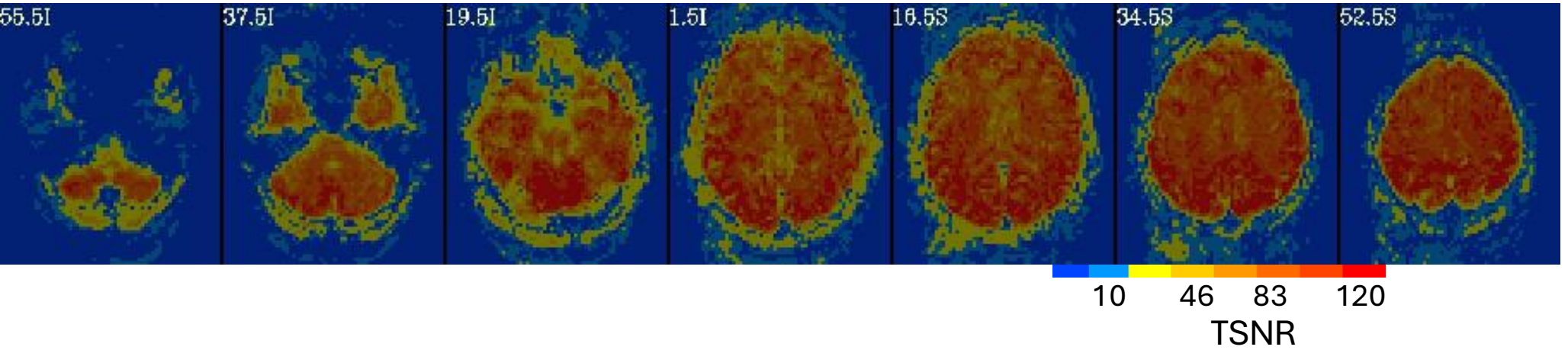
Arbitrary volunteer from the original Human Connectome Project



TR=0.72sec
Multi-slice accel=8
2mm^3 voxels

Temporal Signal to noise ratio (TSNR)

Arbitrary volunteer from ABIDE



TR=2.5sec
In-slice accel=3
~3mm^3 voxels

Context and applications matter

- Successful Research using HCP data
  - ROIs that average across multiple small voxels
  - Correlation or task studies that summarize data across time
  - Averages across the large population
- HCP weaknesses
  - Studies that fully take advantage of the short TR **and** smallish voxel size
  - Brain-wide association studies (BWAS) that require robust signal in individuals' data
- **Note:** This is a broad & not completely fair generalization.
- **Take home message**
  - Great data for one application, might not be great for all applications
  - Identify and view data quality metrics relevant to your application

Approaching QA

- Will these data have the potential to accurately and effectively answer my scientific question?
  - …and future questions others might ask?
- Identify & log data anomalies or unexpected variations that might skew or hide key results
  - Reduce problems through processing or removal.
- **All datasets have problems**
  - Not checking → Incorrect or misleading interpretations of results
  - Checking → **Fewer** unknown problems
- Wang and Strong 1996:

  QC has both intrinsic and contextual measures

Questions for a QA Protocol

Priority	Context & Examples
General	
Which voxels have usable data?	Voxel-wise data quality & coverage*
Are locations of voxels accurately defined?	Distortion & alignment to anatomy & templates*
Define context	Scientific questions & study priorities affect what is or is not good quality data
During study planning	
QC measures to support study goals	Particularly for study-specific QC priorities, this is a good time to seek expert advice
Operation procedures to decrease acquisition errors	Good procedures are critical for making sure data are accessible and consistently documented
Additional measures to collect	Experimenter notes, behavior logs, respiratory & cardiac traces
Organization & sharing QC measures	Inaccessible information is not useful
Piloting acquisition & processing	Evaluate and improve a QC protocol as part of study piloting
During Acquisition	
Real-time monitoring of severe image distortions, head motion, task non-compliance	Observing problems during acquisition can give time to recollect data or fix problems for the current or future scans
Monitor peripheral measures	Respiration, cardiac, eye tracking
Soon after acquisition or download	
Expected data are all present and properly documented	Missing, duplicated, or corrupted files, incomplete runs.* For MRI data, behavioral logs, and peripheral measurements
Data consistency & documented parameters match data	Consistent MRI field of view, contrast, orientation, number of runs, & run lengths match documentation*+
Documentation on QC during acquisition or pre-sharing exists	No documentation means there are undocumented problems
Data plausibly useful for study goals	Regions of interest should have full coverage. No substantive temporal artifacts that affect connectivity measures
Atypical brain structures, acquisition artifacts, drop out, and distortion	May still be fine`, but might require altered processing. AFNI's instacorr can be useful for assessment

During and after processing	
Scripts ran properly	Expected logs, QC metrics, & outputs created*
Appropriate voxels retained or removed	Voxels with good SNR in brain are within mask and voxels outside of brain are removed.*
Voxels lost to dropout or field of view	Check that similar voxels are retained across the population+
Consistent measures of temporal signal-to-noise and intrinsic spatial smoothness across population	Sessions with non-trivially lower TSNR or different smoothness can be a warning sign of other problems*
Automatically removed data	Number of censored volumed and DOF lost from noise regression, temporal filtering, & censoring*
Artifacts like ghosting, phase wrapping, or leakage	Instacorr is useful for checking if the temporal signal from an article is folding over into other brain regions
Partially-thresholded activation maps	Are areas with the largest model fits in anatomically plausible patterns inside the brain?*
Task correlated head motion or breathing	Not commonly checked and can bias results.* (AFNI automatically checks motion, but not breathing.)
Skull properly masked for anatomical & functional data	Can cause problems with alignment. Part of report from AFNI's SSwarper
Intensity inhomogeneity	Brighter signal on the surface can be expected, but can cause problems with masking and alignment*
Good anatomical to functional alignment & alignment across days/runs	Can be a serious hidden problem if one just looks at group maps.+
Left & right hemispheres flipped between anatomical & functional data	More common than it should be & requires excluding data unless the true left/right can be determined*
Good anatomical to anatomical alignment across participants	Often correctable and causes problems if not corrected+
Group coverage across population	A summation of aligned functional masks highlights brain areas missing in part of the population+
Processed peripheral data are good	Plausible behavioral timing files, good peak detection in respiratory & cardiac traces

- General
- Study planning
- (Hardware QC)
- During acquisition
- Soon after acquisition
- During processing

Phases interact

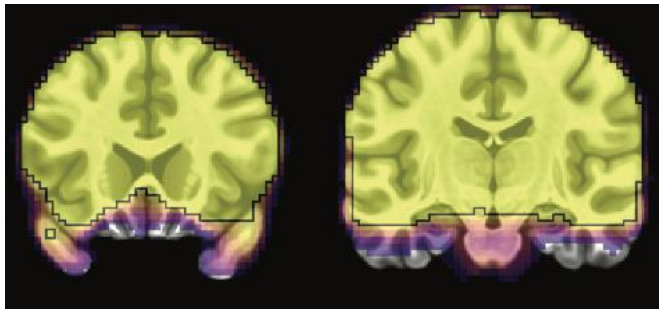
Questions
Not a
checklist

Appendix of: Teves et al "The art and science of using quality control to understand and improve fMRI data" Front. Neurosci. (2023) https://doi.org/10.3389/fnins.2023.1100544

Study planning

- What QC measures matter for my study?
  - Brain coverage? Distortion? TSNR?
  - Stable connectivity maps?
  - Avoiding problematic artifacts?
- Operating procedures:
  - Documentation and real-time checks
  - Data & meta-data to store. What **and** how
- **Piloting**
  - Testing QC measures
  - Testing operating procedures

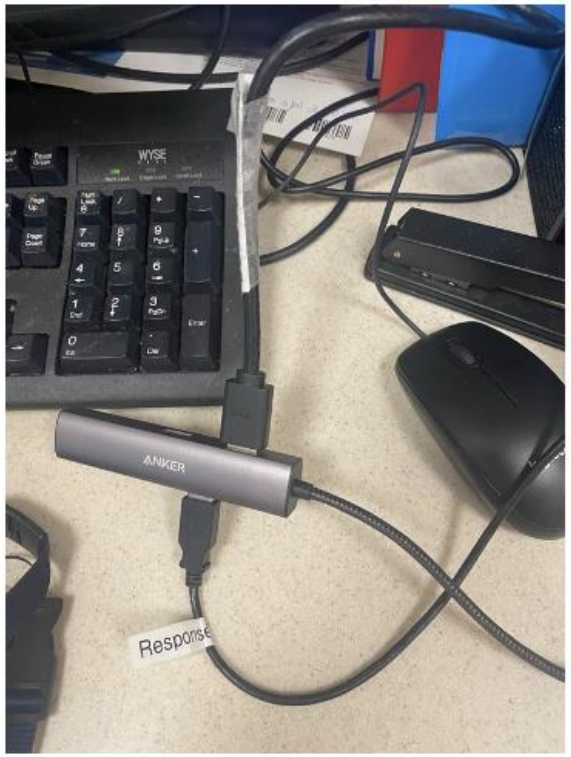
Coverage



See 7/7/2026 talk on parts of an fMRI study
https://youtu.be/RjA3cSz4ZR8?si=y_qpRq38znO-I-qN&t=2321

Study planning
Operation procedures to decrease acquisition errors

2. Set up experiment laptop:
    a. Turn second screen to have HDMI input
    b. 2 cords that need to be set up: Response Box Output and BOLD Screen HDMI

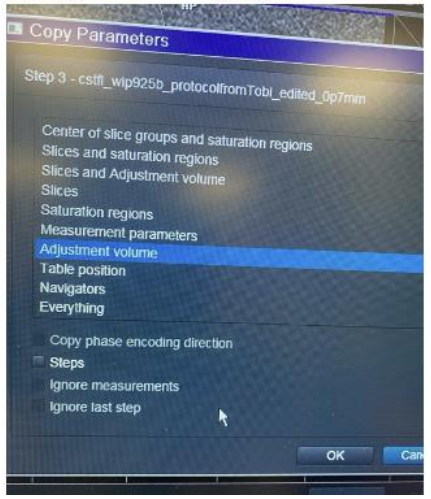


Running the scan

1. Patient Registration:
    a. Patient → Browser → Scheduler → find patient, and make sure you choose the appropriate order (if multiple)
        i. Will open registration and auto-populate all information, just need height, weight and patient position: Head first, supine

h. Cstfl_wip925b_protocolfromTobi : structural scan
    i. FOV (yellow box): get the full brain, keep nose on inside of box (otherwise it'll wrap around and be funky)
    ii. Copy shim parameters from previous functional scans: open prescription for structural, then right click on the functional scan → Copy Parameters → Adjustment Volume

c. Test that these work by making sure laptop is getting input from them

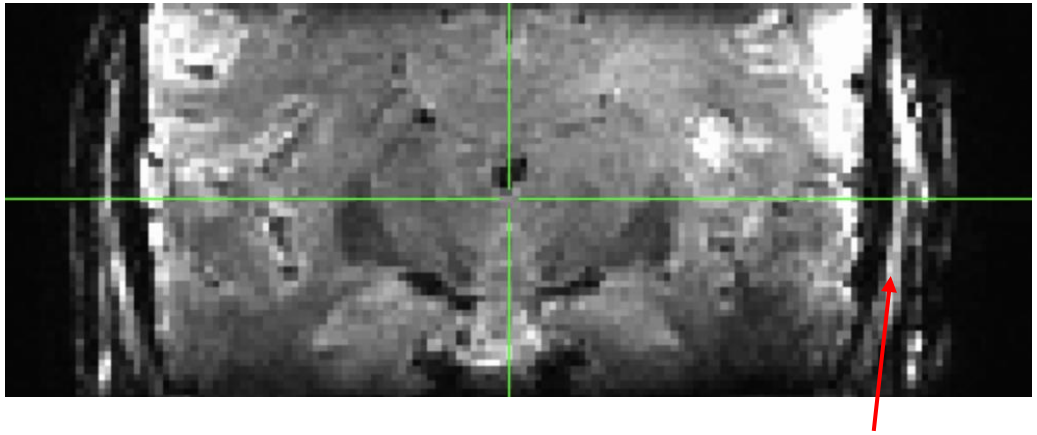


Lists or checklists can be useful

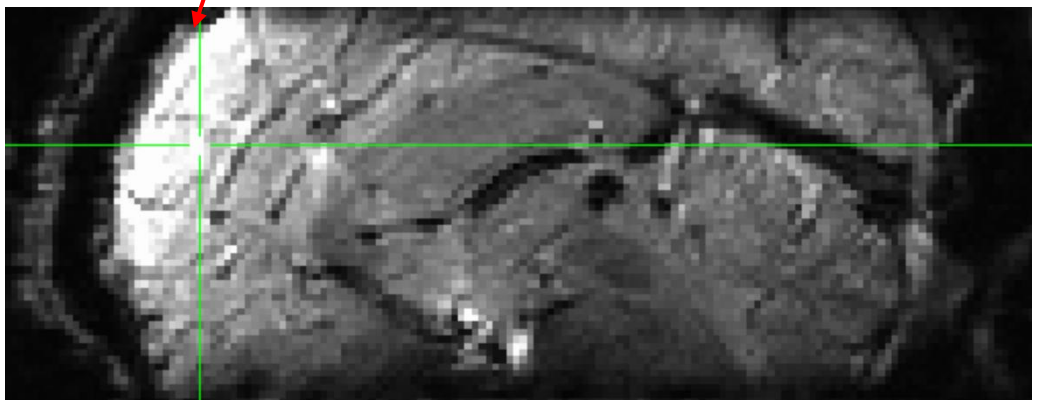
Procedures should include how to log info during acquisition

