Multi-site MRI phantom data collection initiative for QA and assessing effects on morphometric and functional data.

Initiative from the INCF WG on quality control: https://www.incf.org/sig/incf-working-group-neuroimaging-quality-control lead by Pradeep Raamana

Some of this work can/might be turned into a best practice report for research MRI QC that can be discussed by the OHBM best practice committee.

Coordination: Dr Cyril Pernet

Anyone collecting data for this project and later analyzing data can be co-author on any related paper. Simply add names with institutions in section 3.B.

1. Introduction
2. Goal
3. Method
3.A. Sites and protocols
3.B. Data analysis

1. Introduction

Although QC is used to check that scanners are working properly, avoiding artefacts at the subject level, QA can also be used to account for variations in the data, at the group level. This is well exemplified by https://f1000research.com/articles/9-1131/v1, in which phantom SNR explains some of the gray matter volume changes. Since group data acquisition is rarely sampled homogeneously over time, changes in scanner baselines can affect group results in unknown ways.

2. Goal

The main goal is to study the impact of scanner changes on group level results.
1 - looking at SNR and effects on gray matter volumes (replicating https://f1000research.com/articles/9-1131/v1), but also thickness, surfaces, etc. This can be done within and between groups.
2 - looking at fMRI QA (e.g. tSNR) and effects on task BOLD and connectivity - to test if QA is important in evaluating group differences.

A secondary usage of this data is establishing QC norms (*a good backup plan, shall the primary question fail!)*; simply having a set of data and metrics available would allow anyone to compare scanner performances. Although simple, it is reassuring to know how everyone performs.

Incidentally, it will require some data management/BIDS work, which together can be a single paper.

