Errors Associated with Followup Measurments of ADC in Assessing Response to Neoadjuvant Chemotherapy

Shelley Waugh[1], Lukasz Priba[1], and Sarah Vinnicombe[2]

[1]*Medical Physics, Ninewells Hospital, Dundee, Angus, United Kingdom,* [2]*Division of Cancer Research, University of Dundee, Dundee, United Kingdom*

Introduction and Aims: Breast Magnetic Resonance Imaging (MRI) is becoming more routinely used for lesion characterisation and assessment of response to neoadjuvant chemotherapy in breast cancer patients. Morphological changes in lesion size and shape as well as changes in the contrast uptake dynamics can indicate response, without exposing the patient to ionising radiation. However, early changes in diffusion characteristics of the water molecules within a tumour, as assessed by measuring the apparent diffusion coefficient (ADC) have also been linked to treatment outcome. Diffusion weighted imaging (DWI) provides a measure of the diffusion of water molecules, which is restricted in highly cellular lesions. An increase in the ADC after initiation of treatment suggests a decrease in the cellularity of the tumour, which may indicate a response to treatment. There is little data available on reproducibility of such ADC measurements in a clinical environment and therefore it is essential to characterise the likely influences on such follow-up measurements in order to ensure that reported changes in ADC values can be attributed to response, rather than other external factors.

The aim of this work was to investigate the effect of scanner stability, scan-scan variability in ADC measurements and observer reproducibility on the changes in ADC likely to be encountered in clinical patients undergoing neoadjuvant chemotherapy.

Methods To assess scanner stability, an ice-water phantom [1] was used and scanned weekly on each side of the breast coil using the standard clinical imaging sequence over the course of six weeks. ADC values were recorded and stability was assessed using the coefficient of variance (CoV) across all measurements.

Scan-scan reproducibility was assessed in 10 healthy female volunteers (age range: 26-61; mean: 44.7 years) who were scanned twice on a 3.0T MRI scanner (Siemens Trio; Erlangen, Germany) using the standard DWI sequence (b=50, 800 s/mm^2, voxel size: 1.8×1.8×4 mm^2). Volunteers were scanned with a 4-week time interval between examinations to ensure each volunteer was scanned in the same phase of their menstrual cycle, since this can influence ADC measures [2]. Data was anonymised and randomised and analysed by placing a region of interest (ROI) of around 40 pixels within healthy fibroglandular tissue on the ADC map and measuring ADC values. Repeatability between examinations was then assessed using coefficient of reproducibility (CoR, units mm^2/s) and intraclass correlation coefficients (ICC).

For the patient study, a total of 41 patients with biopsy-proven cancer scheduled for neoadjuvant chemotherapy were included. Imaging was performed on a 1.5T Siemens Avanto scanner or a 3.0T Siemens Trio scanner using a dedicated bilateral breast coil. ADC maps were produced from DWI with b-values of 50 and 800 s/mm^2 (in-plane resolution: 1.8×1.8 mm^2; slice thickness: 4 mm). One observer performed baseline and interim analysis after 3-cycles of chemotherapy. Measurements of ADC were made by drawing around lesions on the slice with maximum restriction of diffusion dimension. Whole tumour ADC (ADC_{WT}) and worst tumour (ADC_{min}, using 3mm^2 pixel ROI) were measured. Baseline measurements were repeated again a week later to assess intra-observer variability. Reproducibility was assessed using CoR and ICC. The change in ADC was calculated for both ADC_{WT} and ADC_{min} and patients categorised as responders or non-responders according to RECIST criteria [3] according to their final end of treatment response MR examination.

Results: Scanner stability, as assessed using the phantom measurements, demonstrated excellent stability within the breast coil with an average ADC of 1.089×10^{-3} mm^2/s and a CoV of 6.6% (0.07×10^{-3} mm^2/s).

Analysis of the healthy volunteer data demonstrated that there was a slight difference in ADC measures between the left and right breast, with average values of $ADC_{LEFT}=1.499\times10^{-3}$ mm^2/s and $ADC_{RIGHT}=1.473\times10^{-3}$ mm^2/s, however these were not found to be significantly different across all volunteers at both time-points (p=0.887; students t-test). When comparing the baseline and follow-up examinations, a CoR= 0.122×10^{-3} mm^2/s was calculated, with average ADC values of $ADC_{baseline}=1.503\times10^{-3}$ mm^2/s and $ADC_{follow\text{-}up}=1.544\times10^{-3}$ mm^2/s leading to an overall estimation of CoR of 8.0%. Intra class correlation coefficients were found to be good, at 0.811.