Catherine Walsh's study

Study planning
Piloting showed fMRI to structural alignment failures



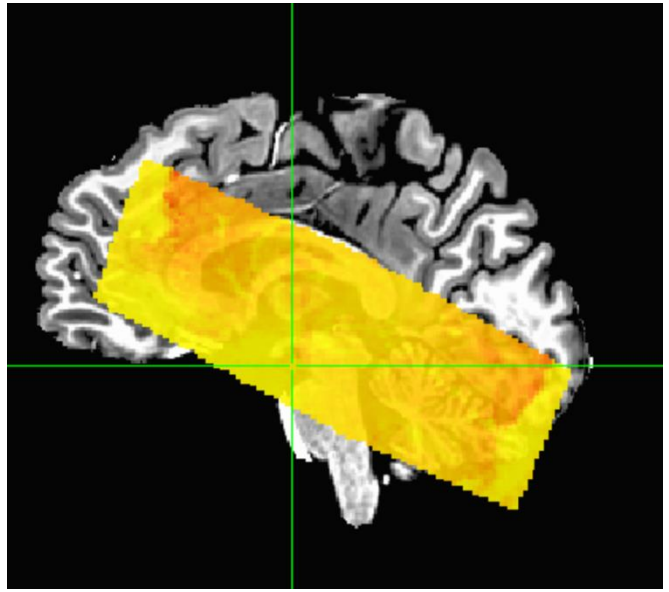
fMRI image inhomogeneity & bright skull

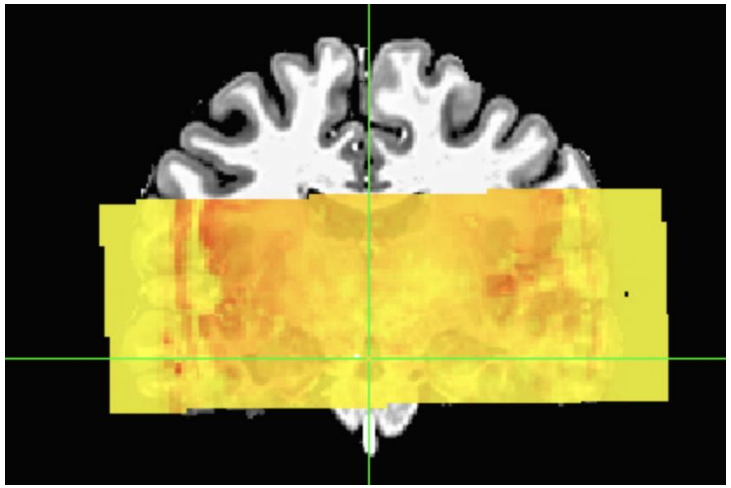


Catherine Walsh's study

Initial alignment options did not work

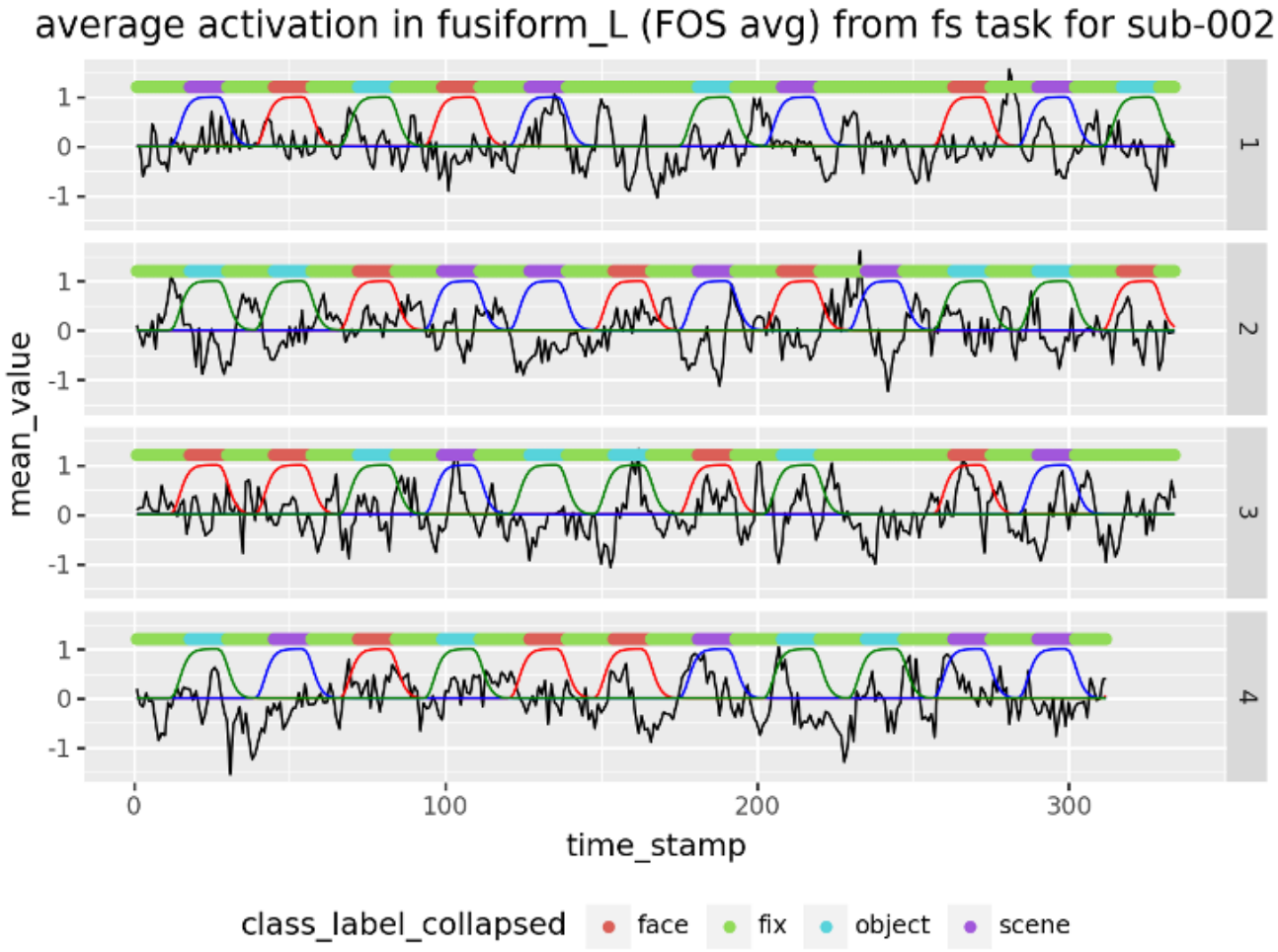
Was solvable with changing options rather than needing to alter acqusitions

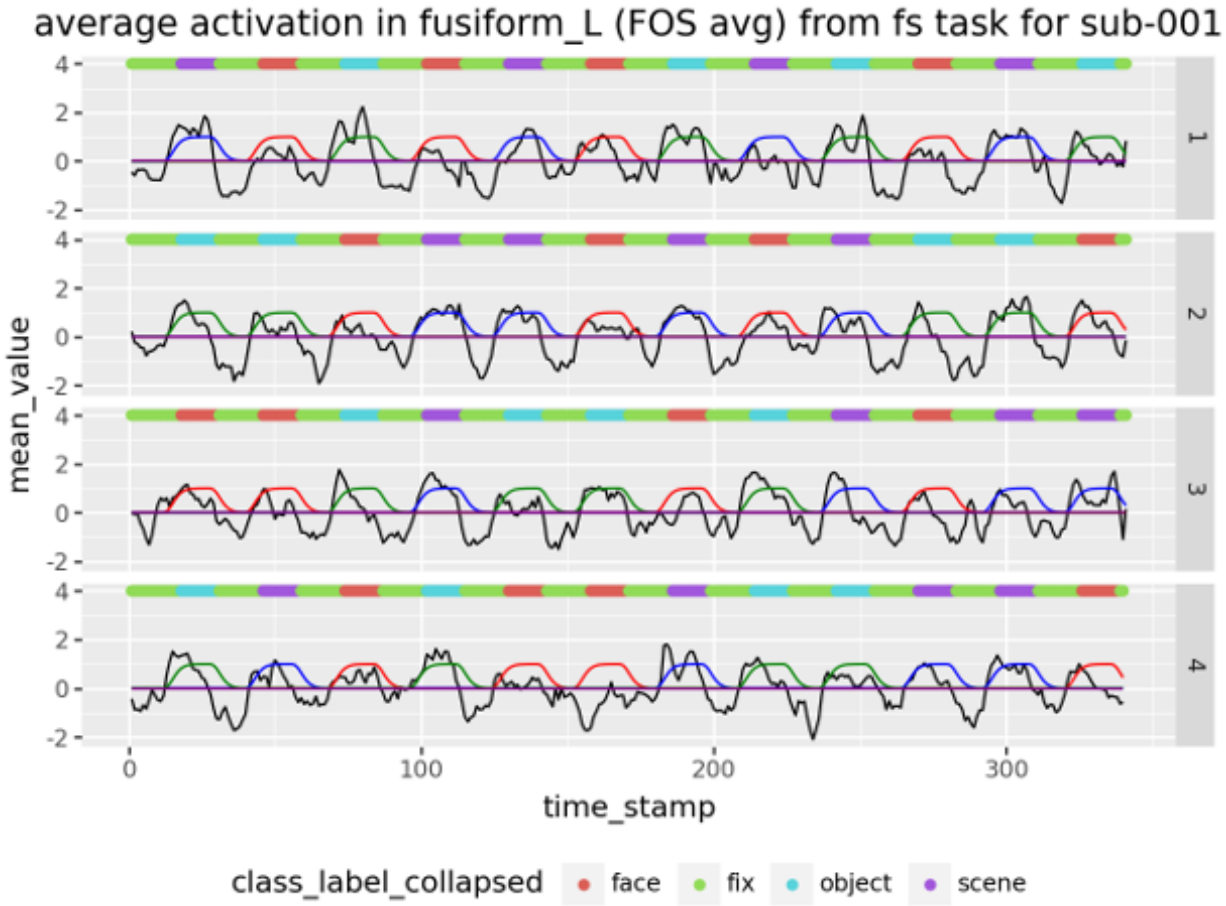




Study planning

Piloting: Realized wrong run was being loaded in analysis script

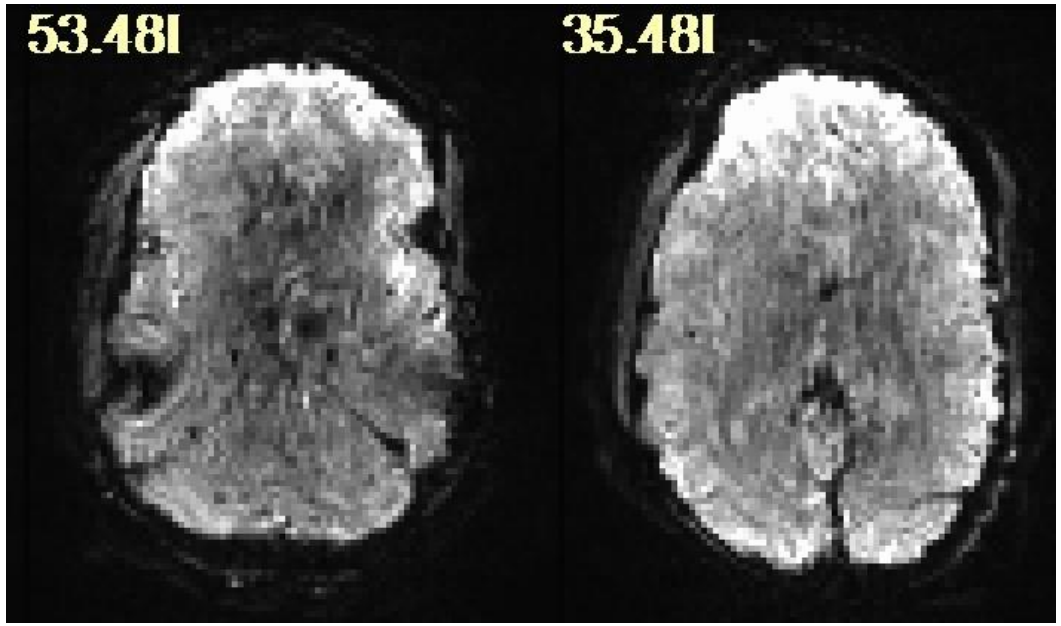


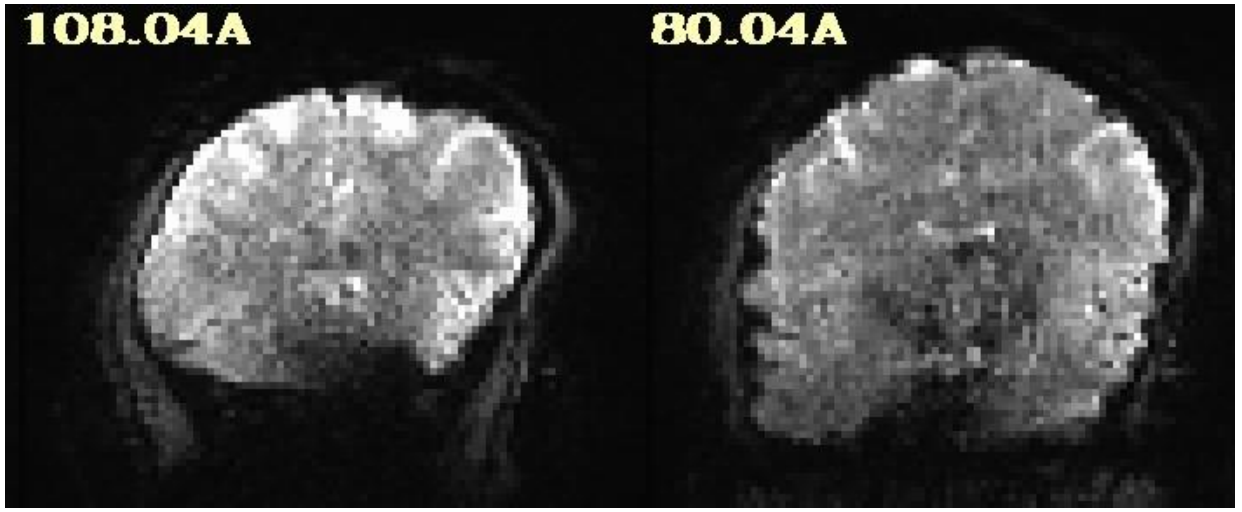


Also gives a QA benchmark for what good runs should look like

Catherine Walsh's study

Study planning & soon after acquisition



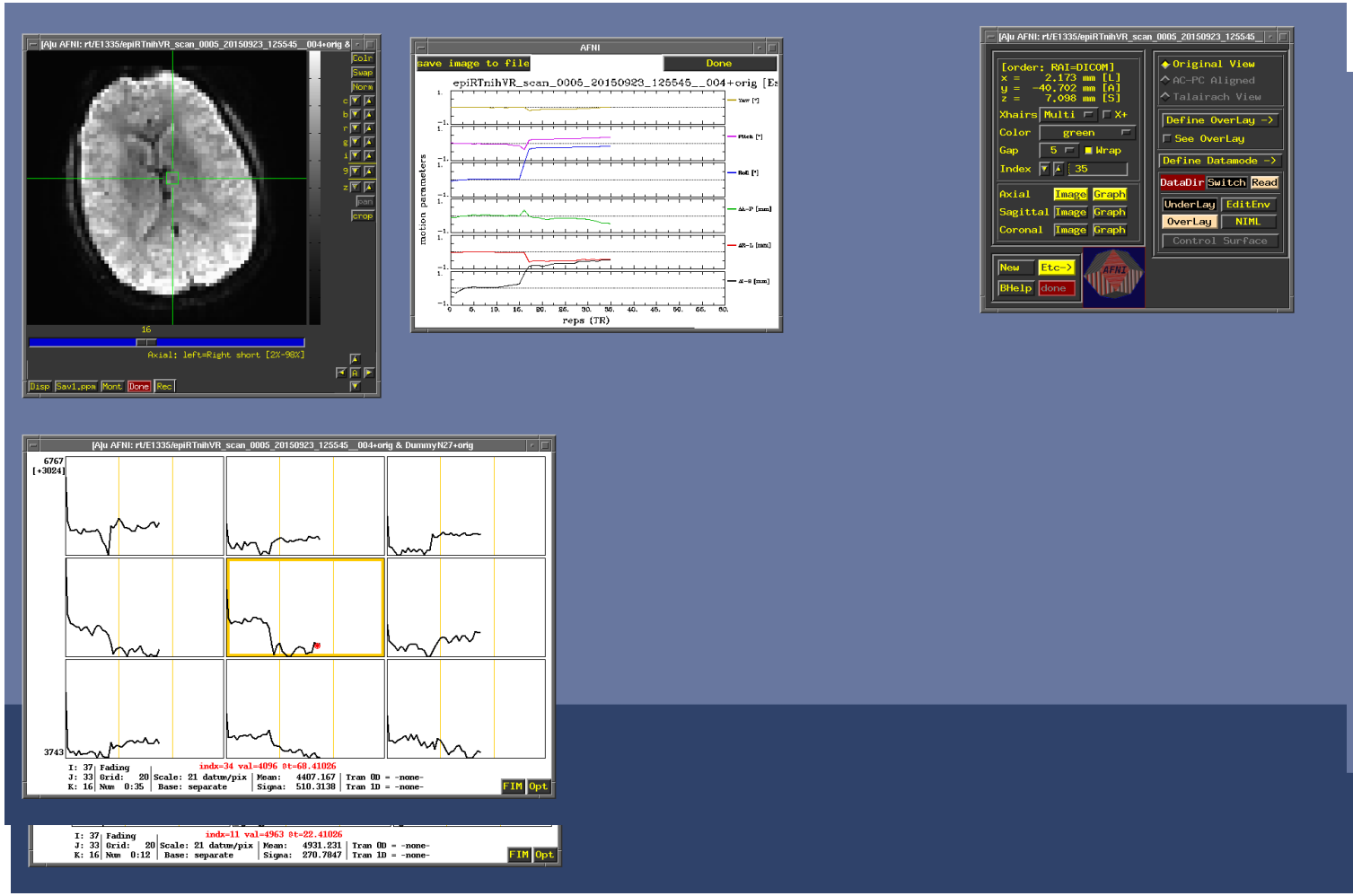


Look for fMRI distortions and dropout that might affect anatomical localization and alignment

HCP data

Data acquisition

- Check outputs as soon as possible!
  - Neuroimaging, peripheral measures, and behavior
- Document anything atypical or different from acquisition plan
  - Altered run order or bad runs, faulty response recording
  - Volunteer sleepiness/non-compliance
  - Obvious artifacts or atypical brain structures



