A2CPS Imaging Quality Assurance

Plans for assuring and evaluating A2CPS imaging quality

Quality Assurance/Quality Control

Loose but useful definitions for these in our case might be
Quality Assurance (QA): Planning, setup, and process-based means of avoiding problems. In our case, this means making sure the scanners are being maintained, the pulse sequences are appropriate, etc. Much of QA will be handled during initial setup. Another way of looking at QA is that we want to be following "best practices", including those for QC. This document is a key component in the QA plan.
Quality Control (QC): Testing, inspection, and other methods for discovering and dealing with any quality problems. This QA plan specifies some specific QC measures, though others may be decided on and implemented at later stages. QC will not consist of one step, and we will continue to have a number of tools for QC we can use at various stages: for example, testing phantoms to verify that the scanner is operating correctly, checking transfer sizes to identify problems with raw data, testing for motion and other issues in the data, and comparing the final results of processing to identify potential outliers prior to running an analysis.

Acquisition Team Activities

- Note exceptional circumstances on REDCap log during the acquisition
- Check image quality during the acquisition
  - Repeat runs with acquisition failure, if scheduling permits, according to A2CPS Tech Manual
- Transfer images to TACC, according to A2CPS Data Transfer and Management Guide

Analysis Team Activities

There is a need for a rapid decision on acceptability so that sites can delete their copy of scan data after a successful transfer. The proposed framework separates a "fast" quality evaluation track from a more comprehensive "slow" one. A third track is included here to list other related evaluations that will not be ongoing. See Imaging Check Wishlist for details of the implementation of checking.

Track 2 evaluations need not occur on every scan. In addition to manually evaluating every scan flagged by Track 1 checks, we may wish to randomly sample from other scans.

Track 1 ("Fast Track")	Track 2 ("Slow Track")	Track 3 ("Off Track")
Goal: fast, simple, and actionable quality measures generated without human intervention	*Goal: comprehensive evaluation of usability, including manual review, and refinement of Track 1 measures*	*Other targeted checks*

Raw	As with ABCD and UK Biobank, only flagged scans are expected to receive a review (plus all early cases, and potentially a random sampling)	Scanner-related quality checks (weekly phantoms)
Based on BIDS conversion, verify individual acquisitions, and flag missing or truncated scans	Features of interest include	Reliability evaluations using controls
Verify scan parameters	• T1w gray/white contrast, motion artifact, and incidental pathology (if suspected incidental pathology is detected, report details to PI for action)	Outlier detection
Pipelines	• T1w segmentation quality • Visual inspection of fMRI RF artifacts, dropout, and distortion • Inspection of MRIQC/fMRIPrep task fMRI and rsfMRI reports	Comparisons of scans across upgrades
MRIQC	• rsfMRI: calculate connectivity for standard seed locations (e.g. DMN), assess correlations with canonical maps	Comparisons between sites/scanners
fMRIPrep		
QSIPrep		
CAT12		

Analysis Notes: Track 1

Data Conversion and Organization/Naming

We will use HeuDiConv, which uses dcm2niix internally, for the conversion of DICOM to BIDS. The main quality measures will come from MRIQC, with others also drawn from QSIPrep, fMRIPrep, and CAT12. There is a draft publication for CAT12 quality measurement (Dahnke in prep), and another group has published on this as well (Gilmore et al. 2021).

Raw Data Integrity Checks

Scans should result in BIDS data files with identical dimensions every time, and imaging parameters should be the same every time, both of which are testable with a BIDS validator. This is especially an issue after a scanner upgrade, when settings specified as "default", "shortest", etc. in the saved configuration may lead to changed values.

These checks are performed as the last step of BIDS conversion, using https://github.com/a2cps/mri_imaging_pipeline/blob/master/heudiconv_app/assets/check_acq.py. Checked values include a subset of the fields in the JSON sidecars output during BIDS conversion, as well as the bvals and bvecs. The complete reference values are listed in https://github.com/a2cps/mri_imaging_pipeline/blob/master/heudiconv_app/assets/acq-params.tsv. Some numerical parameters are expected to exhibit small differences. These parameters and the magnitude of acceptable deviations are defined in the check_acq.py script. Scans that fail this check proceed through the rest of the analysis pipeline, but an automated warning is sent to the #notifications-imaging-pipeline channel in Slack.

Failures of these checks are investigated manually. Some failures have simple explanations (e.g., a functional scan is short because it was terminated when the participant pressed the squeeze ball). Others may involve extended communication with the imaging site. Discussions start in the slack channel #dirc-case-followups. The status of each investigation is recorded in the table Imaging Log - Problem Cases.