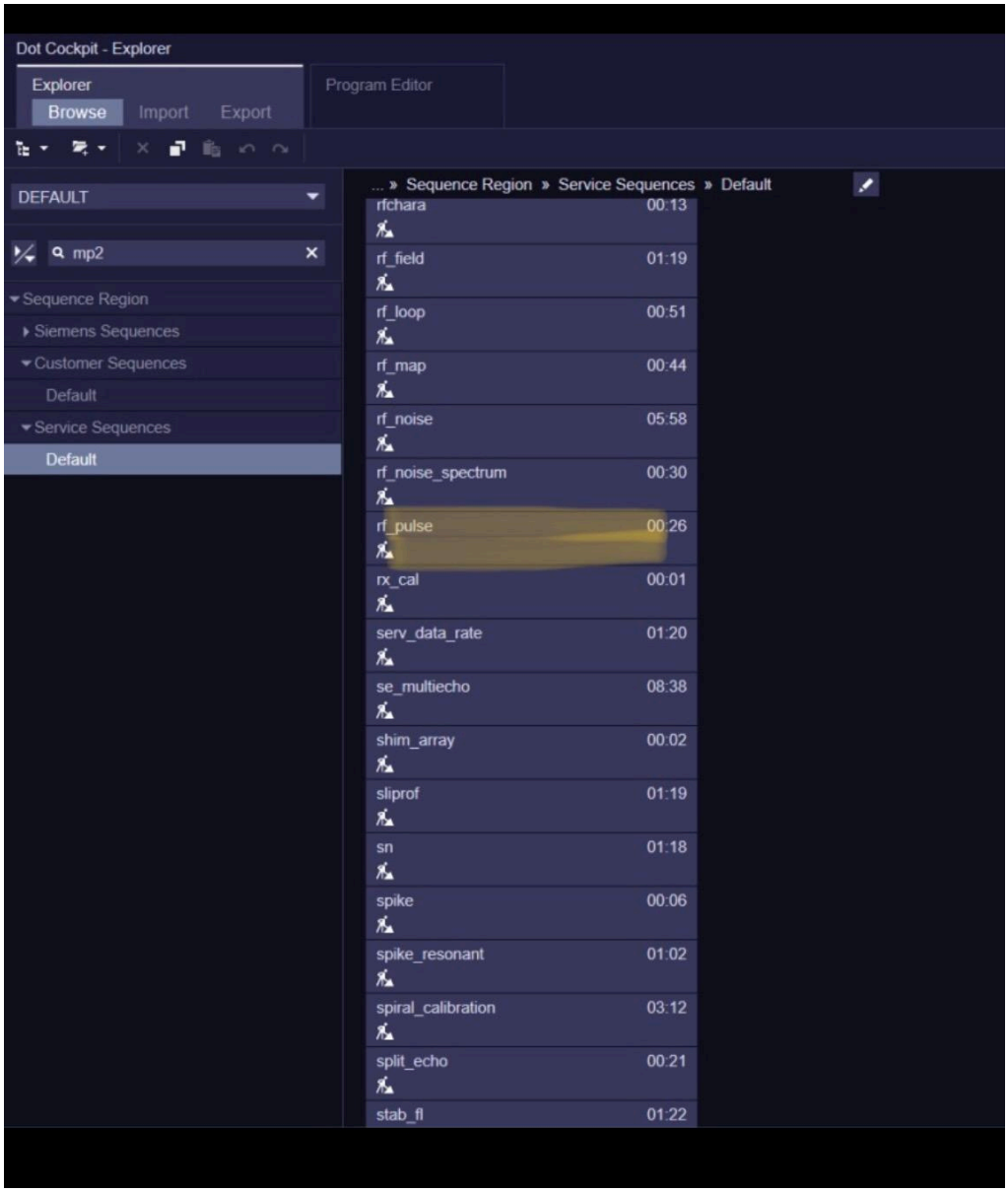
3. Method

'Standard' phantom data will be collected over 18 months, at least once a week, with a schedule specific to each site. For instance 1 time per week on 'random' days. In Copenhagen, the planned schedule is once a week, plus a few times during the 18-month period we will collect data daily to check for higher frequency fluctuations.

Each site uses scanning parameters that approximate/average what is typically done by fMRI studies at their centre. Data collection is standardized within each centre such as (1) the same phantom is used (2) positioning is always the same, using a phantom holder, further avoiding vibration induced motion - an example is given in Vogelbacher et al. (2019) - see figure 7. It is recommended to use either a fixed or reference mode rather than iso, limiting changes in shimming and hence any additional fluctuations. Each center should localise to the center on the phantom.

When 2 or more fMRI runs are collected, order should follow research practice (always 'resting state' fast sequences 1st for instance if that's what people do). Both schedule and fMRI run ordering can be investigated on their own, and regressed at the mega-analysis level. This is important as for long QA sequences, there is a 'natural' drift because the body coil heats up and therefore it is recommended to keep the scan order consistent within site for image quality purposes.

Before measuring, it is recommended to run an RF noise scan: an amplitude scan running at a range of frequencies which outputs a graphical representation allowing visualization of random noise and spikes. This would therefore remove any doubt in the environment if we witnessed any fluctuations in SNR in the scan protocol.



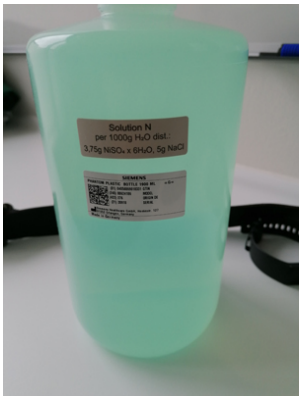
RF scan default for Siemens

3.A. Sites and protocols

Indicate below (1) your site (2) scanner, coil and phantom use (3) names and emails (4) protocol details in tables 1 and 2

Sites

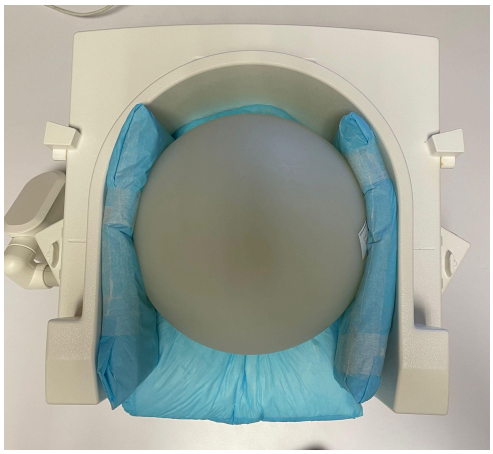
Neurobiology Research Unit, Copenhagen - Siemens Prisma (syngo MR VE11E) with 32-channel receiver head coil - Siemens 'bullet' phantom. Dr Cyril Pernet (cyril.pernet@nru.dk), Dr Patrick Fisher (patrick.fisher@nru.dk), Prof. Gitte Knudsen (gitte.knudsen@nru.dk) + a number of PhD students TBA.