AFNI real time interface Image from Vinai Roopchansingh

Data acquisition

Real time monitoring – task working the way it's supposed to

Run 1 – all is good

Time	Category	Description
0.0275	EXP	wait_for_scanner_text: autoDraw = True
62.9514	DATA	Keypress: t
62.9582	EXP	wait_for_scanner_text: autoDraw = False
62.9582	EXP	pad_block_fix: autoDraw = True
63.8259	DATA	Keypress: t
64.8173	DATA	Keypress: t
65.8187	DATA	Keypress: t
66.8174	DATA	Keypress: t
67.8173	DATA	Keypress: t
68.8181	DATA	Keypress: t
69.8180	DATA	Keypress: t
70.8342	DATA	Keypress: t
71.8193	DATA	Keypress: t
72.8185	DATA	Keypress: t

Run 3 – extra triggers before block padding started

Time	Category	Description
1428.6347	EXP	wait_for_scanner_text: autoDraw = True
1504.8348	DATA	Keypress: t
1505.8418	DATA	Keypress: t
1506.8165	DATA	Keypress: t
1506.8229	EXP	wait_for_scanner_text: autoDraw = False
1506.8229	EXP	pad_block_fix: autoDraw = True
1507.8207	DATA	Keypress: t
1508.8196	DATA	Keypress: t
1509.8192	DATA	Keypress: t
1510.8180	DATA	Keypress: t
1511.8160	DATA	Keypress: t
1512.8188	DATA	Keypress: t
1513.8182	DATA	Keypress: t
1514.8270	DATA	Keypress: t

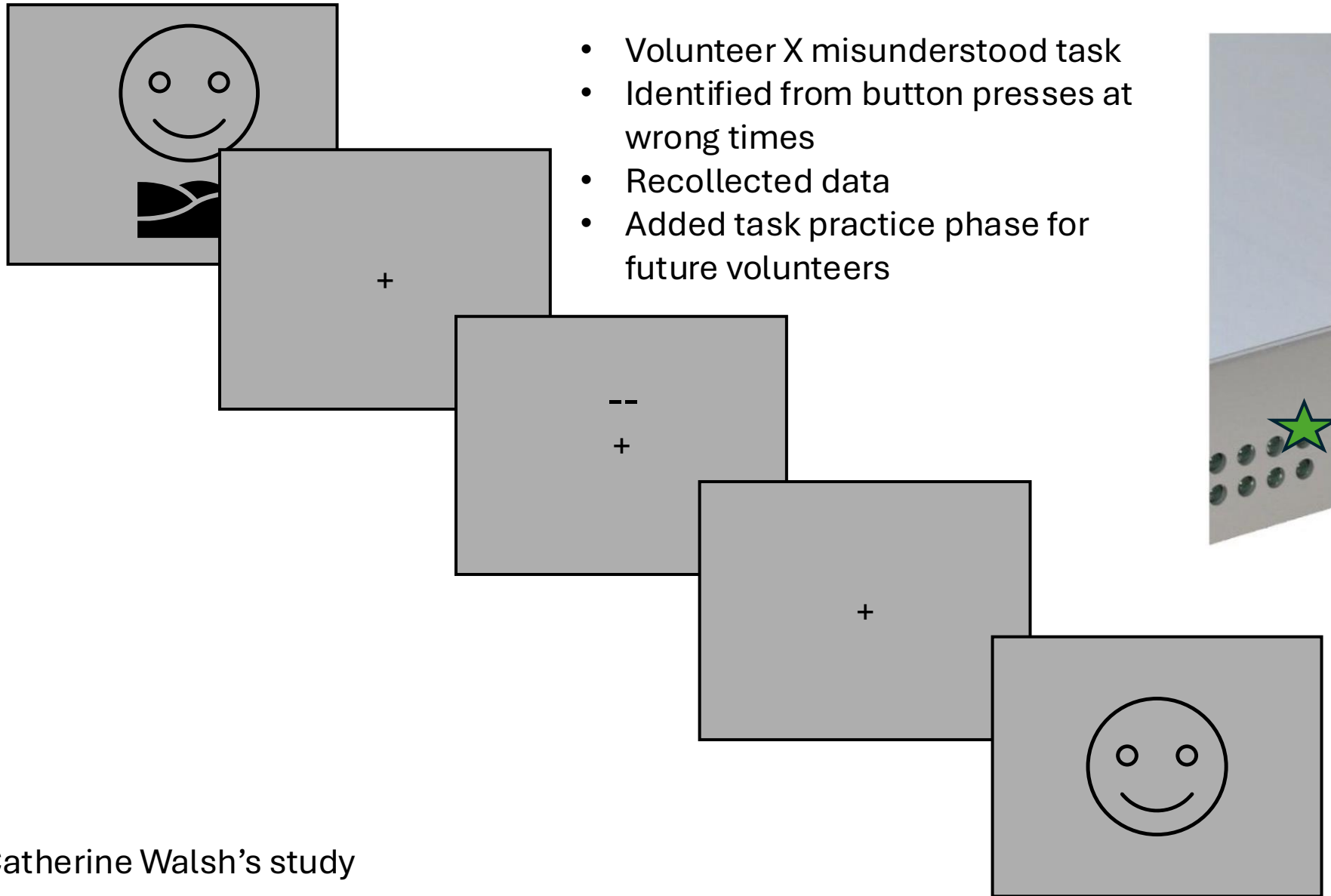
Example log of PsychoPy stimulus presentation software

Note: Need to make sure info you care about is logged

Catherine Walsh's study

Data Acquisition

Real time monitoring – task compliance



- Volunteer X misunderstood task
- Identified from button presses at wrong times
- Recollected data
- Added task practice phase for future volunteers

Catherine Walsh's study

Data acquisition

Real time monitoring – logging & motion

Scan	Name	Notes
1	localizer	
2	loc tfl	
3	DR B - run 1	moving around a lot
4	DR B - run 2	scan crashed because of recon
5	DR B - run 2	cardiac data started getting weird.
6	DR B - run 3	wiggly again
7	loc B - run 1	
8	loc B - run 1	
9	loc B - run 2	
10	loc B - run 3	wiggly
11	loc B - run 4	big movement between runs 3 and 4

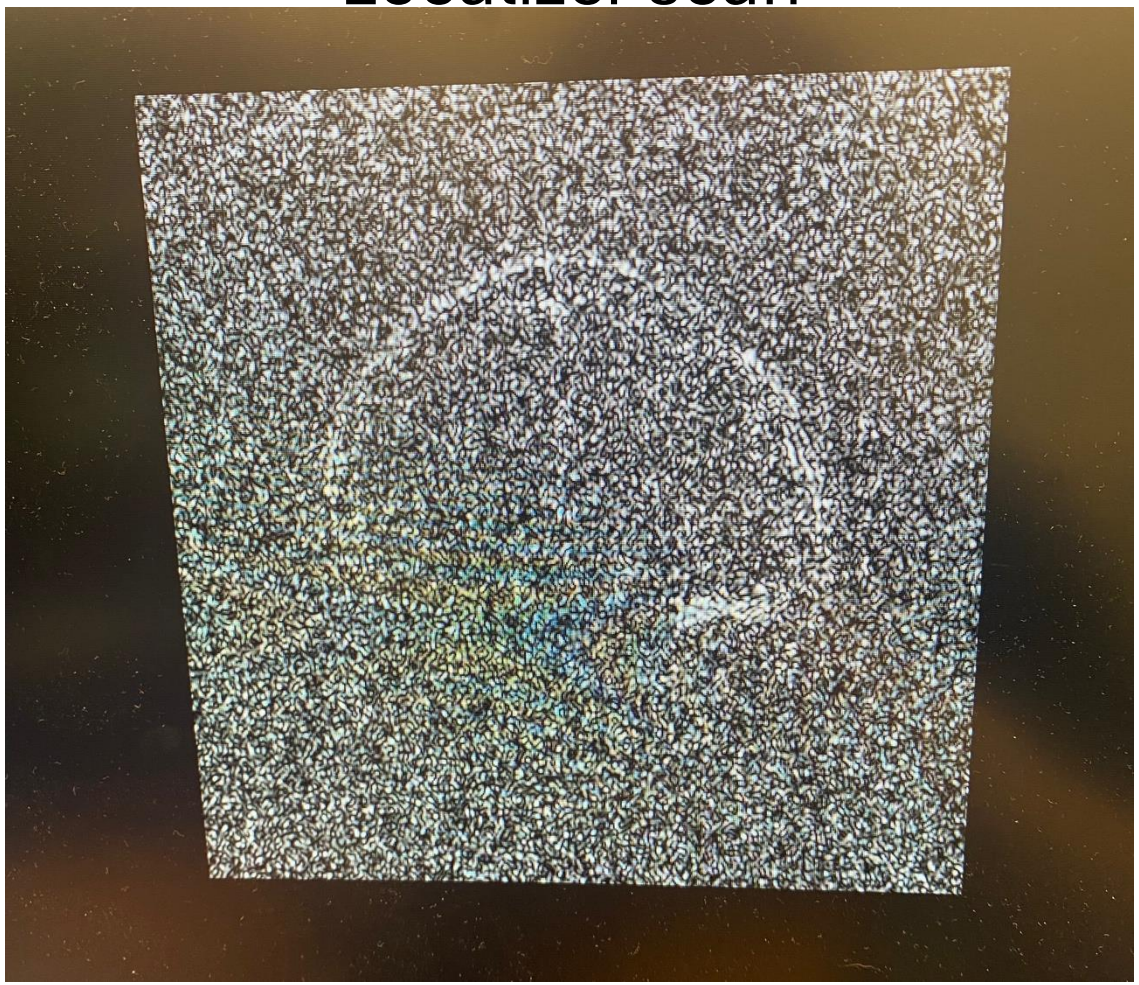
Notes taken by hand and then typed

Catherine Walsh's study

Data Acquisition

Real time monitoring – severe image distortions

Localizer scan

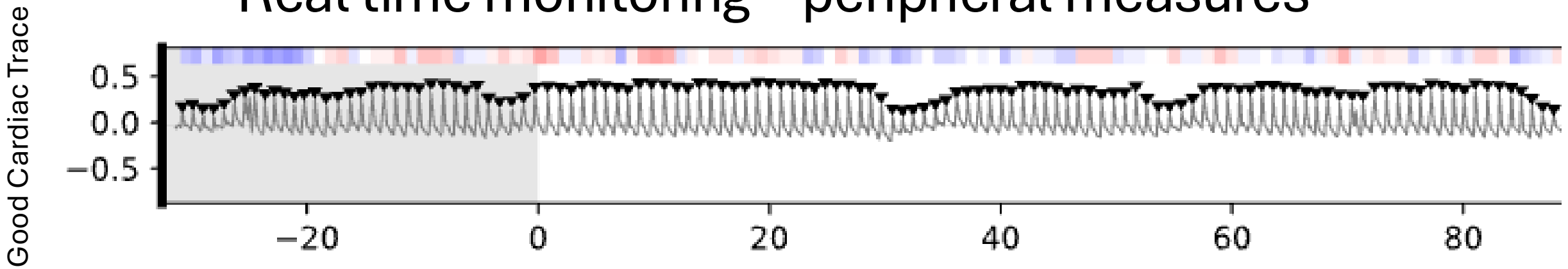


Rebooted scanner & problem resolved

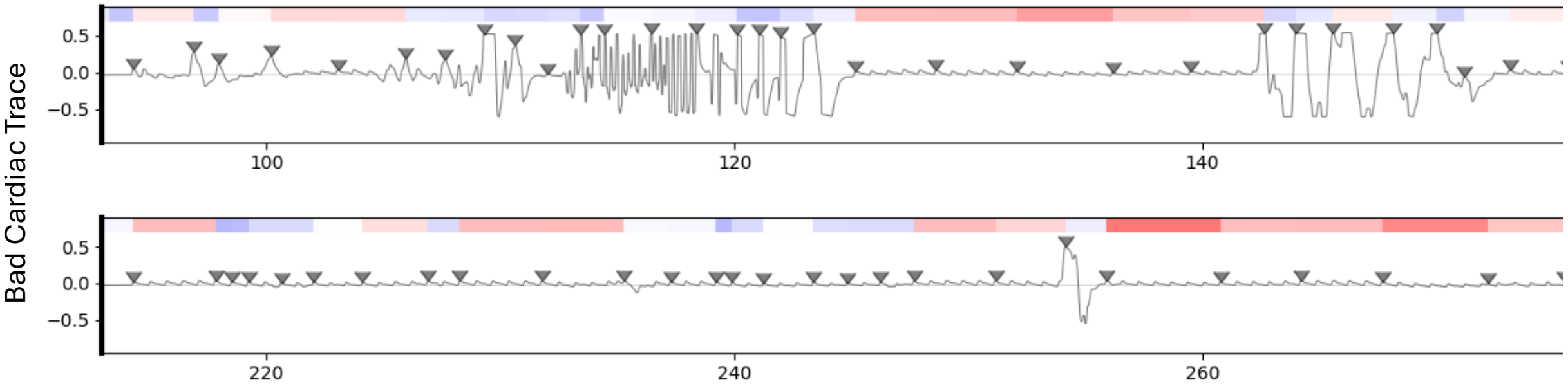
Catherine Walsh's study

Data acquisition

Real time monitoring – peripheral measures



Report and QA interface from AFNI's physio_calc.py

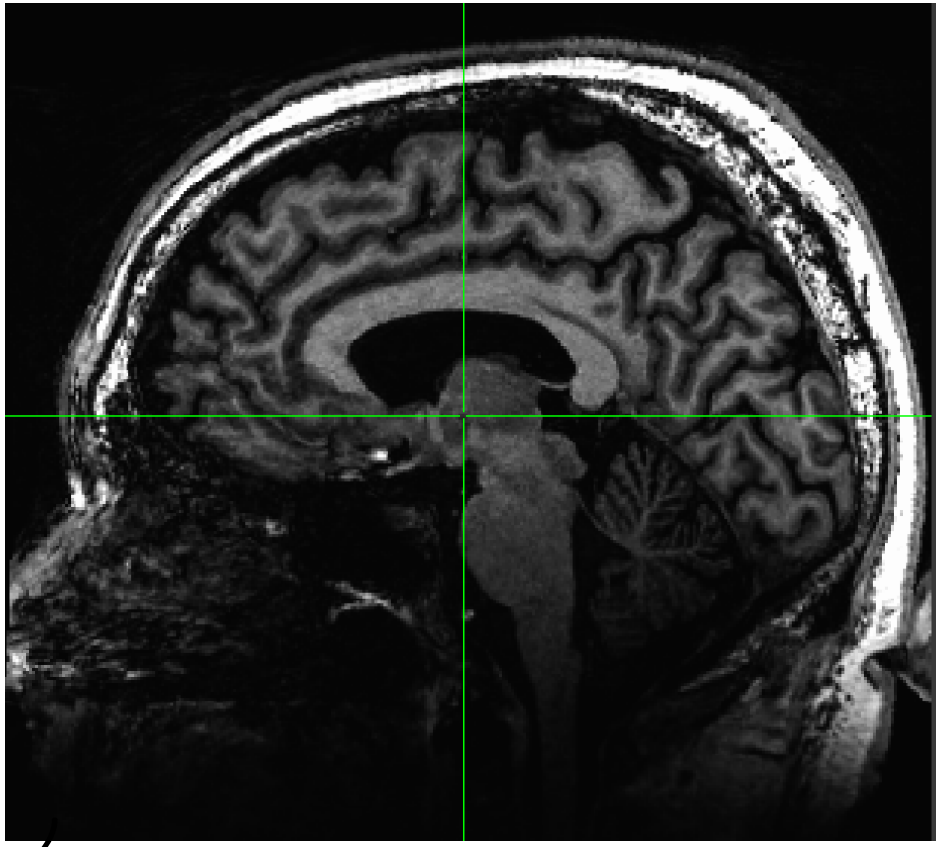


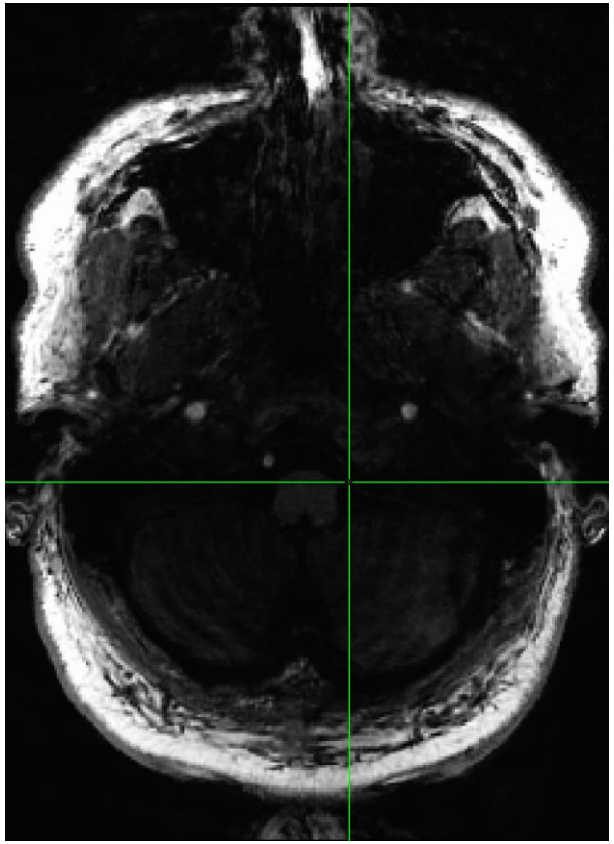
Bad traces can sometimes be fixed if noticed during the scan

