PREDICTING THE QUANTITATIVE ACCURACY OF IN-VIVO ADC USING AN ICE-WATER PHANTOM

David M. Morris[1], Hossein Ragheb[1], Neil Thacker[1], Naomi H.M. Douglas[2], and Alan Jackson[3]

[1]Centre for Imaging Sciences, University of Manchester, Manchester, United Kingdom, [2]Institute of Cancer Research, Sutton, United Kingdom, [3]Wolfson Molecular Imaging Centre, University of Manchester, Manchester, United Kingdom

Target audience – This work is relevant to investigators using ice-water phantoms to validate Apparent Diffusion Coefficient (ADC) measurements in patients. Multi-center trials using diffusion biomarkers to monitor tumors in the abdomen will benefit from these findings.

Purpose – The QuIC-ConCePT project is a multi-center European study to develop methodologies for the measurement of clinical biomarkers. ADC has shown promise for the monitoring of tumors in response to treatment or for oncological decision support. To do this the uncertainties identified previously[1] as restricting these studies need to be quantified. Key to this is the development of calibration and QA methods which allow for the assessment and comparison of scanner performance and the possibility of combining the data across different centers in large trials. A single acquisition phantom will be required across all platforms; due to the temperature dependency of diffusion an ice-water phantom[2] is required. As part of the preliminary work on the project, measures were taken using an ice-water phantom developed for the project[3]. The structural scan in Figure 1a shows the tubes with different ADCs and some ice at the top of the image. A simple process for the assessment of machine performance was compared with measurements in the livers of normal volunteers to establish the relationship between phantom measurements and the quality of the clinical data acquired *in-vivo* in a clinical study.

Methods – Data was provided by centers participating in the development of scanning protocols. This represented four centers (A, B, C and E) in different countries with 1.5 T scanners representing Siemens, GE and Philips. Sequences were designed with a common specification of EPI readout, matrix, field of view, parallel imaging, bandwidth, Partial Fourier, shortest TE, minimum TR and diffusion weighting (b values)[4]. The differences in individual scanner implementations are mitigated by optimisation by local experts as would occur in a clinical trial. For each site three multi slice phantom measurements (Figure 1) were acquired on separate days and five normal volunteers had their livers scanned twice at each site. Data were analysed by a single center fitting mono-exponential signal decay to the b values (100, 500 and 900 s mm^{-2}) including a first order correction for Rician noise. This was on a pixel by pixel basis for the region of interest; mean values were returned following artifact rejection. ROIs were defined in the cylinders by pattern recognition while in the volunteer scans two representative regions of interest (ROI1 and ROI2) were defined by a single expert observer. Data analysis estimated several variables intended to characterise important aspects of performance. The routine assessment of spatial signal to noise was performed over the phantom compensating for the inhomogeneities introduced by the cylinders and variable ice densities. Also goodness of fit, representing the quality of the ADC obtained, and regional inhomogeneity were evaluated. The absolute ADC measurements on the different scanners had systematic offsets which affected these variables so they were nornalised out. These were compared with standard deviation of the tube ADC values to establish the phantom parameters indicative of the uncertainty observed and these were compared with regional values from the volunteer ROIs. A variability of the ice packing densities was observed and this was included in the analysis as a possible confound.

Results – The adjustment for absolute scaling for the standard deviation in the phantom improved the correlation with measures of image quality across the scanners. The goodness of fit and the standard deviation of measured ADC values to a mean scaled prediction are found to be the most reliable forms of assessment as shown in Figure 2a, representing the best agreement of measured parameters of the uncertainty, the dotted line being the best fit line. The assessments of inhomogeneity and noise were disrupted by the presence of the ice while it was observed that less ice resulted in the ADC scaling being affected. If ice packing was above 40% there was no effect on the reproducibility parameters. There was no relationship observed between any phantom values assessed and the variability *in-vivo* as seen in Figure 2b; the variation in ADC being indistinguishable between scanners. Reported values are Chi squared values for phantom goodness of fit (Chi values normalised by the mean signal and mean standard deviation of image noise) and the standard deviation of the mean ADC for the variability assessments which are then compared across sites and visits.
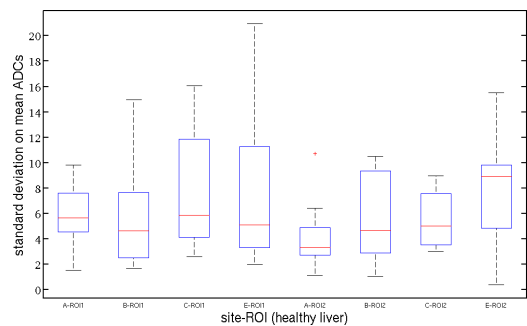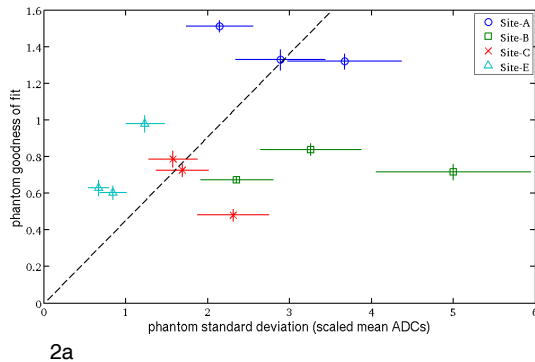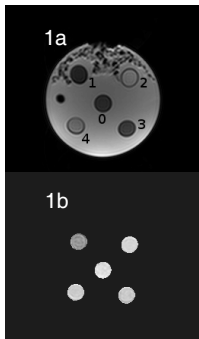

Figure 1a: Structural scan of the phantom
Figure 1b: ADC maps

Figure 2a: phantom goodness of fit versus the corresponding standard deviation on scaled ADC.
Figure 2b: the spread of standard deviation % on ADC from normal liver for each site A, B, C and E for ROI1 and ROI2.

Discussion – The current phantom design can be used to obtain a degree of confidence in the measurement of ADC on a given scanner. However, several processes modify the way data are acquired when imaging a phantom rather than a human subject. In-vivo measurement accuracy is affected by biological variation, imperfect fat suppression and motion resulting in order of magnitude differences between the repeatability of ADC measurements in different individuals. Even fundamental parameters such as signal-to-noise, corrected for any system bias, are not directly comparable with the performance in-vivo. If we wish to be able to use an ice-water phantom to assess the likely quality of clinical measurements we will have to understand the consequences of these differences. The current phantom design needs to be modified to better represent in-vivo measurement and to allow for the improved assessment of uniformity and spatial noise. At the same time the methodology for ROI measurements in-vivo needs modification to improve repeatability. We are currently investigating alternative analysis strategies and phantom modifications in order to address these issues. Further investigation on the effects of different ice packing on thermal equilibrium on absolute ADC measurements is required.

Conclusion – While considerable differences are seen in the estimated parameters between vendors (consistent with the size of effects reported[5]), we see no correlation of any phantom parameters with in-vivo reproducibility. While we can differentiate scanner performance this does not indicate that the data will be of the quality required for a particular study. Future work on accurate measurement of inhomogeneity might explain the missing correlation.

References – [1] Padhani A et al Neoplasia. 11(2): 102–125. 2009. [2] Chenevert T et al JMRI 34: 983–987. 2011. [3] Collins D et al QA Phantom for DW-MRI 23/02/2012. [4] Douglas N et al IMI WP3 Task 2.3 Reproducibility Study 11/05/2012. [5] Malyarenko D et al JMRI 37:1238–1246. 2013.