Siemens bullet phantom

Centre for Translational MR Research (TMR), Yong Loo Lin School of Medicine, National University of Singapore.
Prof Michael Chee (michael.chee@nus.edu.sg), Dr. Helen Zhou (helen.zhou@nus.edu.sg), Stevia Ng Gogna (stevia_ng@nus.edu.sg), Constance Lin (cons.lin@nus.edu.sg), Annalissa Tiu Munoz (anna.munoz@nus.edu.sg), Koh Li Yang (kohly@nus.edu.sg), Dr. Soon Chun Siong (cs.soon@nus.edu.sg), Dr. Cisy Liu (cisy.liu@nus.edu.sg).

Siemens PRISMA FIT and PRISMA scanner (syngo MR VE11E) with 32-ch head coil - Siemens spherical water phantom.



Spherical phantom position

Hardware info

Site	Scanner Vendor	Model	Software version	Coil	
Neurobiology Research Unit,	Siemens	Prisma	Syngo MR VE11E	32-channel receiver head	

Copenhagen					
Centre for Translational MR Research (TMR), Singapore	Siemens	Prisma FIT	Syngo MR VE11E	32-channel receiver head	

Protocol

(1) Rf noise scan
(2) MPRAGE
(3) fMRI (at least run 1, up to 3)
(4) FieldMap
(5) T2
(6) MP2RAGE (optional)
(7) DWI (optional)
(8) SVS (optional)
-- manual record of temperature and humidity?

	T1	T2	fMRI run 1	fMRI run 2	fMRI run 3
Neurobiology Research Unit, Copenhagen	MPRAGE (IR) TE: 0.00258, TR: 2, IT: 0.972, FA: 8, 0.9mm^3 Done 1st	Spin Echo TE: 0.408, TR: 3.2, FA: 120, 0.9mm^3 Done 5th	Gradient Echo TE: 0.014,0.03,0.052 TR: 2, FA: TBC, MB: 2 3*3*3mm 100 volumes (3min33) Done 2nd	Gradient Echo TE: 0.03, TR: 2, FA: 70, 3*3*3mm 100 volumes (3min33) Done 3rd	Gradient Echo TE: 0.037, TR: 0,8, FA: 52, MB: 8 2m^3 250 volumes (3min33) Done 4th
Centre for Translational MR Research (TMR), Singapore	MPRAGE (IR) TE:2.45 TR:2200 IT:900 FA:8 1.0mm isovoxel	Spcir TE:393 TR:7000 IT:2100 FA:120 1.0mm isovoxel	epi TE:12,29.75,47.5 TR:1000 FA:50 MB:4 3.0mm isovoxel	epi TE:30 TR:719 FA:50 MB:6 2.5mm isovoxel	

Table 1: anatomical and functional phantom data collected across sites.
TE echo time, TR repetition time, IT inversion time, FA flip angle, MB multiband acceleration factor

	FieldMaps	MP2RAGE	DWI	SVS	
Neurobiology Research Unit, Copenhagen			32 directions QA done on Siemens FA, ADC, Trace maps		
Centre for Translational MR Research (TMR), Singapore	We commonly use the inverted phase encoding direction of each fmri run with one measurement				

Table 2: other acquisitions - data are shared but not used for these planned analyzes (could be used later for other exploratory purposes)

3.B. Data analysis

Note, analyses can be carried out locally and results aggregated - likey running a container to minimize hardware/software differences.

Phantom QA metrics

For T1, T2, FA, ADC, Trace images:

$$SNR = \frac{mean\ of\ the\ 50\%\ most\ central\ voxels\ of\ the\ phantom\ (\sim=most\ intense)}{mean\ of\ the\ 50\%\ noisiest\ background\ voxels\ (i.e.\ those\ showing\ the\ highest\ std)}$$

For fMRI: static spatial noise (SSN - Friedman et al 2006) and temporal SNR (based on Liu 2016 and Wald &Polimeni 2017)

$$SSN = mean(sum\ of\ odd\ images - sum\ of\ even\ images)\ for\ the\ 50\%\ most\ central\ voxels$$

$$tSNR.phantom = \frac{mean(mean\ in\ time\ of\ the\ 50\%\ most\ central\ voxels\ of\ the\ phantom)}{mean(std\ in\ time\ of\ the\ 50\%\ most\ central\ voxels\ of\ the\ phantom)}$$

$$tSNR.zero = \frac{mean(mean\ in\ time\ of\ the\ 50\%\ most\ central\ voxels\ of\ the\ phantom)}{mean(std\ in\ time\ of\ the\ 50\%\ noisiest\ background\ voxels)}$$