Catherine Walsh's study

Soon after acquisition

Visual checks of data





Wrap-around of noise on structural image

This is not a major problem (might affect skull stripping)

but wrap-around or aliasing into the brain can be a major problem

Soon after acquisition

Expected data are all present

Not yet collected sessions

Not yet processed data

Regions of interest with possibly too few useable voxels

```
++ [INFO]: Starting validation at 2026-04-23 16:18:07[++INFO]: Task data not collected yet:
Likely haven't collected task version B for subject sub-011 yet: missing 54 items
Likely haven't collected task version B for subject sub-001 yet: missing 54 items

[++INFO]: Processing not run yet:
Likely have not processed data using retroicor from version A for subject sub-003: missing 18 items

[++INFO]: ROIs that may be too small in general (no data for any processing)
L MTL.PRC.35, task version B, subject sub-003
R MTL.PRC.35, task version B, subject sub-003
R MTL.PHC, task version A, subject sub-010
L MTL.PRC.35, task version A, subject sub-010
R MTL.PRC.35, task version A, subject sub-010
R MTL.PHC, task version B, subject sub-010
R MTL.PRC.35, task version B, subject sub-010
L MTL.PRC.35, task version B, subject sub-010
L MTL.PRC.35, task version A, subject sub-011
R MTL.PRC.35, task version A, subject sub-011
```

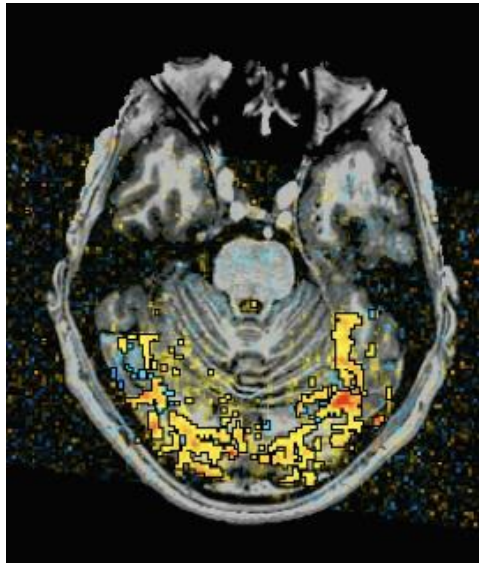
Output of script to run on all data to check if unprocessed & processed data for participants are present

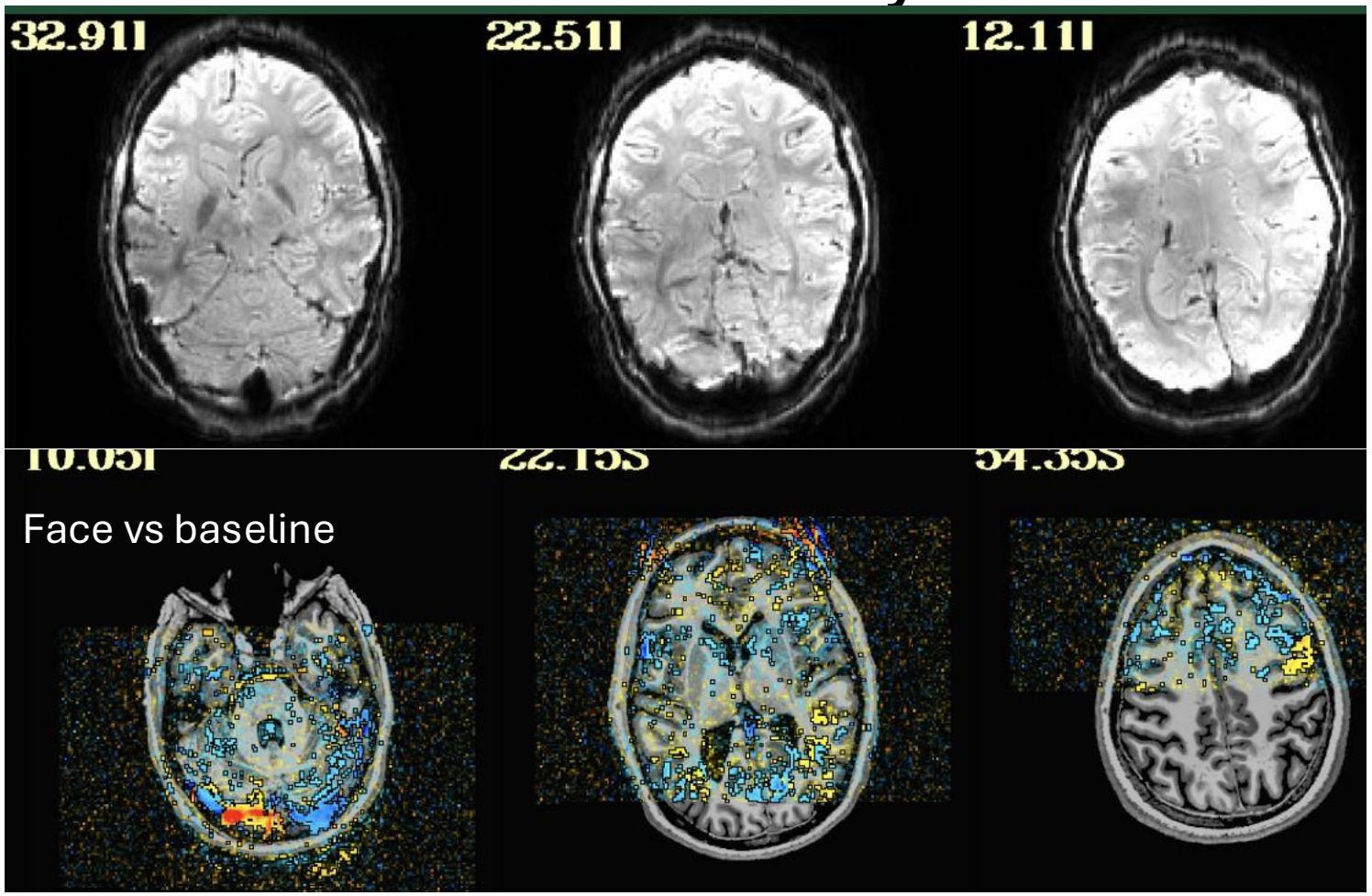
Catherine Walsh's study

Soon after acquisition

Anatomical variability

Face vs baseline





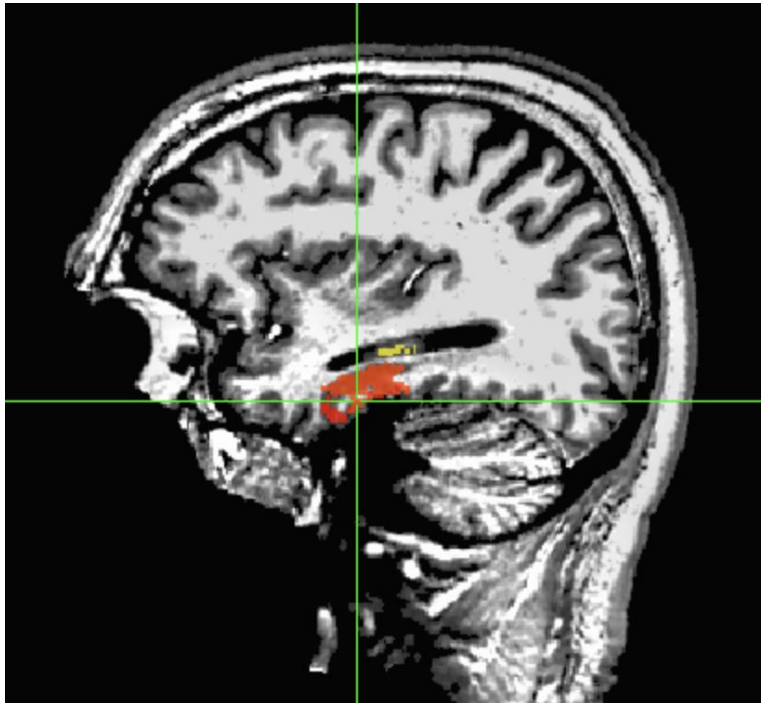
Incidental findings in healthy volunteers are common

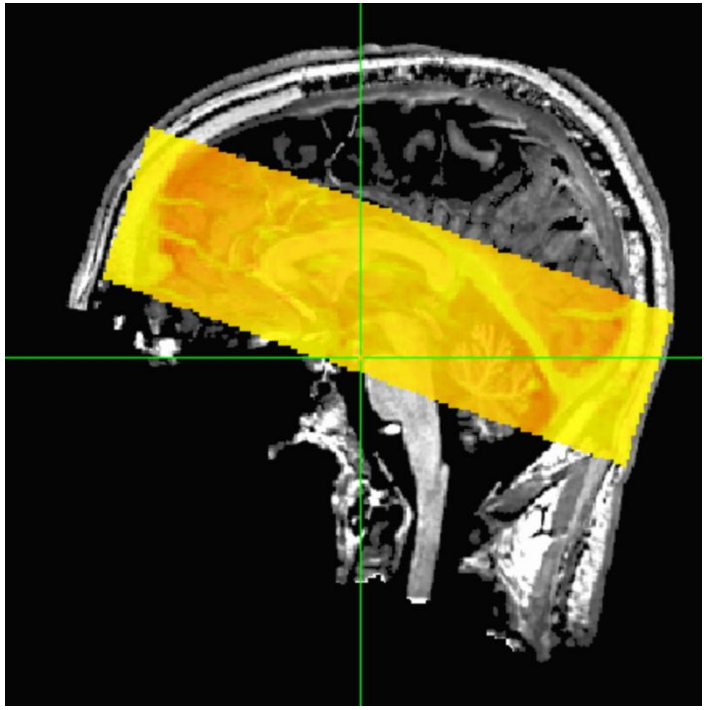
This volunteer has good data for a with-subject ROI analysis,

Analyses requiring alignment to a common template might not work

Soon after acquisition

Data collected in critical areas



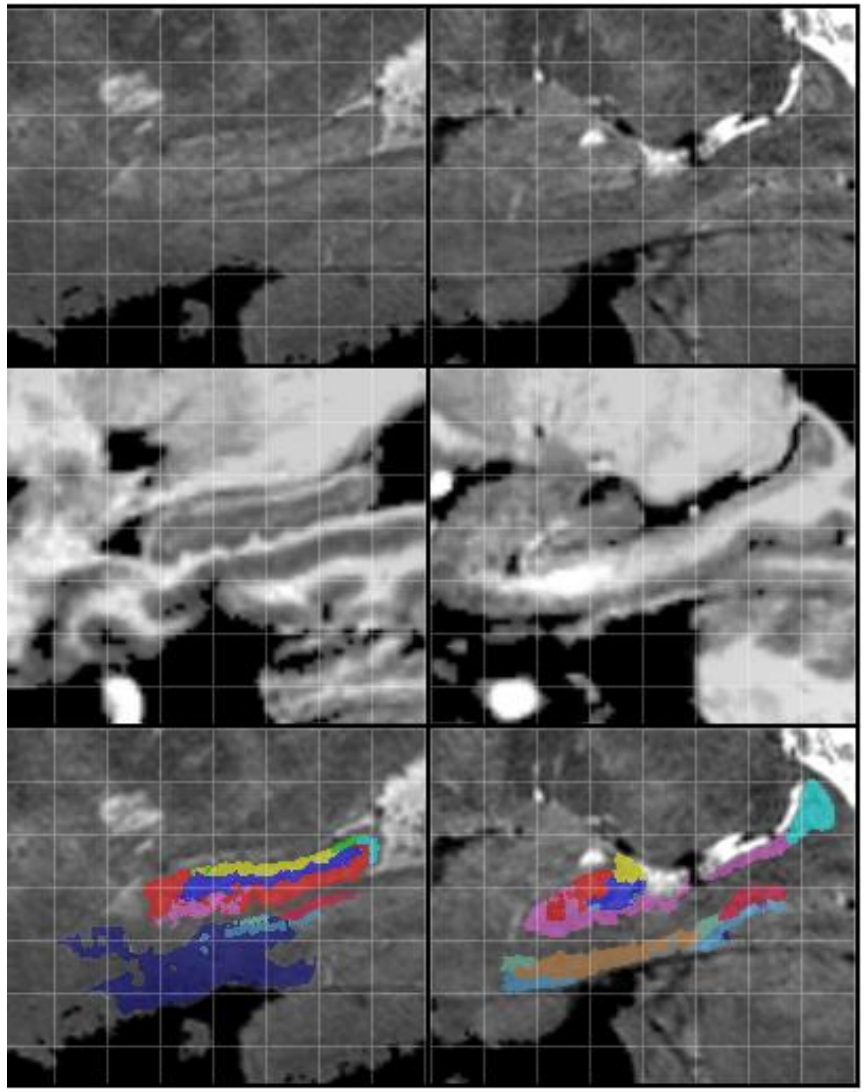


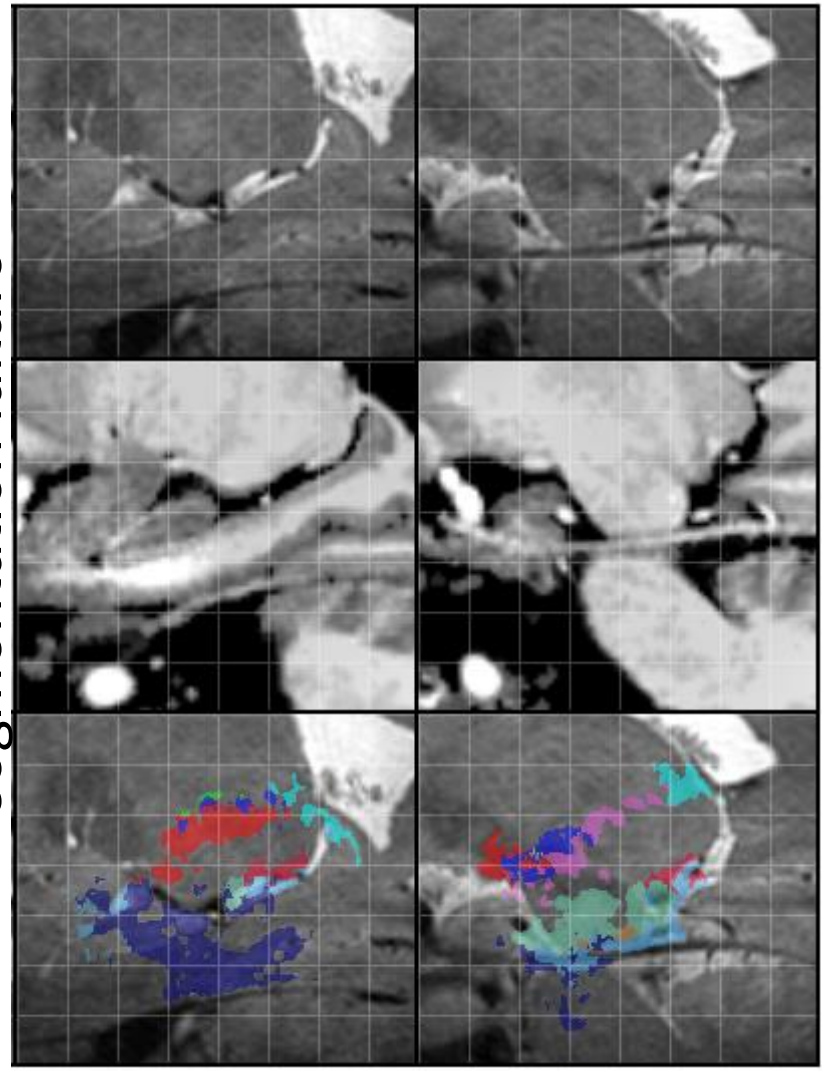
Particularly with a limited field-of-view, identify voxels in key ROIs with good data in FOV at all timepoints.

