Whole Body Diffusion Weighted Imaging (WB-DWI) for Assessing Treatment Response in Myeloma: Preliminary Observations

Sharon L. Giles, Christina Messiou, David J. Collins, Veronica A. Morgan, Catherine J. Simpkin, Sharon West, Faith E. Davies, Gareth J. Morgan, Nandita M. deSouza.

Original Research

Advances in Knowledge:

1. Whole body diffusion-weighted (WB-DW) images assessed by visual inspection and observer scoring are potentially useful for response evaluation in myeloma.
2. Quantitative measurement of bone marrow Apparent Diffusion Coefficient (ADC) using a volumetric segmentation technique is highly repeatable in normal volunteers and myeloma patients.
3. ADC histograms derived from segmented tumour regions on WB-DWI show potential as a biomarker of response in myeloma.

Implications for Patient Care:

1. WB-DWI can potentially be used in conjunction with conventional laboratory assessments to assess treatment response.
2. WB-DWI may reduce the requirement for serial bone marrow biopsies for longitudinal assessment in patients with myeloma.

Abstract

Purpose: To determine the feasibility of using whole body diffusion-weighted imaging (WB-DWI) for assessing treatment response in myeloma.

Materials and Methods: This was a HIPAA compliant prospective, single-institution study with local research ethics committee approval. Written informed consent was obtained from each subject. 8 healthy volunteers (Cohort 1a) and 7 myeloma patients (Cohort 1b) were scanned twice to assess repeatability of quantitative Apparent Diffusion Coefficient (ADC) estimates. 34 further myeloma patients (Cohort 2) underwent WB-DWI before treatment; 26 completed a post treatment scan. WB-DW data were compared pre and post treatment using qualitative (observer scores) and quantitative (WB segmentation of marrow ADC) methods. Serum paraproteins/light chains or bone marrow biopsy defined response.

Results: WB-DW image scores were significantly different between observers ($p<0.001$), but change in scores between observers following treatment was not ($p=0.49$). Sensitivity and specificity for detecting response by observer scores were 86% (18/21) and 80% (4/5) for both observers. ADC measurement was highly repeatable: mean coefficient of variation was 3.8% in healthy volunteers and 2.8% in myeloma patients. Pre treatment ADC in Cohort 2 was significantly different from Cohort 1a ($p=0.03$), but not from Cohort 1b ($p=0.2$). Mean ADC increased in 95% (19/20) of responding patients and decreased in all (5/5) non-responders ($p=0.002$); a 3.3% increase in ADC identified response with 90% sensitivity and 100% specificity; an 8% increase (>repeatability of cohort 1b) gave 70% sensitivity, 100% specificity. There was a significant negative correlation between change in ADC and change in laboratory markers of response ($r=-0.614$, $p=0.001$).

Conclusion: Our preliminary work demonstrates WB-DWI is a repeatable, quantifiable technique for assessment of treatment response in myeloma.

Introduction

Multiple myeloma, a malignant plasma cell disease, mainly involves the axial skeleton and proximal long bones. X-ray and MRI techniques are therefore routinely included in the diagnostic workup and staging of myeloma patients [1-4]. However, International Myeloma Working Group (IMWG) response criteria [5] are largely based on analysis of serum and urinary levels of paraproteins and light chains but are unsuitable where these are unmeasurable [6], necessitating painful serial bone marrow biopsies, which may not be representative of the extent or histological severity of disease [7]. Despite these limitations, imaging techniques are not recommended for assessing treatment response; x-ray and CT are unhelpful because appearances of lytic lesions are usually unchanged despite response to treatment and conventional MRI sequences are limited in their ability to measure response [8-12]. MRI and ^{18}FDG Positron Emission Tomography (PET) have not been incorporated into response criteria because their role remains unproven, although further studies to evaluate these techniques have been recommended [13].

In soft tissue tumours, Diffusion Weighted Imaging (DWI) is proving invaluable for assessing response [14]. The capability of achieving WB-DWI provides high contrast between normal and pathological marrow [15] and offers potential to quantify response within individual lesions and the whole skeleton [16,6]. The aim of this study therefore was to determine the feasibility of using WB-DWI for assessing response in myeloma.