$$tSNR.signal2termal = \sqrt{(tSNR.phantom/tSNR0)^2 - 1}$$

Phantom QA and group level analyses of grey matter

For a set of T1 (and T2 for multispectral segmentation) compute the usual voxel-wise volumetry (SPM12 - Dartel) and thickness/surface area (Freesurfer 7 recon-all with DKT atlas), i.e. we have 6 measurements (unimodal/multimodal segmentation * volume/tickness/surface area) and regress out age, sex, total gray matter volume, SNR and redo the analysis without SNR (see Pernet 2018 for choice of covariates). The obvious contrast of interest is on SNR but it might not

show a lot. More interestingly, we can look at the usual effect of age and sex on ROI (hippocampus, amygdata and basal ganglia for age (Dima et al., 2021) and anterior cingulate, hippocampus, amygdala, fusiform gyri for sex Liu 2020) and investigate how the SNR covariate influence the results.

Super-additive or interaction effect with in-scan QC: QC metrics from T1/T2 scans can also be computed and see itself how this related to the phantom data - and how morphometric results are affected by those relative to the phantom's data.

Phantom QA and group level analyses of BOLD fMRI

Basic Analysis: Because each centre runs different studies, parallel to the QC protocols, basic analyses must be carried out locally (also avoid any data transfer issues). Taking data from one or more studies, in collaboration with the study(ies) PI(s), extract ROI values for an fMRI effect - being differences on a task activation, difference in connectivity, differences between groups. The more diverse 'effects' we have the better. From these ROI, replicate the observed effect and regress out QA to investigate how the tSNR covariates influence the results. A meta-analysis of the amount of change will be performed. In addition, we can try a mega-analysis since we will focus on ROI only (i.e. should not be an issue for everyone to share a csv file of values) with sites, studies and observed effects (pre- post- QA covariation).

Super-additive or interaction effect with in-scan QC: QC metrics for each time series scan (dropout, ghosting, motion, temporal standard deviation) can also be computed and see itself how this related to the phantom data - and if regressed together, it has more or less effect on results.

Older discussion moved here for record keeping

Great paper tool+review from Christoph
https://www.frontiersin.org/articles/10.3389/fnins.2019.00688/full#h3

Each site has different objects to scan as phantom, and may have different objectives - however a common core can be found. Assuming we still want to show that QA can be used to regress out stuff on studies, parameters should reflect what studies do locally - caring about reproducing/replicating effects rather than having the same set of acquisitions (which would be more useful and scanner QA itself)

Initial Draft Protocol

- Types of phantoms
    - ADNI Phantom?
    - Siemens Bullet?

Data acquired on 15 September 2021 - with the intention that this reflects what people do with their data, therefore variations on the phantom could influence results (ie no need for us to have the same TR, TE, FOV, etc, across sites - IMO we want to replicate that the phantom QA influences human results not that the variations values are similar across sites) → access our nifti here https://drive.google.com/file/d/17h8JCebPCKSIyESv3fwiChGYLr_NnJ1N/view?usp=sharing NOTE - not sure we will use that, looking at the data this is not very homogenous

- Positioning
    - Fixed
        - How?
    - How to identify orientation correctly and unambiguously?
        - Per scanner/model instructions on how to copy/transfer (ideally not "enter") positioning information from prior scan. (This was Yarik before login)
    - How long should we let it sit before scanning, to deal with bubbles etc?
- Sequences included and their parameters
  → expected total duration? For copenhagen, we would be at ~1/2h max (also we have two scanners, Prisma and Biograph so that's ~1h in total - likely including set-up)
    - T1w
        - Parameters?
            - Try using parameters that match with the local users projects as much as possible?
            - Try use as many advanced settings as possible
    - T2w
        - parameters?
    - T2*EPI
        - 2 runs minimum
        - Other parameters/requirements?
          In Copenhagen, 2 runs with different parameters reflecting task ('slow') vs resting state (SMS)
    - DWI?
        - Parameters?

- Schedule
    - Fixed (e.g. every tuesday) or
    - Pseudo-random (different days in different weeks) or
    - Both (tuesday and another day)?
    - What time of day?
        - Before operations start (cold scanner)
        - Middle of the day ("warmed up" scanner)
        - End of day (after full use)
        - Or randomized?

Copenhagen - pseudorandom,1 time a week + maybe at random a few weeks with daily sampling in case we have some higher frequency scanner changes

- Stability
    - Procedures and methods to identify if the phantom itself is stable or not?
- Formats and data management
    - Scripts to make them as much BIDS compatible as possible?
    - We can easily retain DICOMS on NITRC
    - These outputs should work (or easily transformable) as inputs to fBIRN and other QA pipelines
-