Catherine Walsh's study

Soon after acquisition

Successful anatomical segmentation





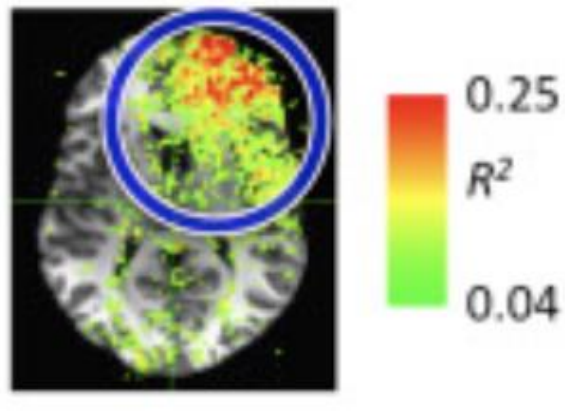
Improved to useability with data preprocessing changes

Catherine Walsh's study

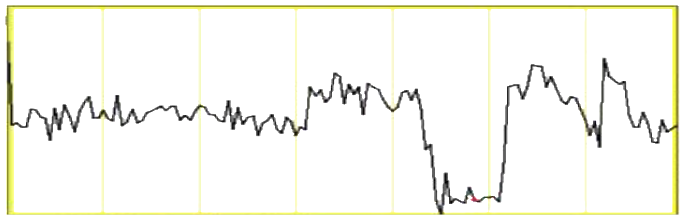
Soon after acquisition

- Result quality similar to pilot scans?
- Data useful for study goals:
  - Good signal quality in brain regions of interest
  - No artifacts that will affect analysis
    - Signs of scanner problems
    - Warnings to reconsider acquisition protocol

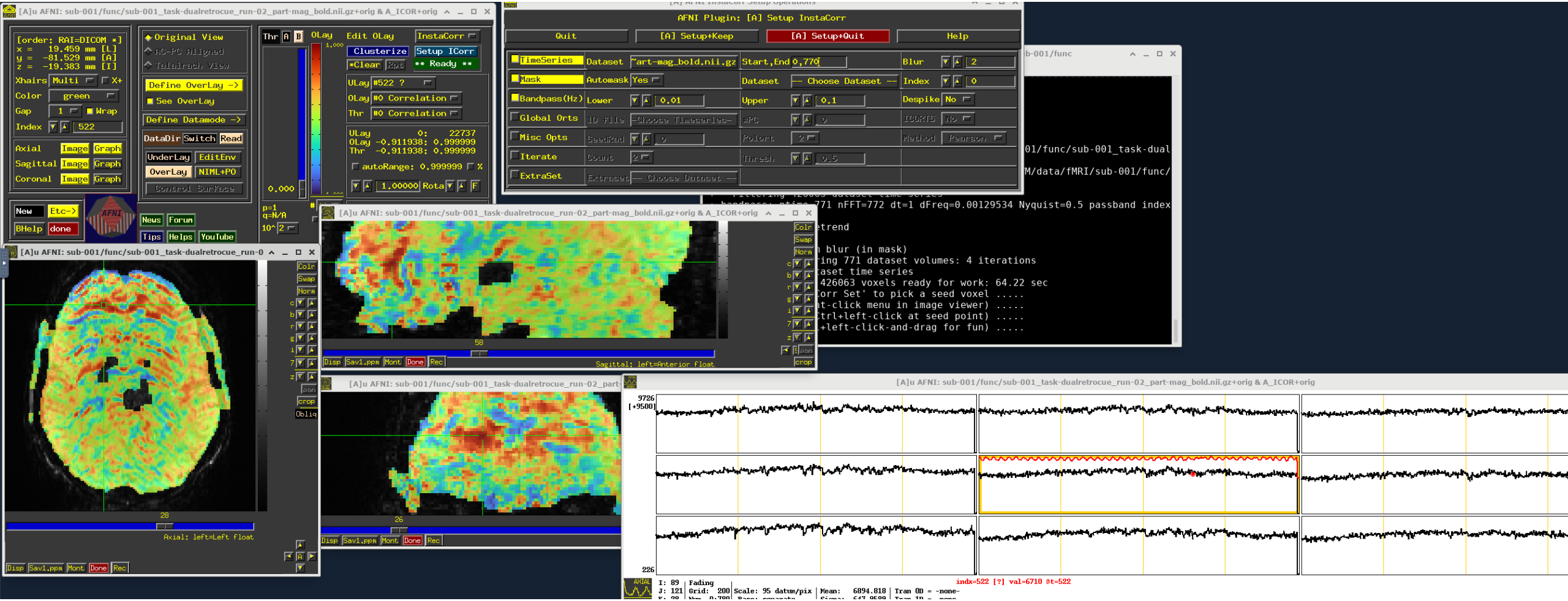
Temporal instability in one receiver coil



ANATICOR
Local correlation to white matter
Jo, Comp Bio & Med (2020)



Soon after acquisition



Potentially concerning, but reduced after preprocessing and results looked reasonable

Catherine Walsh's study

During processing

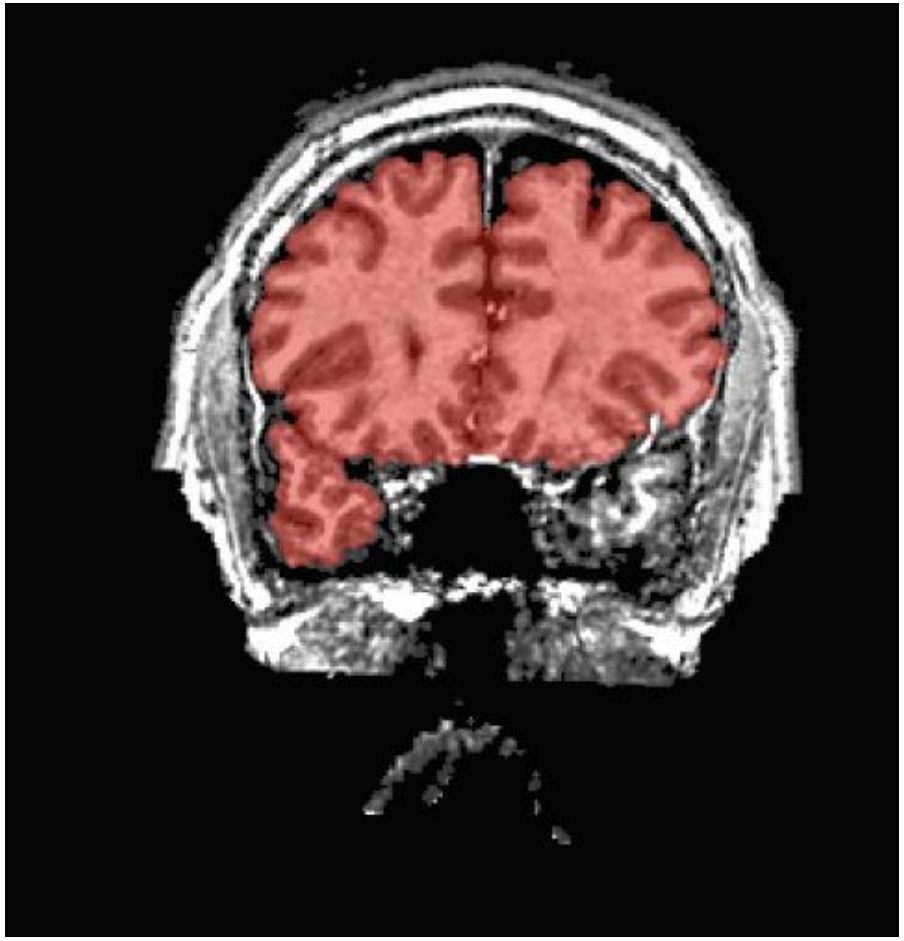
Scripts ran appropriately

```
++ [INFO]: Starting analysis at 2026-06-12 16:08:47
++ [INFO]: Analyzing sub-015, ROI: L MTL.PRC.36, errts.acompcor.noevents with a logistic classifier
++ [QC]: pulling previously masked data from /vf/users/SFIM_fastWM/data/fMRI/derivatives/ROI_data/data_3d/sub-015/acompcor
++ [QC]: pulling onsets from /vf/users/SFIM_fastWM/analysis/onset_files/sub-015/sessionA_3d
++ [QC]: pulling events from /vf/users/SFIM_fastWM/data/fMRI/sub-015/func/sub-015_task-dualretrocueA_acq-3D_events.tsv
++ [QC]: pulling onsets from conditions: ['face', 'scene', 'object']
++ [INFO]: Loading previously masked data
++ [QC]: Found 45 trials over 3 runs
++ [QC]: using 48 training samples
++ [QC]: face: 16, scene: 16, object: 16
++ [INFO]: Training classifier on all data
++ [RES]: Average accuracy in training set: 0.7291666666666666
++ [RES]: Best C: 0.1
++ [RES]: Best number of Voxels: 250
++ [QC]: decoded at 1908 time points
++ [QC]: mean percent signal change -0.018
++ [INFO]: Finished analysis at 2026-06-12 16:09:00
++ [INFO]: Elapsed time 0:00:12.338334
```

In addition to afni_proc execution log, a simplified file showing what was run and on what data it was run

Useful for immediate checks and “provenance” info when revisiting data in the future