Methods

This was a prospective, HIPAA compliant, single-institution study from August 2011 to May 2013 with approval from the local research ethics committee. Written informed consent was obtained from each subject.

Volunteers and Patients

To conduct a repeatability pilot study of the WB-DW technique, eight healthy female volunteers with visible marrow on DWI (aged 25-40) (Cohort 1a) and seven patient volunteers (mean age 53 years, range 37-69 years [4 men: mean age 44 years, range 37-51 years, 3 women: mean age 65 years, range 63-69 years]) with a proven diagnosis of myeloma (an elevated level of paraproteins in the blood (>30g/L) or urine (no specific concentration required) and/or the presence of clonal plasma cells (>10%) in a bone marrow biopsy sample [3] (Cohort 1b) were recruited. In each case, WB-DWI was acquired at two time-points within a week. Lack of previously reported ADC repeatability of WB-DWI in patients with bone marrow disease necessitated a pilot repeatability study of the technique in this group of patients.

For the study of response to treatment, 34 further patients with active myeloma (mean age 61 years, range 42-73 years [18 men: mean age 61 years, range 42-73 years, 16 women: mean age 62 years, range 45-72 years]) (Cohort 2) were recruited and scanned prior to treatment, including those at first presentation, relapse or progression on current therapy. Exclusion criteria were suspected spinal cord compression, patients requiring radiotherapy during treatment and contra-indications to MRI.

33 Cohort 2 patients underwent treatment (one died): eight received the same

treatment regime of Cyclophosphamide, Lenalidomide and Dexamethasone, seven received Cyclophosphamide, Bortezomib and Dexamethasone, five received Lenalidomide and Dexamethasone, whilst 13 others all received different combination treatment regimes. 15 of the 33 patients also received growth factors, e.g. Granulocyte Colony Stimulating Factor (GCSF), during treatment to boost white cell count [17]. Treatment was initiated at a median of 6 days after the pre- treatment scan (Inter Quartile Range (IQR) 0 - 26 days).

A post treatment scan was acquired in 26 of the 33 patients at a median of 13 weeks after commencement of treatment (IQR 12 - 13.1 weeks). Of the other seven patients, four declined, one died, one was excluded due to artefacts on the baseline scan and one received radiotherapy. Serum paraproteins and light chains were measured at both time points and bone marrow biopsies were acquired in all patients pre treatment, and post treatment in non-secretory disease or as part of another therapy trial (n=9). Paraproteins, light chains, or bone marrow histology were used to classify patients' response status according to IMWG response criteria [5].

Image Acquisition

Using an Avanto 1.5T system (Siemens Healthcare, Erlangen, Germany) a WB study was achieved by the serial acquisition of contiguous body regions. All subjects were scanned supine with arms by their sides. Coil elements were positioned from skull vertex to knees in patients but limited to the abdomen and pelvis for normal volunteers. Axial T1W spin echo (slice thickness 5mm, no gap, FOV 430mm, phase direction AP, repetition time (TR)/echo time (TE) 386/4.8 ms, flip angle 70^0, matrix

256/154) and coronal VIBE Dixon 3D gradient echo breath-hold sequences (52 slices per slab, FOV 470mm, TR/TE 7/2.38,4.76 ms, flip angle 3^0, matrix 192x192) were acquired, followed by axial DW sequences (single shot double spin echo echo-planar technique with STIR fat suppression in free breathing). b values of 50 and 900 s/mm^2 were applied in 3 orthogonal directions and combined to provide isotropic trace images. The choice of the higher b value reflected ADCs expected from aqueous soft tissues such as tumor infiltrated marrow. DW sequences were acquired in blocks of 50 slices (slice thickness 5mm, no gap, FOV 430mm, phase direction AP, parallel imaging (GRAPPA) factor 2, TR 14800 ms, TE 66 ms, Inversion time (TI) 180 ms, voxel size 2.9 x 2.9 x 5 mm, number of signal averages acquired 4, matrix 150 x 150, bandwidth 1960 Hz per pixel). The optimised scanner carrier frequency offset used for the top station was applied for all other stations [18]. The same shim gradient currents were applied for each station [19]. Total acquisition time was 50-60 minutes.

