The effects of region of interest errors on estimates of whole-body tumour volume and ADC in patients with bone metastases.

Matthew D Blackledge[1], Nina Tunariu[1], David J Collins[1], Anwar R Padhani[2], Martin O Leach[1], and Dow Mu Koh[1]
[1]*CRUK and EPSRC Cancer Imaging Centre, Institute of Cancer Research, Sutton, Surrey, United Kingdom,* [2]*Paul Strickland Scanner Centre, Mount Vernon Hospital, Northwood, Middlesex, United Kingdom*

Introduction: Whole body DWI (WBDWI) has shown high sensitivity for detecting metastases in the bone marrow and soft tissues, and high diagnostic accuracy for disease staging. Due to high disease/background-tissue contrast, definition of Regions of Interest (ROIs) is possible throughout the body, a feat that has been accelerated using semi-automatic segmentation techniques [1]. From the defined regions, it is possible to derive two quantitative imaging biomarkers: (i) *total diffusion volume*, or tDV (in milliliters), and (ii) *global apparent diffusion coefficient* (gADC) (in mm^2/s) [1]. Such metrics may provide valuable alternatives for evaluating treatment response in patients with metastatic bone disease, where the use of standard morphological imaging is unreliable. However, current practice requires manual correction and/or removal of ROIs, leading to errors and bias. A recent report has investigated the level of agreement between volumes defined by two independent radiologists that were blinded to clinical outcomes of patients diagnosed with bone metastases from a range of primary malignancies [2]. Although this report concluded that good inter-observer agreement could be achieved for tDV and median gADC estimates, the authors did not investigate the *intra*-observer variability, the knowledge of which is vital to understand the repeatability in derived metrics from WBDWI.

Purpose: To derive the intra-observer repeatability of two radiologists, O1 and O2, reporting median gADC and tDV estimates derived from WBDWI imaging in a cohort of nine patients with bone metastases.

Methodology: *Study population:* Seven patients with metastatic prostate (N = 3) and breast (N = 4) cancers, one patient with malignant melanoma and one patient with multiple myeloma. *MR image acquisition:* Axial, fat-suppressed, free-breathing EPI WBDWI performed on a 1.5T MR system (Siemens Avanto, Erlangen, Germany) from skull base to mid-thigh in each patient (TR = 14000ms, TE = 67ms, matrix size 128x104x150, 430 cm field of view, slice thickness 5mm, receiver bandwidth 1628 Hz/pixel, STIR fat suppression, 6 signal averages, b-values 50 and 800 s/mm^2). ADC maps for all stations were generated covering each voxel location using a mono-exponential model for the two b-value data. *Image analysis:* Two independent radiologists, with eight and four years experience in body DWI respectively, reviewed the WBDWI data alongside T1-weighted morphological MR-images for anatomical reference. A pre-processing step was applied to all data to correct for discontinuities in WBDWI imaging stations [3], followed by semi-automatic segmentation to provide preliminary definition of ROIs [1]. Resultant regions were manually corrected using the following exclusion criteria: regions of necrosis with ADC > 2.0 mm^2/s (T2 shine-through), regions that included incomplete fat suppression and any regions above the C4 vertebrae. Analysis was repeated twice by each radiologist, ensuring at least a six-month interval between each review, resulting in two volumes if interest: V_{i1} and V_{i2} for each radiologist *i*. Bland-Altman analysis [4] was used to report the level of repeatability for tDV and median gADC estimates derived within V_{i1} and V_{i2}. Furthermore, the difference between regions V_{i1} and V_{i2} was quantified by the *Percentage of Agreement* (*PoA*) score, defined as the area of the intersection of the two regions as a percentage of the area of their union:

$$PoA = \frac{A\left(V_{i1} \cap V_{i2}\right)}{A\left(V_{i1} \cup V_{i2}\right)} \times 100\%$$

Results: Limits of repeatability, as percentages of baseline value, are shown in columns 2 and 3 of **Table 1** for median gADC and tDV respectively. Furthermore, the mean PoA score across the entire patient cohort for each observer is shown in column 4. Regions outlined for two patients are demonstrated in **Figure 1** and **Figure 2** for observers O1 and O2 respectively. These repeatability limits can be applied to clinical data for assessing significant changes in gADC and tDV estimates after treatment. This is demonstrated on a similar pilot cohort of 13 patients with bone metastases, using the same imaging parameters (**Figure 3**).

