QIBA ASL Profile: TF2 Meeting
15/09/2021

Agenda

- Introduction
  - Aims of taskforce:
    - Establishing QA procedures to be performed to qualify that a site can achieve the baseline performance level for ASL data as described in the profile
    - Phantom experiments to validate the assumption of linearity in an ASL measurement.
  - Current QIBA profile claim (https://docs.google.com/document/d/1rgaUqTLw1h3dYXGemkNAj3XADuPtB430oomZAf-_Muo/edit)
    - A measured change in CBF using ASL of 21% or larger, in a grey matter region with a minimum size of 1mL, indicates that a true change in CBF has occurred in that region with 95% confidence.
- Focus of today's meeting is site qualification:
  - What are sites currently doing for QA of ASL scans?
  - What are the important factors to assess and the baseline thresholds - SNR, tSNR, readout distortions, blurring, ghosting/artefacts? Or focus on the hardware, e.g. gradients, RF, coils etc.
  - Also consider the human factors.
  - Is a perfusion phantom necessary for this?
    - Advantages - can make the process very simple as results can be easily compared against thresholds.
    - Disadvantages - availability of phantoms, expense.
  - Should anything else be assessed in addition, for example a T1w structural scan for registration?
- Get input from as many people about what they currently do
- Discussion about what is important for site accreditation
- Establish people who are interested in doing the work to take this forward

Meeting Outcome

Summary about what QA is currently being done

In general human phantoms are used for ASL Quality Assurance. ASL is inherently challenging and so if images with sensible perfusion are acquired that is a good enough test. There are ways that this assessment can be made quantitative and objective - a Quality Evaluation Index can be used, for example the GM/WM ratio should be approximately 3:1.

It does however require the hardware to be well set up. This can be assessed using a static phantom - fBIRN phantoms (spherical gel phantom) + test are often used for this.

Is a perfusion phantom necessary?

At present there is no evidence to support that a perfusion phantom would do better than a human phantom for these purposes. Furthermore perfusion phantoms do not fully reproduce human physiology or anatomy, so scanning a volunteer will likely be required anyway.
For a cross-sectional claim however, a perfusion phantom will likely be necessary to establish bias.

Image analysis

A digital reference object (DRO) will be necessary for testing that analysis software is performing correctly. One does exist for ASL: https://github.com/gold-standard-phantoms/asldro

Target strategy for site accreditation

Based on the outcome of this meeting, this is the approach we will adopt within the taskforce.

- Standardised static phantom test to assess the MRI hardware - fBIRN test proposed.
- Standardised human volunteer procedure:
    - Specify MRA for label plane planning
    - Specified ASL protocol for each vendor and implementation (Siemens will soon be releasing their product sequence for platforms other than VIDA, meaning all 3 vendor product sequences will be pCASL with 3D readout will be standard across current generation MR systems).
    - Formalised analysis of visual quality. Ideally this should be automated, but a manual procedure should be described.

The next steps are:

1. Groundwork to establish the minimum level of performance to support the claims (either experimental or literature):
    a. Hardware - fBIRN test + other measurements on the same static phantom?
    b. Human phantom - what are the metrics and the thresholds for baseline performance.
    c. Linearity across CBF values important
2. Define the human phantom procedure
    a. Subject requirements - age, sex, health, other factors important (e.g. coffee) - do any need to be standardised/restricted?
    b. Protocol
        i. MRA
        ii. ASL
        iii. T1w Anatomical
        iv. Any other scans required? Field maps? PCA at label plane?
        v. Any requirements for how long it should take - 30 minutes?
    c. Define the analysis
        i. Preprocessing
        ii. Calculation of the QA metrics
3. Validate that the procedure/accreditation test works
    a. What is the gold standard to compare against?