Comparison to Established Norms

The output of MRIQC, or potentially other tools like QSIPrep, can be compared to distributions of values from reference datasets. The MRIQC project provides the MRIQC Web API (https://mriqc.nimh.nih.gov/) for this purpose, and Scott Peltier at UMich has offered to supply data.

Anatomical

For the purposes of understanding the progress of the project and identifying problems, anatomical scans will receive a rating of "green", "yellow", or "red". Scans with a green or yellow rating are expected to be viable sources of biomarkers. Scans are marked yellow rather than green when they exhibit a minor quality issue that may need special consideration in some analyses. Scans marked red are not necessarily unusable, but extracting biomarkers from them would likely require bespoke processing.

Labels are assigned initially as follows. By default, an image is marked as green. In Track 1, that rating can be downgraded in two ways

1. The technologist may label the image as red or yellow.
    a. For examples, please see the A2CPS Tech Manual
    b. Images that received a rating of red will have been repeated once. The label used here will be based on the final anatomical scan only.
2. An image may be downgraded from green to yellow if it receives an IQR from CAT12 is below a relaxed threshold (80%), and further downgraded to red if the IQR is below a stringent threshold (e.g., 60%).

The resulting rating for the image is therefore the lowest rating that was given by the technologist or CAT12. These ratings will be refined in Track 2.

Functional

As with the Anatomical images, Functional images will receive an initial rating of "green", "yellow", or "red". The technologist does not rate these scans. Instead, the automated downgrading from green to yellow or red is based on the following:

1. Motion
    a. Images will be downgraded to yellow if the framewise displacement measures produced by fMRIPrep are such that any of the following is true
        i. average displacement > 0.25mm
        ii. more than 20% of the individual displacements are above a threshold, where the threshold is 0.3mm for REST and 0.9mm for CUFF.
        iii. any individual displacement is greater than 5mm
    b. Images will be downgraded to red if the average framewise displacement is above 0.55mm
    c. These motion thresholds are based on Parkes et al. (2018).
        i. Note that there are two common but different methods of calculating framewise displacement: the "Power" method and the "Jenkinson" method. The thresholds from Parkes et al. are based on the Jenkinson method, which is also what is reported by FSL's MCFLIRT. MRIQC produces only motion according to the "Power" method, which is in general much higher (that is, scans appear worse). fMRIPrep produces both, and it refers to the Jenkinson method as rmsd.
2. Integrity Checks
    a. Truncated scans are downgraded to red

Analysis Notes: Track 2

Anatomical

In this track, the rating from the "Fast Track" is confirmed, and more detailed information is provided.

All images that fail the Raw Data Integrity Check will be reviewed manually. Scans that only fail these checks but exhibit no artifacts will be rated on a case-by-case basis.

All images receiving a rating of yellow or red will be reviewed manually. These reviews are not expected to alter many ratings, but are instead primarily to confirm the lack of obvious error in the Fast Track (e.g., in case the technologist is recorded as having rated an image as "red" but there is no obvious artifact, suggesting a data-entry error or CAT12 gives a low IQR due to correctable segmentation failure).

Additionally, any image will be reviewed manually if it exhibits at least one image quality metric that is anomalous for the distribution of metrics available at that site. Metrics are calculated by MRIQC, outliers are defined in https://github.com/a2cps/mri_imaging_pipeline/blob/master/aggregator_qc_app/assets/check_qc.py, and images with an outlying metric are listed at QC Aggregation.

Reviews rely on the HTML reports produced by MRIQC, but those reports may be supplemented by the HTML report from fMRIPrep or the NiFTI itself. The product of a manual review is a JSON file produced by the MRIQC "Rating Widget". This widget provides a standardized list of artifacts and a summary of the image in a four-category numerical rating (1: Exclude; 2: Poor; 3: Acceptable; 4: Excellent). The summary should be based on both A) the worst rating from Track 1 (after confirming that the rating is not due to data entry error or correctable processing failure), and B) the severity of the worst artifact present in the image, so that the summary rating is based on the minimum (valid) of Tracks 1 and 2.