Observer ID	Median gADC (%)		tDV (%)		Mean *PoA* (%)
O1	-13.5	+15.6	-31.4	+45.7	71.5
O2	-7.2	+7.8	-44.2	+79.3	65.1

Table 1: Repeatability analysis (columns 2 and 3) and mean PoA (column 4) for both observers.

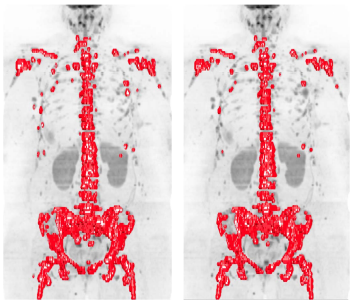

Figure 1: Volumes defined by O1 for a patient diagnosed with metastatic breast cancer.

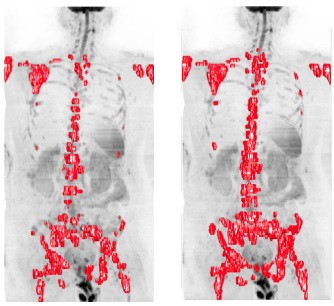

Figure 2: Volume defined by O2 for a patient diagnosed with metastatic prostate cancer.

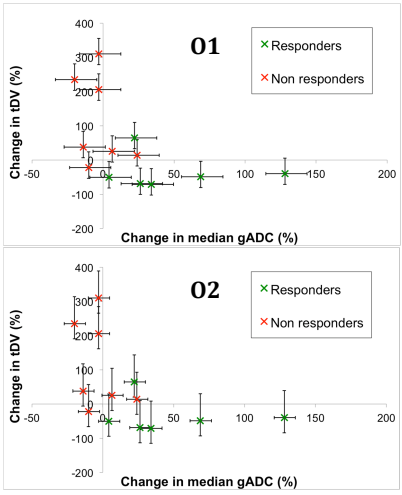

Figure 3: Changes in tDV and median gADC for a cohort of 13 patients diagnosed with bone metastases from a range of primary malignancies. Error bars represent repeatability limits (**Table 1**).

Discussion: User repeatability limits for percentage changes in tDV and gADC have been derived for two observers defining regions of interest in WBDWI data. Observer 1 (O1) demonstrated improved repeatability for tDV estimates, in agreement with their higher *PoA* score (**Table 1**). Conversely, O2 demonstrated improved repeatability in median gADC estimates. Close inspection of results revealed an outlier in one patient for O1 where the second reading included diffuse bone marrow disease within ROIs, whereas these regions were not included in the first reading. This highlights the importance of establishing inclusion/exclusion criteria in WBDWI studies before ROI delineation, especially deciding whether or not to include regions of necrosis and on the minimum size of lesions to include. We observe that in a pilot cohort of 13 patients, changes seen clinically are substantially greater than those required for significance indicating the feasibility and utility of such metrics in assessing disease response.

References: [1] Blackledge *et al*., Proc. 20th Annual Meeting ISMRM 2012, 255, [2] Blackledge *et al*., Proc. 20th Annual Meeting ISMRM 2011, 4102, [3] Blackledge *et al*., Proc. 20th Annual Meeting ISMRM 2011, 4111, [4] Altman and Bland, Statistician, 32:307-317, 1983. **Acknowledgments:** CRUK and EPSRC Cancer Imaging Centre in association with the MRC and Department of Health grant (England) C1060/A10334; NHS funding to the NIHR Biomedical Research Centre; and also EPSRC Platform Grant EP/H046526/1. Dr T. Feiweier (Siemens Medical Sector) for developing the DWI sequence.