Inline Meeting Notes

Item	Who	Notes
QA Experience	John Detre	Have always used human phantoms. Large studies have used a travelling human phantom. Looking for the appearance of the perfusion maps - do they look as expected. Checking for appearance of GM/WM ratio (=3/1). More recently, development of a Quality Evaluation Index (QA metric) for comparison in many sites: from visual inspection to number. Also develop phantom, but more for sequence development (optional for QA ASL in the field). No affordable easy solution. Main issue with phantoms: not really reproducing physiology and human anatomy.
	Karl Lovblad	Also use of humans as ‘phantom’. Scan - rescan as control, mainly in volunteers. Difficult to differentiate between technical and clinical issue.
	Xavier Golay	Multiple studies have also used regular scanning of volunteers, and in addition to establish some visual QA for assessing acquired study data. QUASAR study (n=30 sites), certain sites had very poor reproducibility, and these sites did not

		pass the standard QA of the manufacturer. Hard to separate effects.
	Anthony Liu	Use of ASL for brain tumours. No phantom available now. Main question: post-op: recurrence vs. treatment effect. Comparison between ASL, DCE * DSC (complementary to each other). Use of DSC as QA to ASL!
	Patricia Clement	Have a phantom (QASPER). Problem of quantification (air bubbles). Setting up a QA protocol, but not easy. It could be good as part of a QA (including volunteer scanning). Collaboration with Henk Jan Mutsaerts also on visual QC.
Outcome	Aaron OT	Is the use of a 'human phantom' something that can be prescribed by the QIBA profile? Does it need specific expertise, or is it too dependent on the local expertise for it to be used? What about the use of a simpler phantom? ACR test-like for ASL?
	John D	No other phantom measures the same as what is in the human. Generally, HW needs to work well, and this can be assessed using a simpler phantom. The main challenge though is that implementations are still relatively varying from manufacturer to manufacturer. So, a metric would be difficult to specify.
	Aaron OT	What are the tests used?
	John D	A simple fBIRN test. Some places run a QA daily.
	Anthony Liu	ACR test once a year. Picks up trends only. Generally, issues are picked up by patient scanning.
	Michael Boss	Generally, phantoms discussed (ACR, fBIRN) are static. Only QASPER is dynamic. Fix pitfalls? Use proxy measurements? Physical phantoms are only one part of the equation. Analysis needs to be assessed as well. Also use of DROs.
	Aaron OT	GSP established a DRO, which is distributed for free (Open Source) as part of OSIPI. When we used it in a 'round robin' trial, GSP personnel used SOPs. Test of HW + volunteer: can this be sufficient?
	John D	ASL is tough, you guys! It is a good test in itself. It is not a bad test to be able to get a perfusion image out of a volunteer. Compare first vs. second half of time series. Formalisation of a visual quality test might be the better way to go. It does not have to depend on the actual CBF value. Important is the reproducibility.
	Anthony L	Provide detailed instruction on how to plan and scan a 'human phantom'. E.g. MRA at the labelling plane to optimise.
	Aaron OT	Define a basic test (e.g. fBIRN?) + human phantom with detailed protocol and image processing.
	Anthony L	Siemens: 3D-pCASL soon available on more than VIDA scanner.
	Aaron OT	How can we demonstrate that this is sufficient to fulfill the QIBA claims? Should we try to stress the system (i.e. change shims)?
	Michael Boss	Need to do some groundwork to establish the minimum level of performance to support the clinical claim, in the absence of supporting literature. At present we have best practises, there may need to have something that is more concrete and quantitative to support.
	John	A physiological manipulation could be used to support the physiological change,

	Detre	which is what we are looking at in the claim - finger tapping exercise for example.
	Xavier Golay	There exists lots of test-retest studies published. Was this sort of groundwork done for diffusion?
	Michael Boss	For diffusion there was some preliminary groundwork. The immediate need was a profile with a longitudinal claim (for a set of organs). In the future there would be a cross-sectional claim, which requires establishing bias. The ice-water phantom at 0°C needs to be measured to within 4% to demonstrate. In the future this will move to the range of ADC's. A round robin was performed, and this had a high CoV, and adding an additional b-value helped to reduce this.
	Xavier Golay	So only a phantom measurement is required? Can you prove/demonstrate that if you obtain within 4% of the phantom value then you will get good results in a patient?
	Michael Boss	The present QA procedures are a sanity check - are the gradients working correctly, is the scanner well set up. Ensuring that things aren't way-off by 50%, and showing that you are likely to have good reproducibility.
	Anthony Liu	Static phantom and expect no signal.
	Michael Boss	The QA procedures in the diffusion profile involve multiple measurements and this is something to consider.