Within our patient study, the intra-observer repeatability was found to be excellent with a CoR $ADC_{WT}=0.113\times10^{-3}$ mm^2/s and CoR $ADC_{min}=0.131\times10^{-3}$ mm^2/s (ICC=0.983 and 0.976 respectively). This corresponds to an error on the average ADC of around 11% for ADC_{WT} and 16% for ADC_{min}.

Within our patient population, complete data was only available for 37 patients. Of these, 6 were deemed to have had a complete response (CR), 25 had a partial response (PR) and 6 were categorised as stable disease (SD), by standard RECIST criteria. The average change in ADC value for each group is shown in Table 1 for both whole tumour and worst case tumour values. The percentage change in ADC was calculated relative to the baseline value and results for this parameter are also presented in Table 1. There is clearly a difference in the change in ADC between each category, although these were not found to be significant in this study (p=0.1; Kruskal-Wallis test), which may be due to the small sample size.

	Average Absolute ADC Change (mm^2/s)		Average % ADC Change	
	ADC_{WT}	ADC_{min}	ADC_{WT}	ADC_{min}
CR	0.176	0.202	19.7 %	26.9 %
PR	0.109	0.099	11.1 %	14.0 %
SD	0.045	0.102	6.1 %	11.4 %

Table 1*- Summary of change in ADC values for partial, complete and non-responders to neoadjuvant chemotherapy treatment*

Comparing the ADC changes with our intra-observer repeatability, it can be demonstrated that CR should be readily identifiable when only one reader reports ADC images. Partial response lies just at the boundary of repeatability and therefore using ADC alone to classify reproducibility may be unreliable in some cases. Stable disease cannot be identified using ADC images alone and would need to be characterised in conjunction with other images from the examination.

Discussion: Our phantom measurements were found to be within 1% of the ADC reported for water at 0°C (1.1×10^{-3} mm^2/s). Our repeatability measures suggest that scanner stability has minimal effect on the repeatability of measures of ADC. With just a contribution of 6%, this is found to be in line with the measured scan-scan variation as assessed within our healthy volunteer population.

The coefficient of repeatability for scan-scan variations within the volunteer population was measured to be on the order of 10%. This measure is likely to reflect uncertainty when placing an ROI within a region of healthy fibroglandular tissue, and therefore slight positional differences between analyses of each time-point scan. This is less likely to occur in symptomatic patients where regions of restricted diffusion are more straightforward to identify and analyse between multiple examinations. Absolute ADC values measured were variable, but this is likely due to the age range and breast characteristics of the volunteer population. All women were imaged in the same phase of their menstrual cycle at baseline and follow-up in order to negate any cyclic effects that have been previously noted [2].

Our patient results demonstrate that CR to neoadjuvant chemotherapy can be determined using ADC images alone. When response is categorised according to the RECIST criteria, the change in absolute ADC value for CR results in changes greater that those attributable to intra-observer repeatability or scan-scan repeatability and scanner stability. The change in ADC values for those patients who have a PR to treatment are of an order of magnitude similar to intra-observer repeatability, however well within the changes attributed to scanner stability and scan-scan stability. However, it should be noted that this group is the majority of cases and therefore very diverse. It is possible that by further sub-categorising the response categories that the changes could be dichotomised.

Stable disease results in very little changes to the ADC values, as would be expected. These measured changes are likely to be indiscernible from natural error due to intra-observer and scan-to-scan repeatability. Should it prove possible to further categorise the PR into more robust categories, it is possible that SD could be classified by using these small percentage changes to identify such disease. Due to the better repeatability in measurement, however, it is recommended that ADC_{WT} is used wherever possible. The results from the patients demonstrate a good reproducibility for measurements of large breast cancers in patients scheduled to undergo neoadjuvant chemotherapy, with repeatability being slightly improved for whole tumour measurements compared with worst case values. The slightly poorer reproducibility for ADC_{min} using the small ROI is likely to be attributable to tumour heterogeneity, perception and artefacts.

Conclusions: This study has demonstrated various considerations that should be made of different factors affecting repeatability of ADC measurements in patients undergoing neoadjuvant chemotherapy. We conclude that ADC measures of whole tumour are best for identification of response, and that complete response by RECIST criteria can be identified using ADC images alone. Due to the diverse group of partial response patients, it is more difficult to apply a specific cut-off due to the similarity of changes in ADC measures with those measured for intra-observer repeatability.

References:

[1] Malyarenko *et al.* JMRI (2013) ; 37(5) :1238-1246 ; [2] Partridge *et al.* JMRI (2001) ; 14 : 433-438 ; [3] Eisenhauer *et al.* EJC (2009) ; 45 : 228-247