During Processing
Individual subject coverage map

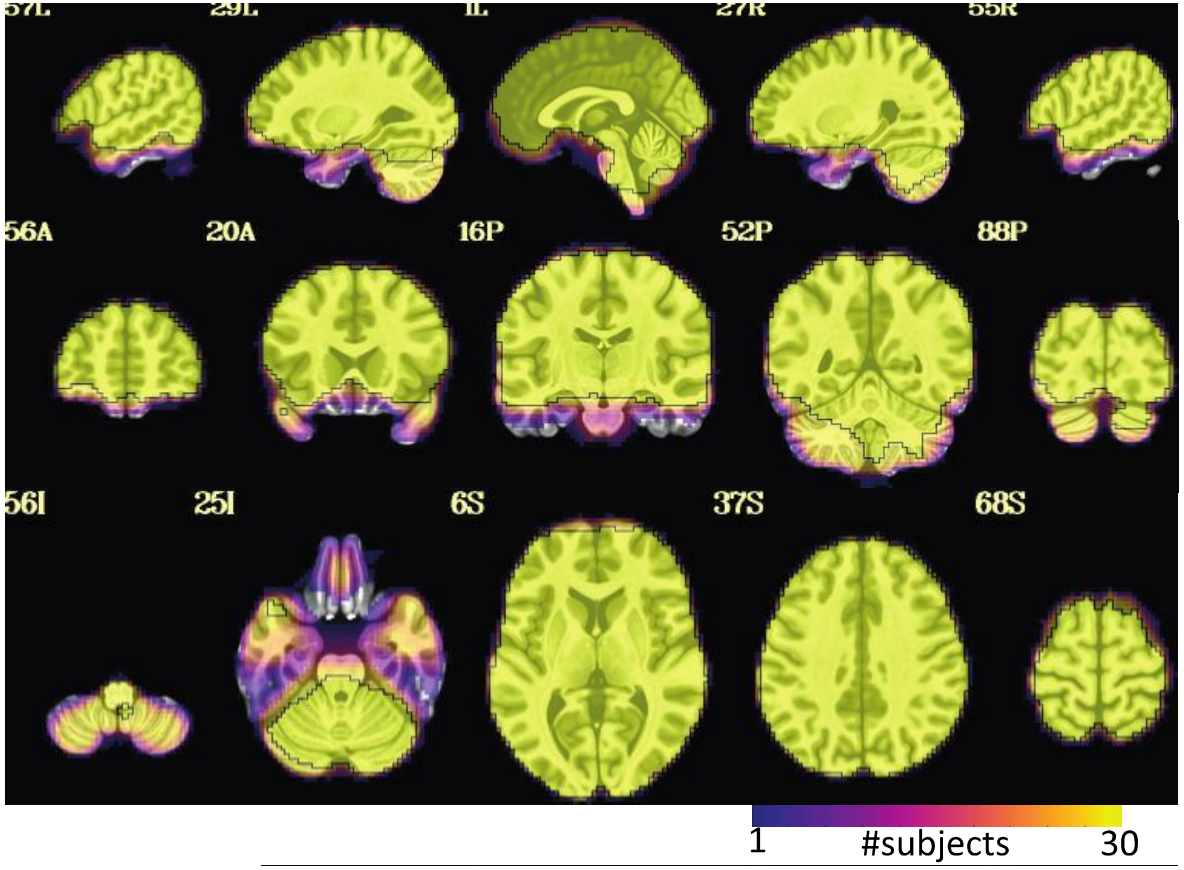


Freesurfer segmentation missed a lobe

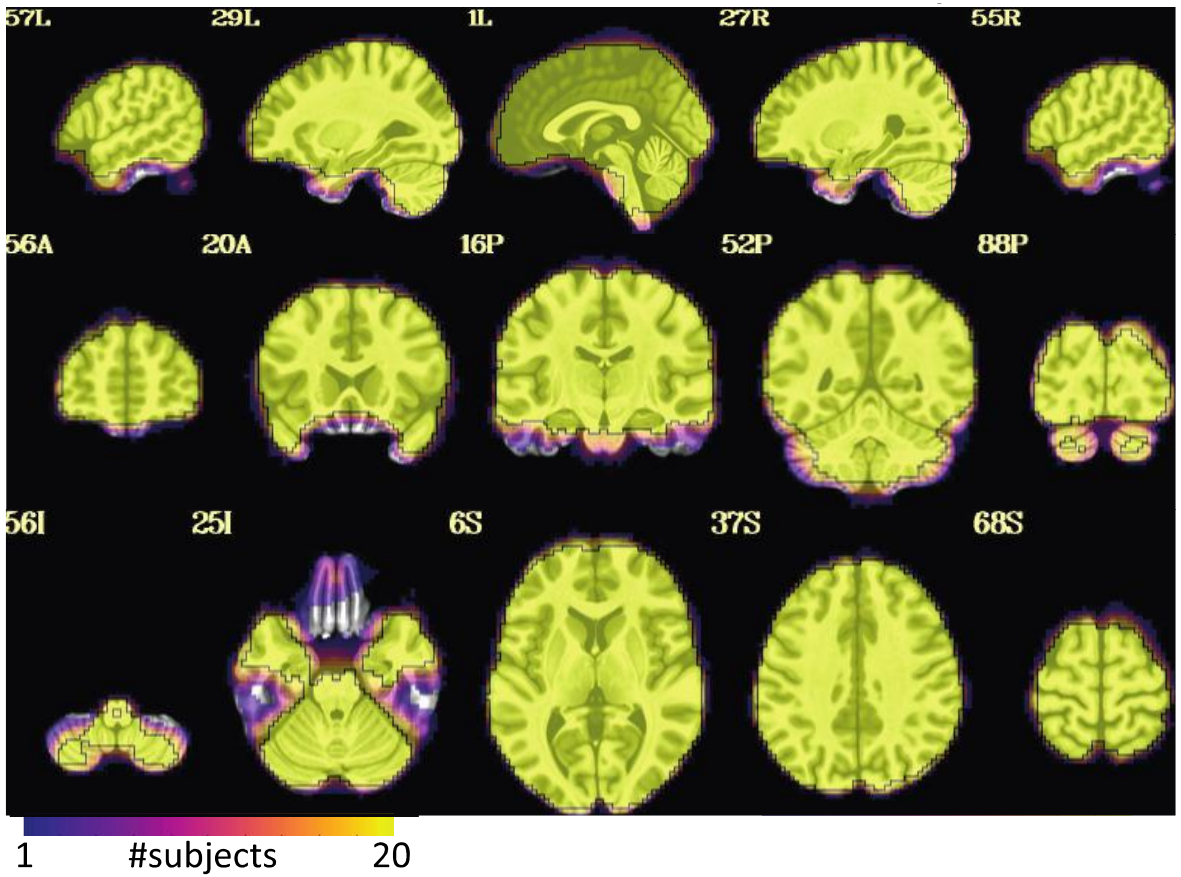
Catherine Walsh’s study

During Processing
Full study coverage map

- Task data



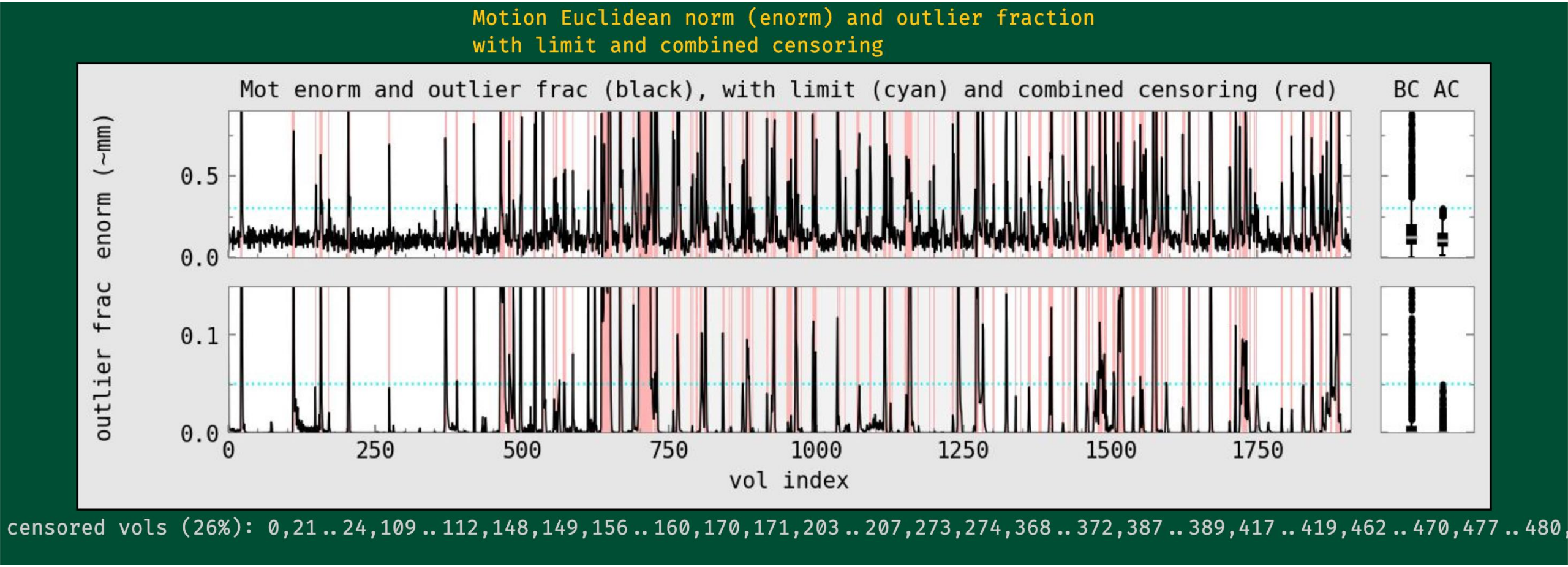
Rest data



Coverage differs between studies: Sufficient" coverage is study-specific

During processing

Head motion & censoring



From afni_proc report

26% of volumes lost

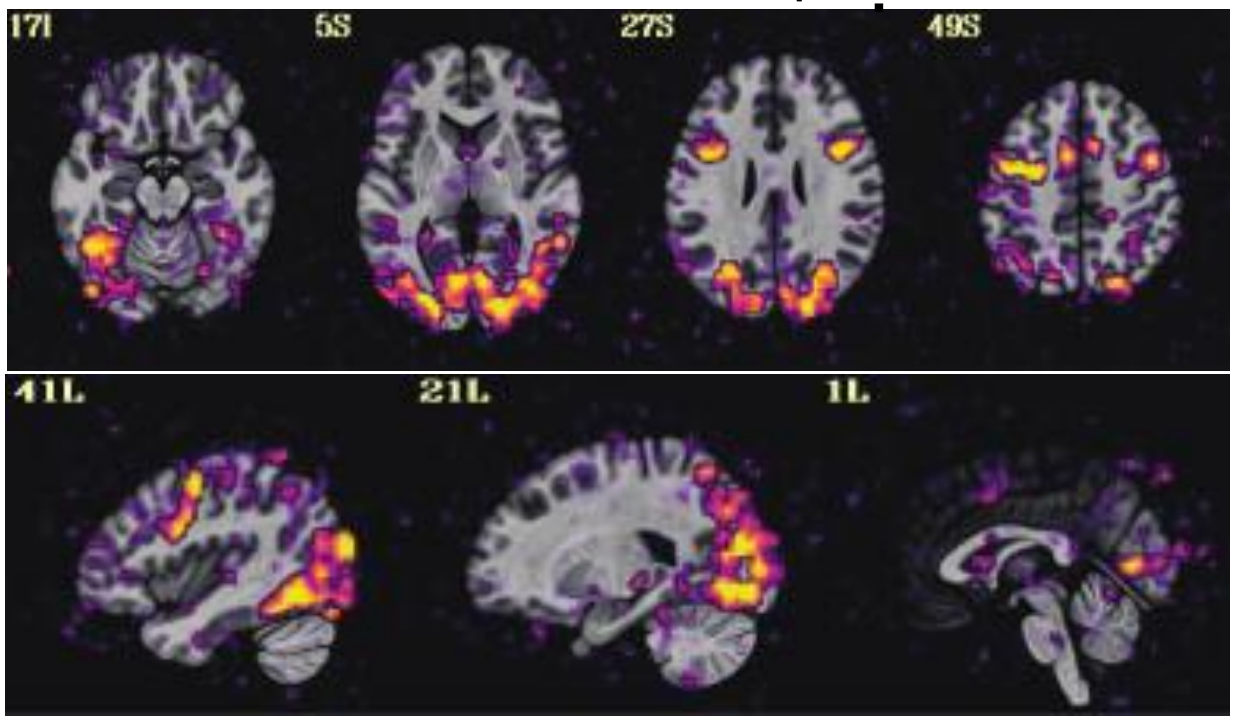
Difficult or impossible to use

Catherine Walsh's study

During processing

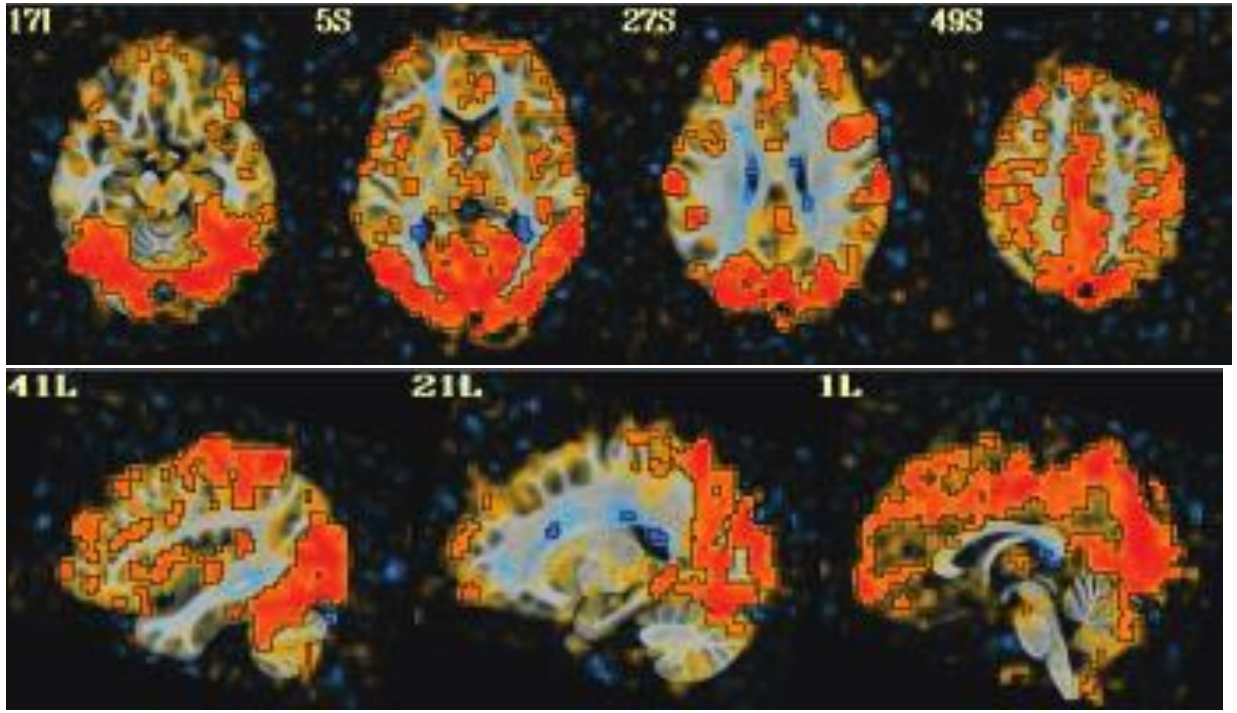
Statistical maps

Full F stat map



- sub-001 (included)

Correlation to white matter ROI

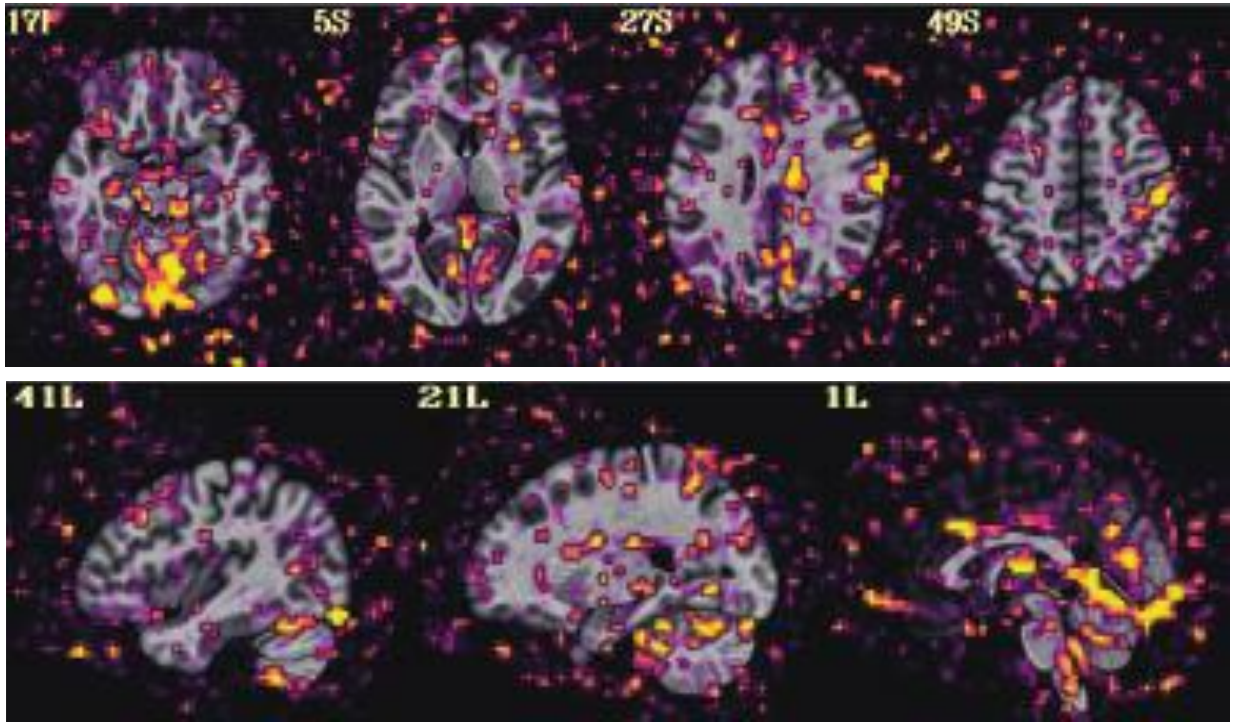


From afni_proc report

Teves et al 2023

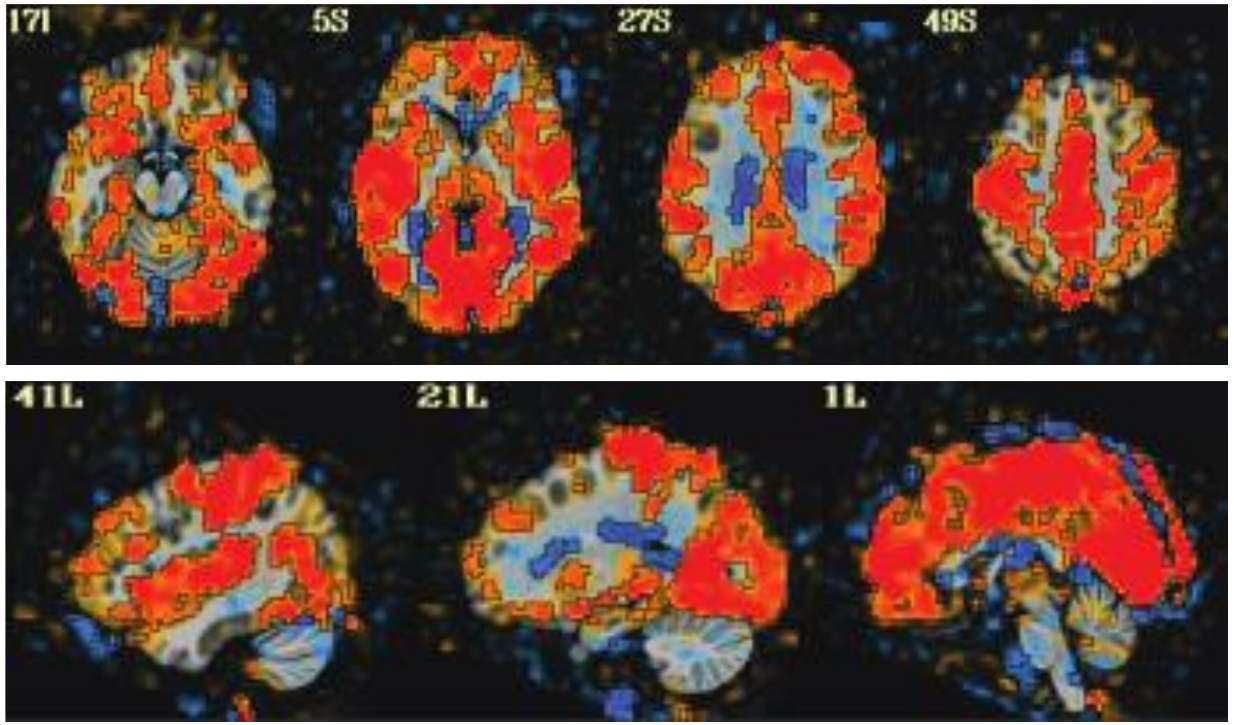
During processing
Statistical maps

Full F stat map



0 6.6

Correlation to white matter ROI



-0.6 0.6

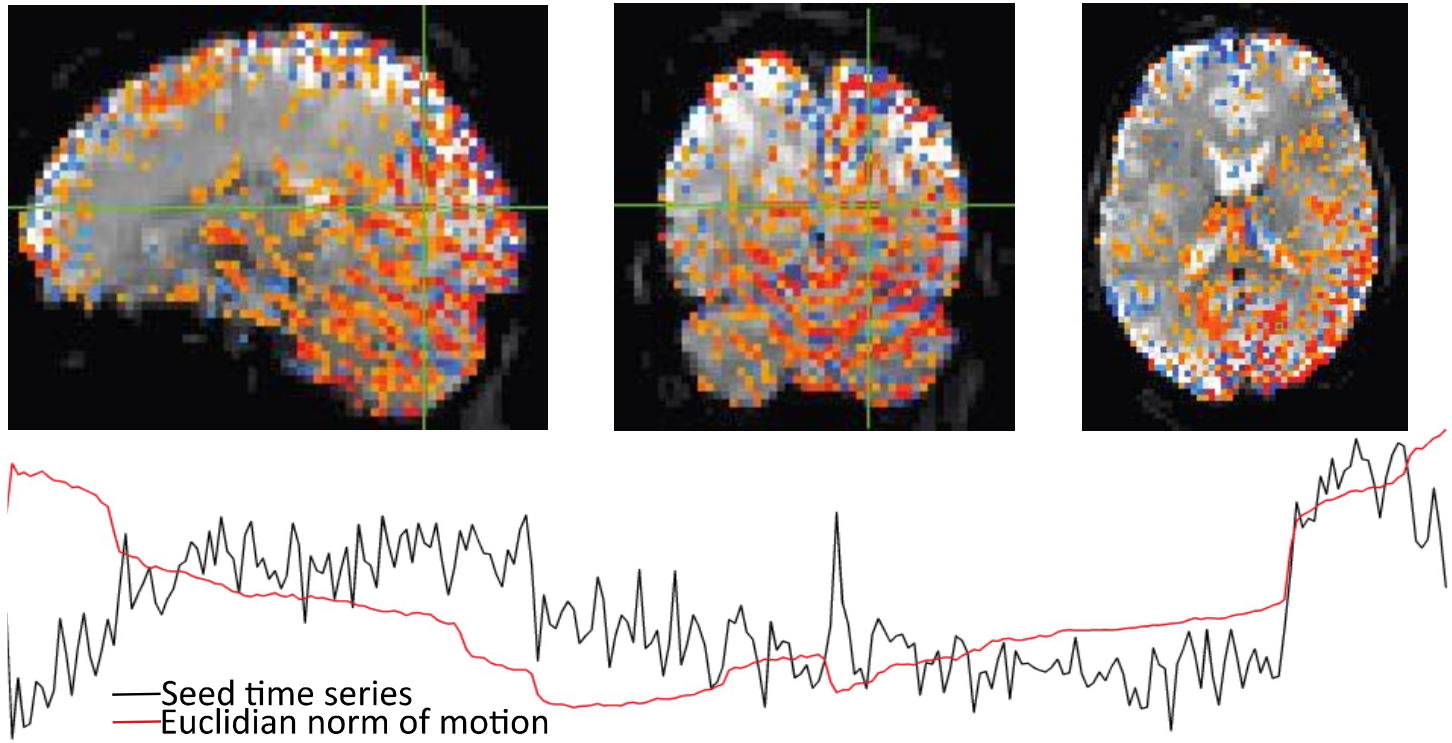
sub-016 (excluded)
Task condition, but not control condition mildly correlated to head motion