The severity for the following artifacts is rule-based

- Field-of-View (including wrap-around)
    - Acceptable: wrap-around that does not overlap with the cerebrum, or a field-of-view that excludes the cerebellum
    - Poor: wrap-around that overlaps with a small portion of the cerebrum, or a field-of-view that excludes a small portion of the cerebrum
    - Exclude: substantial wrap-around or exclusion of the cerebrum
- Eye spillover through the PE axis
    - Acceptable: Not used
    - Poor: Visible spillover
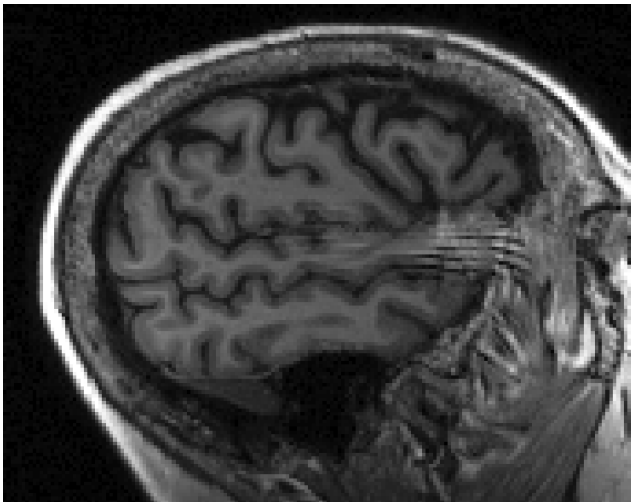
- Exclude: Not used

The severity of the remaining artifacts will require a judgment call. The four category ratings will be mapped to the red/yellow/green labels (0 red; 1 yellow; 2,3 green), and so the color rating should be used to guide choices. For judging the severity of artifacts like ringing, the examples from the A2CPS Tech Manual are used as a reference.

JSON files from reviews are stored on TACC at /corral-secure/projects/A2CPS/shared/psadil/qclog/mriqc-reviews. These JSONS are aggregated nightly with the aggregator_qc_app (specifically: https://github.com/a2cps/mri_imaging_pipeline/blob/master/aggregator_qc_app/assets/update-qclog.py) and recorded in the QC Log.

Note that, because manual reviews can be triggered by a scan exhibiting an outlying metric, individual scans may move on and off the list of outliers. Removing an image from the list of images with an outlying metric will not affect the rating of red/yellow/green.

Individual images receiving a rating of red will be brought to the attention of the DIRC. The final red/yellow/green rating will be presented in the Imaging Overview Report. That report will also contain more detailed information for the individual MRIQC image quality metrics.
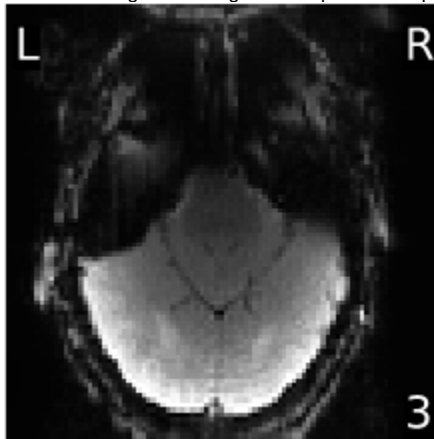
Functional

All images that fail the Raw Data Integrity Check for reasons other than being truncated will be reviewed manually. Scans that only fail these checks but exhibit no artifacts will be rated on a case-by-case basis.

All images receiving a rating of yellow or red due to motion will be reviewed manually. These reviews are not expected to alter many ratings but are instead primarily to confirm the lack of obvious error in the Fast Track (e.g., inaccurate segmentation resulting in poor motion estimation).
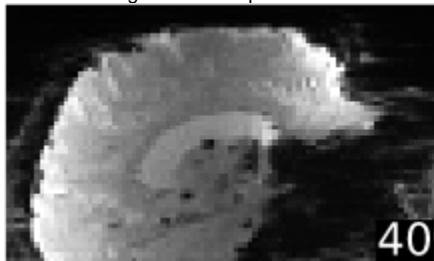
Additionally, any image will be reviewed manually if it exhibits at least one image quality metric that is anomalous for the distribution of metrics available at that site. Metrics are calculated by MRIQC, outliers are defined in https://github.com/a2cps/mri_imaging_pipeline/blob/master/aggregator_qc_app/assets/check_qc.py, and images with an outlying metric are listed at QC Aggregation.