Teves et al 2023

During processing

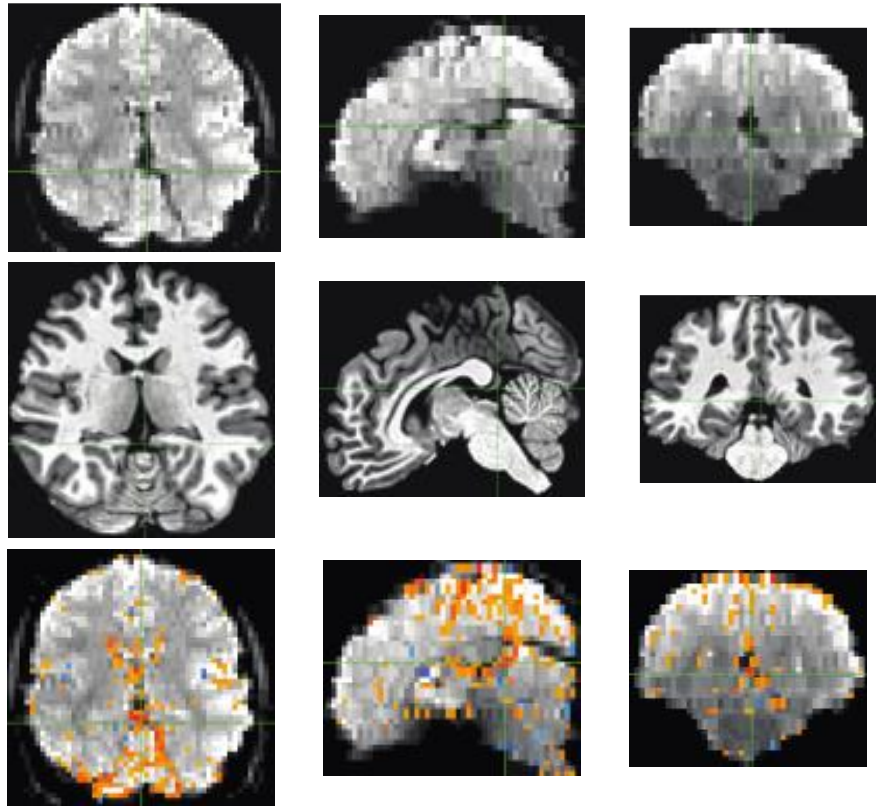
Manual checks with instacorr

A



Sub-018: Investigated artifact with “high EPI variance” warning
Time series follows several jumps in motion

B



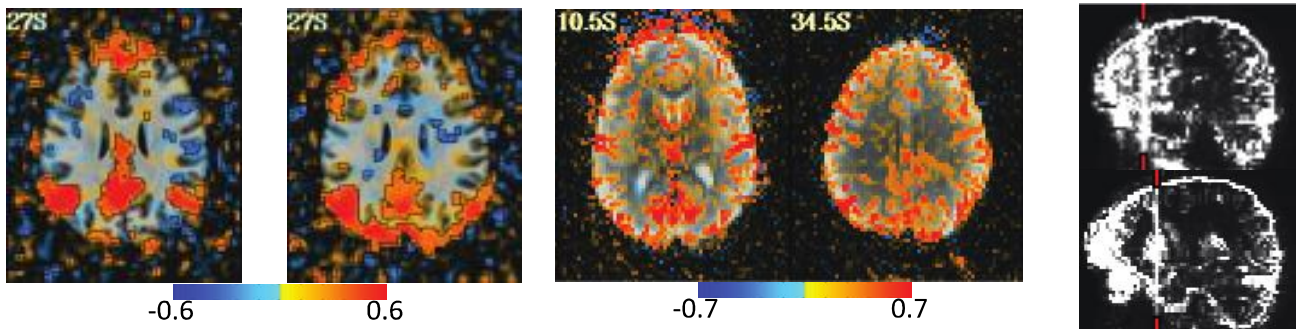
Sub-002: Investigated hypointensity
Correlations mostly to other CSF areas

Teves et al 2023

During processing
More afni_proc report measures

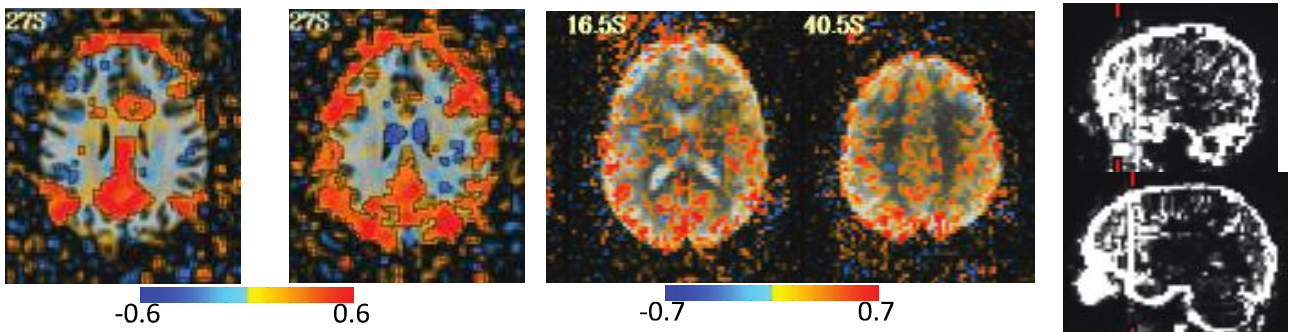
A sub-109 included

Corr to PCC Corr to WM Local corr EPI variance



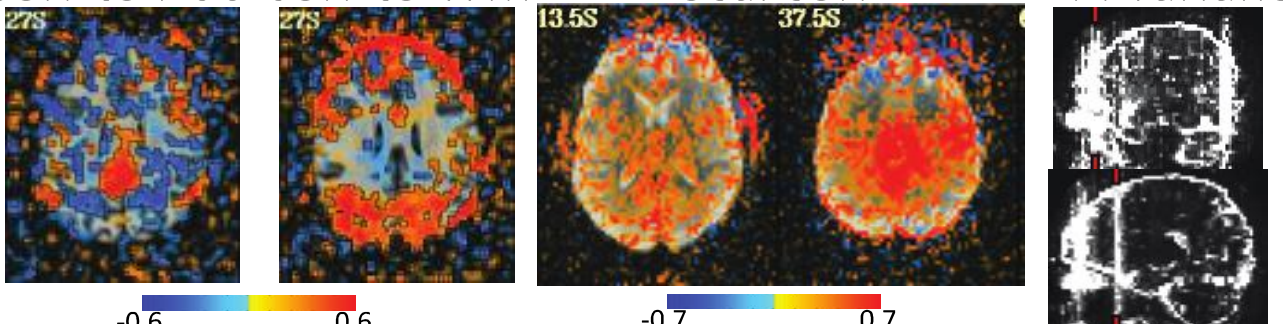
B sub-102 included

Corr to PCC Corr to WM Local corr EPI variance



C sub-114 excluded

Corr to PCC Corr to WM Local corr EPI variance

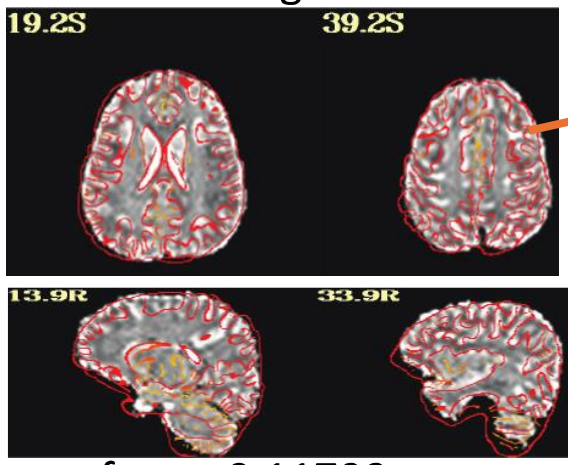


- Expected network plausible?
- Corr to white-matter non-global?
- Local corrs follow anatomy
- EPI variance for follow-up.