The severity for the following artifacts is rule-based:

- Field-of-View (including wrap-around)
  - Acceptable: wrap-around that does not overlap with the cerebrum, or a field-of-view that excludes the cerebellum
    - functional image exhibiting inconsequential wrap-around (NS10128V1 REST1)
   -
  - Poor: wrap-around that overlaps with a small portion of the cerebrum, or a field-of-view that excludes a small portion of the cerebrum
    - functional image with small portion of cerebrum excluded (NS10075V3 REST1)
      -
  - Exclude: substantial wrap-around or exclusion of the cerebrum
- Uncategorized. Scans with any of the following will have the 'uncategorized' tag in the MRIQC rating jsons
  - Banding along z-axis (both narrow and wide)
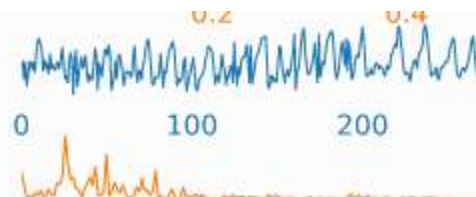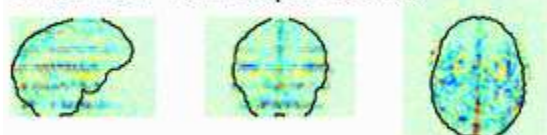   - Acceptable: This category is not used to downgrade the rating
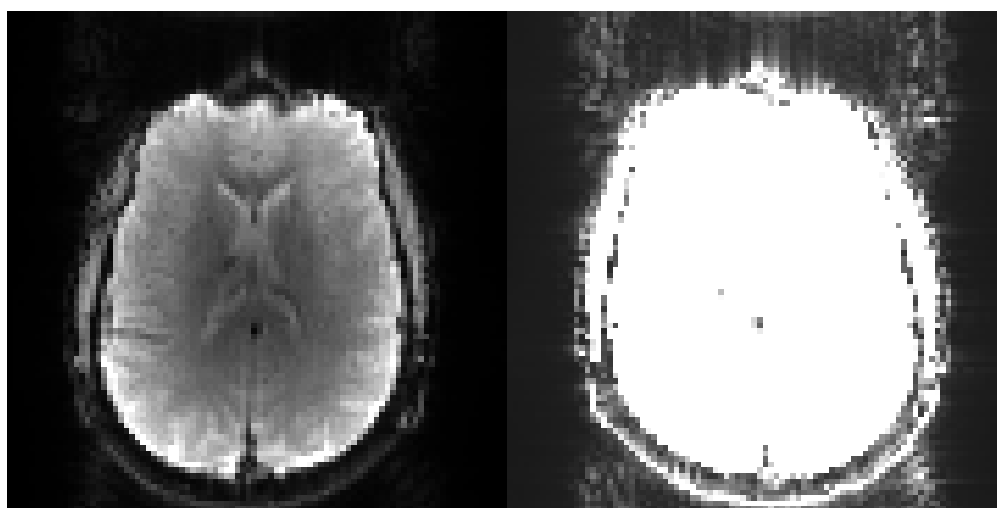
C13: Tot. var. expl. 1.4%
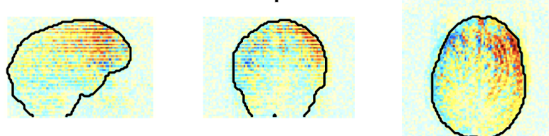
C15: Tot. var. expl. 1.3%

C46: Tot. var. expl. 0.53%

- - Poor: not used
  - Exclude: not used
  - Other Ghosts. Present on most/all images from NS. See 2021-12-10 Imaging Sites Meeting notes
    - Acceptable: This category is not used to downgrade the rating.

- - Poor: not used
  - Exclude: not used

Manual reviews will proceed as for the Anatomical scans, as will their recording and follow-up.

The summary should be based on both A) the worst rating from Track 1 (after confirming that the rating was not due to correctable processing failure), and B) the severity of the worst artifact present in the image, so that the summary rating is based on the minimum (valid) rating of Tracks 1 and 2.

Analysis Notes: Track 3

Scanner-Related Quality Checks

Currently, sites are handling phantom checks, with scan data collected at sites shared with TACC. We run several measures, including the QC regime outlined in (Friedman & Glover 2006), on all such phantom scans. Though the main interest is in changes over time for each site, there is a degree of standardization across sites – all (except UM, which uses the ABCD QA protocol) use a 17 cm diameter spherical volume (DSV) system phantom from GE or Siemens (UC, the sole Philips site, uses a GE phantom and other sites use one from the scanner manufacturer).

Useful Prior Work

ABCD

The ABCD study has developed a QC approach that includes both an automated QC check and a manual review of raw data of all types, the latter with a standardized qualitative rating scheme based on multiple keywords (e.g. "mo" for motion or "sus" for susceptibility-related issues, with a numerical rating of 0-3 for problems: so a possible coding is "mo1 and sus2" for mild motion plus moderate susceptibility artifacts). See attached ABCD RawQC document and attached ABCD QC Release Note.