Teves et al 2023

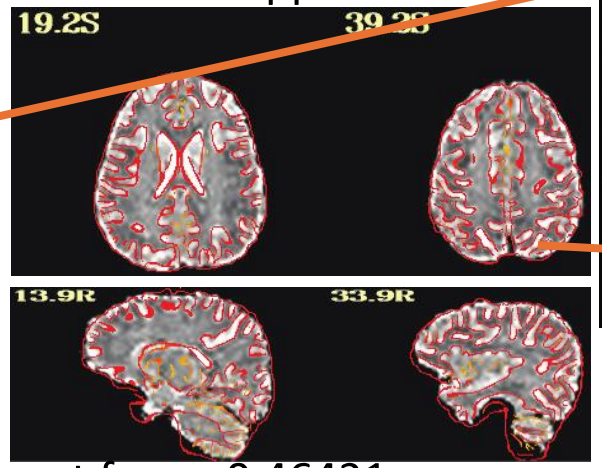
During processing
Alignment

A sub-115 Orig Orientation

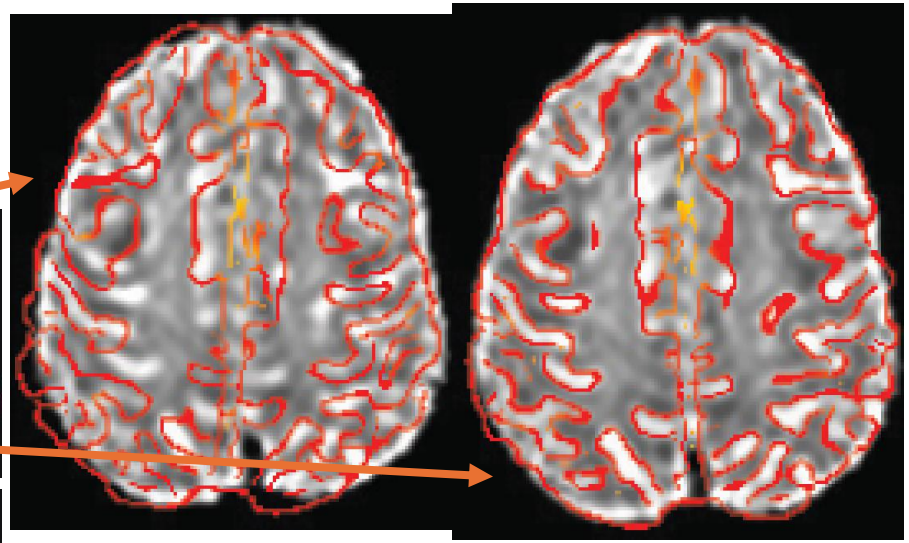


cost func: -0.11799

B sub-115 Flipped Anat

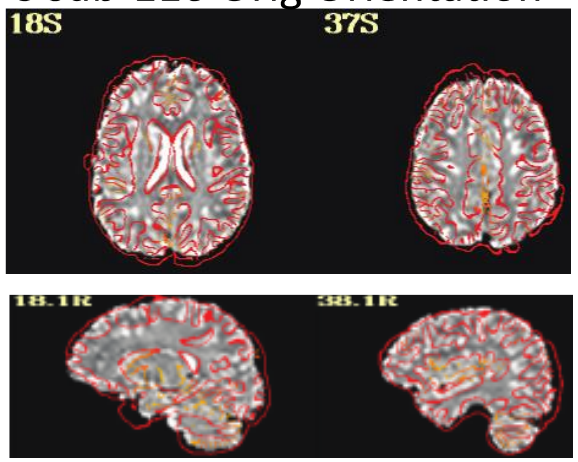


cost func: -0.46431



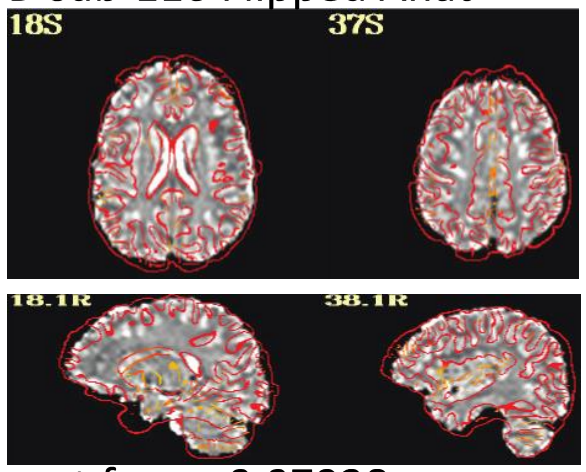
Flipped fits better

C sub-116 Orig Orientation



cost func: -0.07533

D sub-116 Flipped Anat



cost func: -0.07928

Neither fits well

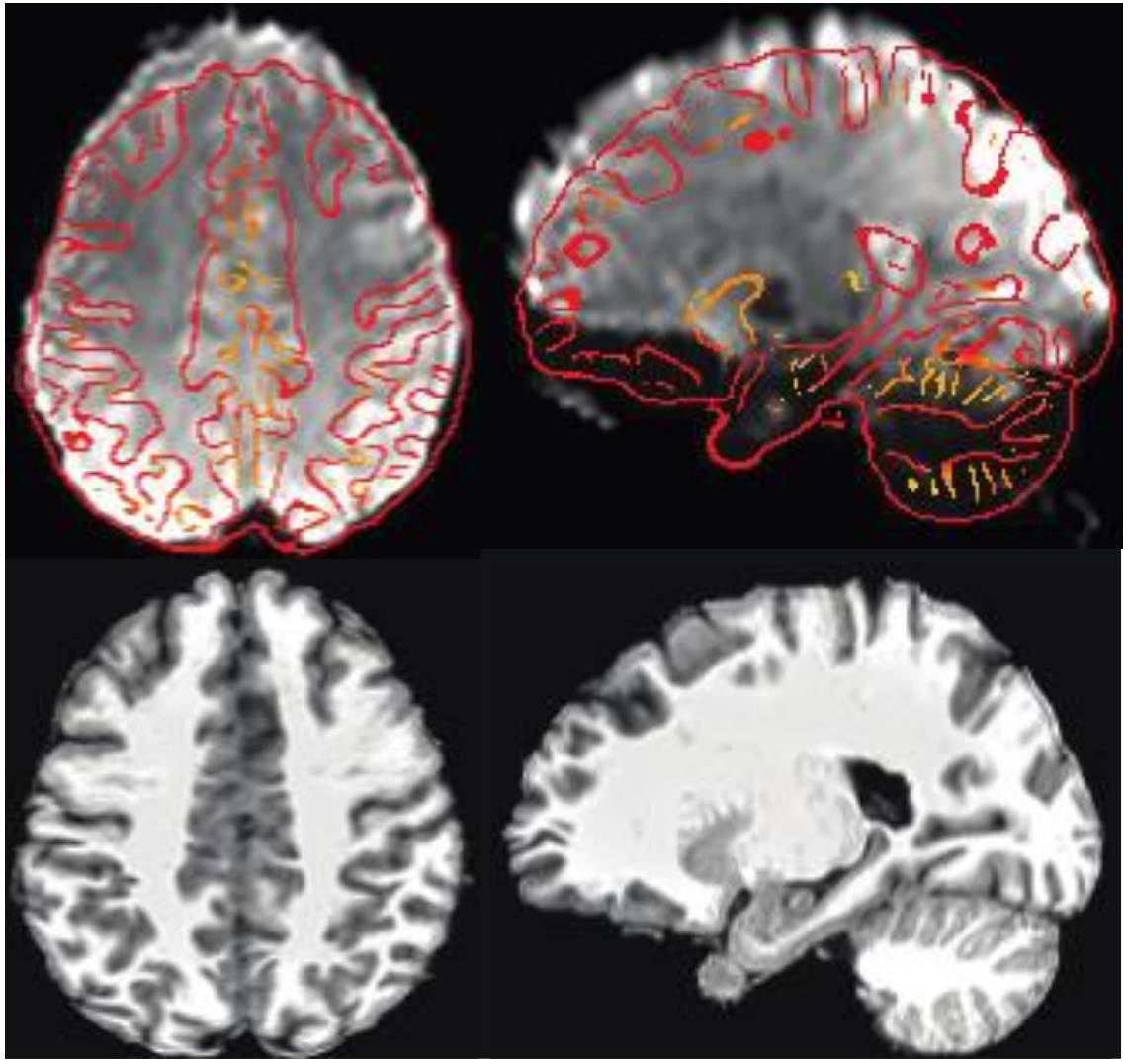
- Anatomical edges over EPI volume is useful for both checking alignment and viewing if there the left & right sides of the brain were flipped

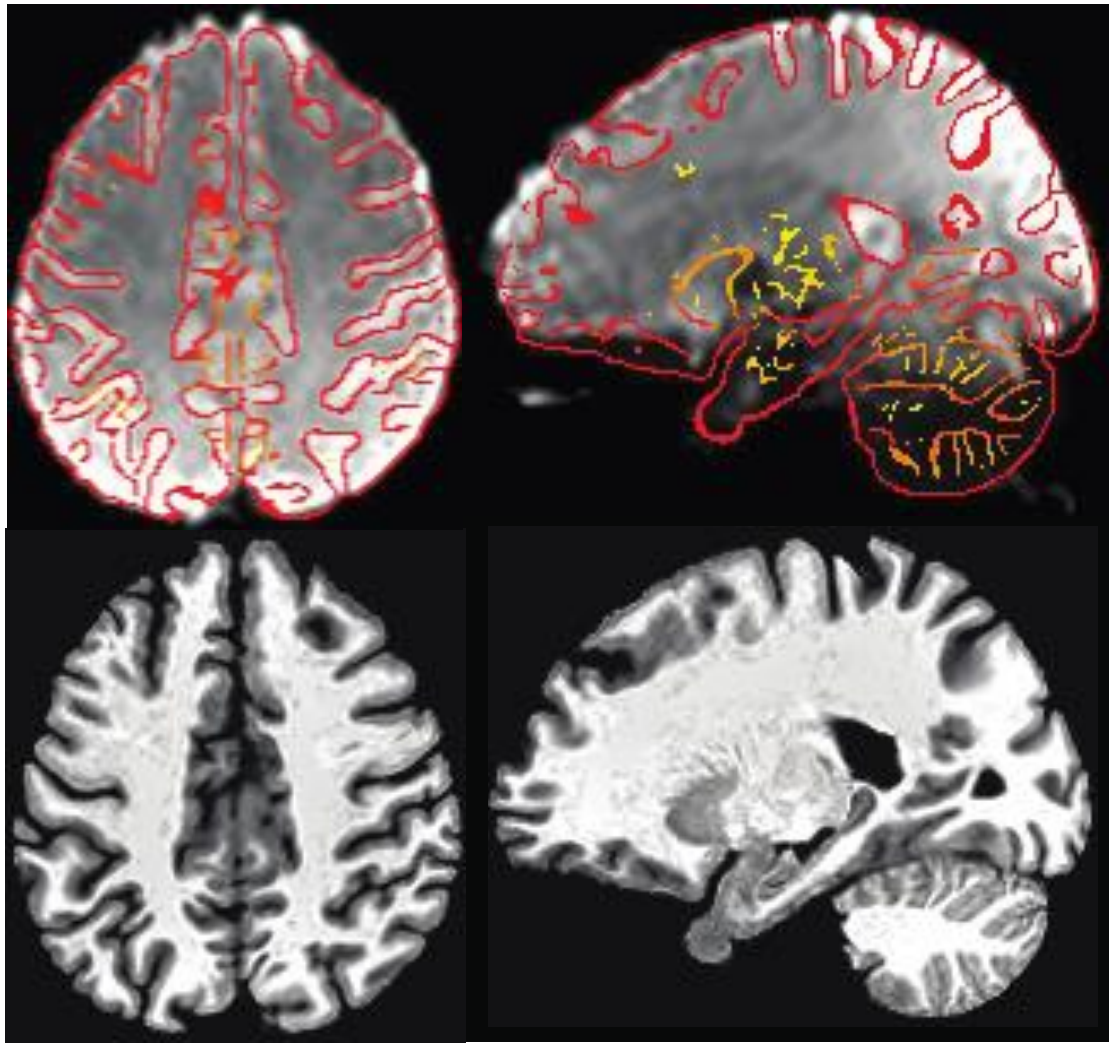
Teves et al 2023

Why you should QA your processing pipeline!

A. Anat & EPI from diff brains

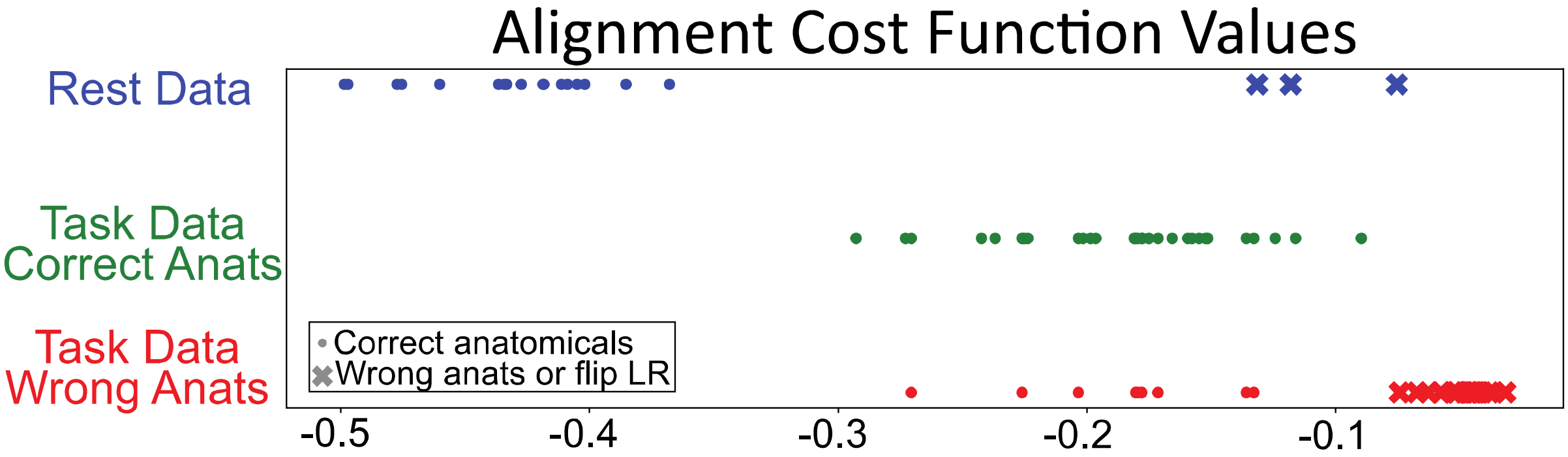
B. Anat & EPI from same brain





Teves et al 2023

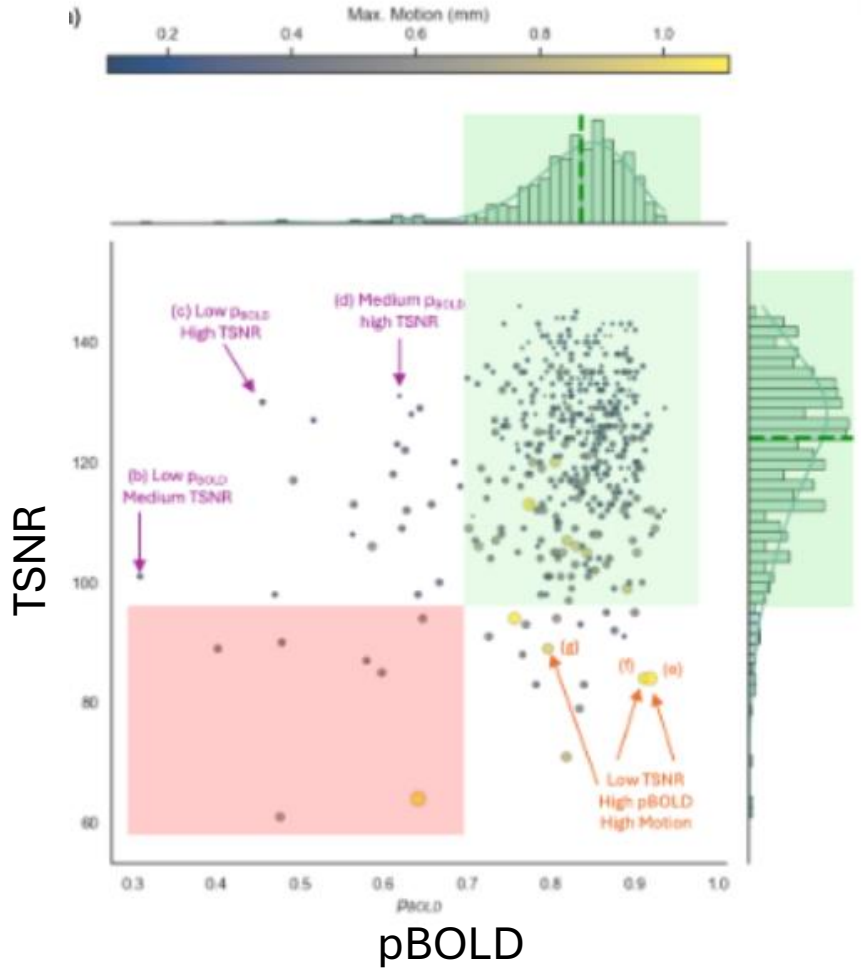
QA is also a research area!



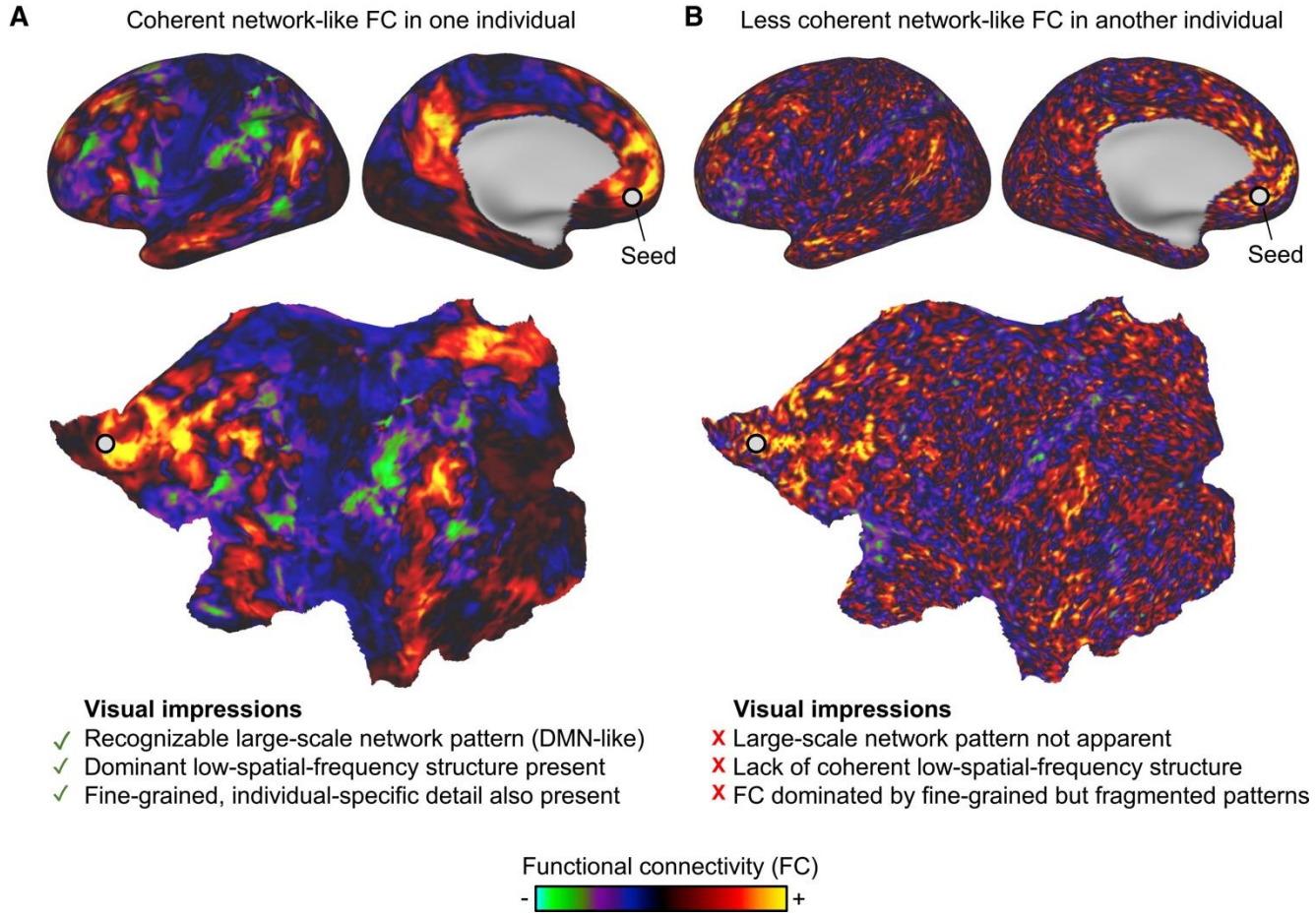
- Alignment cost function used to optimize a relative local minimum.
- The raw values within a dataset seem to highlight bad alignments

Recent QA methods

pBOLD with multi-echo fMRI



Based on assumptions about network connectivity patterns



Gonzalez-Castillo, et al, “A new fMRI quality metric using multi-echo information: Theory, validation and implications” bioRxiv 2026

Lynch, et al “Objective quality assessment for precision functional MRI data” Neuron 2026

Take home messages

Part 1

- Reproducible & relevant science requires knowing your data
  - Pipelines & sample size don’t matter if your data is not good
  - “What is good data” varies with context & application
- Quality assurance **protocols** are central to knowing your data
- Automation is great, but insufficient
- Training on how to do QA needs more attention

Take home messages

Part 2

- Designing a QC process starts with study planning and includes every step of a study
  - Including & especially for using data collected by others
  - Document QC procedures & results particularly for data sharing
- Automated QC must be paired with human observations and interpretations
  - Where human time is most useful is a growing issue for big N studies
- QC is not just keep vs exclude. It's "What questions can I answer with these data?"
- QC is an active area of research & should be more active