The ABCD QC process seems to consist of three basic steps: (1) an initial assessment by the acquiring tech (a decision to accept or reject is made based on their notes), (2) an automated QC evaluation (the ABCD QC Release Note gives suggested thresholds for acceptance), and (3) a manual evaluation process described in both the attached documents above. Here is the description from the online ABCD documentation:

> *QC is performed by scan operators at the time of the scan. Subjects may fail for a variety of reasons whether the subject was moving, began talking, fell asleep, used the squeeze ball, the scan had to be stopped, etc. These operator notes are taken into account and each scan is given a binary pass or fail. This QC information is now provided by the NIMH Data Archive (NDA) and instructions to download it can be found on the ABCD repository for downloading and setting up the BIDS dataset (ABCD-STUDY /abcd-dicom2bids). Only scans that passed this initial QC were considered for processing. Subjects without a passing T1w image were excluded from this dataset. If there were valid anatomical scans, but no functional scans the subjects were processed through an anatomical-only version of the pipeline that excludes any functional image processing.*

UK Biobank

The paper "Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank" (Alfaro-Almagro et al. 2018, https://www.sciencedirect.com/science/article/pii/S1053811917308613) describes an automated QC pipeline. Their reasoning is that manual QC, while useful to have, is simply not feasible for all scans in large-scale studies: "the sheer quantity of imaging data which will be produced by UK Biobank (100,000 subjects) makes QC via visual inspection unfeasible". However, they do suggest that manual QC can be applied after a system like theirs is applied (with the parameterization they describe, the automated QC has a relatively high false positive rate of ~15%, but they reason that checking 15% of a large subject pool manually is probably reasonable).

The basic initial checks in the pipeline occur after source DICOMs are converted to NIfTI (using dcm2niix) and organized into a BIDS structure (using custom scripts?). They consist of the following:

> *After the main file organisation, a basic Quality Control (QC) tool is used to check if every raw dataset in each modality has the correct dimensions. When raw data has the wrong dimensions, is corrupted, missing or otherwise unusable, it is moved into a sub folder called "unusable" (inside the given modality's folder), and not processed any further (apart from defacing applied to the raw T1 and T2 FLAIR). This "unusable" data is included in the NIFTI packages in Biobank database because some researchers may be interested in working with it, for example, to develop methods for detecting or even possibly correcting such data. In the case of unusable T1 data, the raw imaging data for all other modalities is deemed unusable (because the pipeline cannot function without a usable T1). However, as with the T1 data, all such raw data is still available for NIFTI download, but without any processing applied.*

The main QC checks use a supervised learning approach in the Weka machine learning toolbox, in which categorically labeled exemplar data using a scheme that codes both type of issue (or overlap of issues) and severity ("problem" that makes processing impossible or unreliable, or "imperfection" that is less severe). The QC features used to drive the categorization are 190 measures like "asymmetry between subcortical structures" or "volume of white matter hyperintensities".

Problem code	Imperfection code
0 = No problem	1 = Multiple/Unknown imperfections
1 = Multiple/Unknown problems	2 = Bad head movement
2 = Missing or incomplete modality	3 = Movement-related ringing/blurring
3 = Bad FOV	4 = Bias field/contrast problem
4 = Bad registration: Bad head motion/Noise	5 = Structurally atypical
5 = Bad registration: Structurally atypical: Important lesions	6 = Problem on top (brain mask)
6 = Bad registration: Structurally atypical: Big Ventricles	7 = Problem on temp. lobe (brain mask)
7 = Bad registration: General Registration failure	
8 = Bad registration: Bad brain mask on the top	
9 = Bad registration: Bad brain mask on the temporal lobe	
10 = Bad registration: Brain mask out of the brain	

Of note, while their rate of passing QC on T1-weighted scans was 98.1%, only 81.1% of subjects passed QC on all scans (diffusion imaging was worst, with 87.2% passing).

References

Dahnke R, Ziegler G, et al. Segmentation-Based Quality Assessment of Structural T1 Images. In preparation. Preprint from 2019.

Dietrich O, Raya JG, et al. Measurement of signal-to-noise ratios in MR images: Influence of multichannel coils, parallel imaging, and reconstruction filters. JMRI 26:375-385 (2007).

Friedman L; Glover GH. Report on a multicenter fMRI quality assurance protocol. JMRI 23:827-839 (2006).

Gilmore AD, Buser NJ, et al. Variations in structural MRI quality significantly impact commonly used measures of brain anatomy. Brain Inform. 2021 Dec; 8 (1): 7. PMID: 33860392.

Dahnke et al. 2019

Dahnke-RQA.pdf