Image Analyses

Observer Scoring:

Image appearances were scored independently by two radiologists with 5 years' experience of DWI in bone (CM and NDS), blinded to clinical information. Composed WB-DW inverted grey scale Maximum Intensity Projection (MIP) b=900 s/mm^2 images were used to categorise lesion number and largest lesion dimension in each of 7 body regions and then scored according to the system defined in Table 1. Post-treatment data were scored on a separate occasion, using pre-treatment images to

standardize window widths and levels of the MIP images. In diffuse disease, to improve the spatial perception of response and assist scoring, the b900 s/mm^2 images and ADC maps were assessed alongside the MIPs. Patients were initially classified as responding to treatment if their total score decreased by at least 30% but optimal threshold analysis was also undertaken.

ADC Analysis:

Quantitative ADC analysis was undertaken using OncoTreat software (Siemens Healthcare, Erlangen, Germany), which is not yet commercially available. Raw WB-DW data were separated into 3 series of b50 s/mm^2, b900 s/mm^2 and ADC maps for all stations for each time point in all volunteers (TP1 and TP2) and pre and post treatment in patients. Data from each time point were registered using reference b50 s/mm^2 images, applying a deformation field to allow alignment of all images. Multiple volumetric ROIs were outlined against 3D multi-planar reformat (MPR) images of the pre treatment b900 s/mm^2 data using a semi-automated technique, whereby one set of 'seeds' was manually placed inside every region to be included in the analyses, with a second set being defined to exclude surrounding areas. The software then generated outlines of volumes to be included in the segmentation, based on signal intensity values. Segmentations included all areas of visible marrow within vertebral bodies, pelvis, femora, proximal humerii and sternum. ADC values for every voxel within the segmented volume were recorded and displayed as histograms; metrics representing histogram centile values, skew and kurtosis were recorded and compared between TP1 and TP2 for Cohorts 1a and 1b and between pre and post treatment scans for Cohort 2. All segmentations were undertaken by one observer (SG). In 8 volunteers, segmentations were repeated on a separate occasion to

assess intra-observer variability, as well as by a second observer (CS) to assess inter-observer variability of the technique.

Statistical Analyses

Observer Scoring:

The sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) (adjusted for prevalence of response [20]) of the WB-DWI image scores for detecting response were calculated using contingency tables. Standard laboratory assessments were the reference standard. Paired t-tests evaluated differences in scores between observers.

ADC Analysis:

The number of voxels within segmentations was constant between time-points for each subject and was noted. The percentage changes in histogram ADC metrics were recorded and a value for the t-statistic (difference in means divided by the standard error (SE) of those means) was calculated to test the null hypothesis that there was no change in the mean and variance of ADC values between scans at each time-point. This calculation assumed a normal distribution of ADC values for each patient; the very large number of voxels included within the calculations meant that the test could still be used in a non-normal distribution [21].

Repeatability: Repeatability of the Cohort 1a and 1b ADC measurements was assessed by Bland-Altman analysis [22], whereby the respective coefficients (indicating the greatest difference between replicate ADC measurements in 95% of

paired observations) were calculated as r=1.96 x standard deviation of the percentage ADC difference. Bland-Altman plots have been widely used to assess repeatability in a variety of studies [23-26]. In addition, a paired t-test tested for significant differences between the repeat measurements in cohorts 1a and 1b. Intra- and inter-observer variability were assessed with intra-class correlation coefficients using MedCalc for Windows, version 9.5.0.0 (MedCalc Software, Belgium).

Response to Treatment: Statistical analysis was undertaken using SPSS for Windows software (Version 20, SPSS). Normality plots and the Kolmogorov-Smirnov and Shapiro-Wilk tests were used to confirm normality. Independent samples t-tests were used to determine whether the distribution of the measured variables was different in responding and non-responding patients. Receiver operating characteristic (ROC) curve analysis gave an area under the curve (A_z) and was used to calculate the sensitivity and specificity of each variable for predicting response. In addition, a correlation coefficient evaluated the strength of the relationship between the percentage change in measured variable and the percentage change in the laboratory measure of treatment response across the 2 time-points.

Results

Myeloma sub-types in cohort 2 were IgG (n=23: 15κ,8λ), IgA (n=4: 3κ,1λ), light chains only (n=4: 2κ,2λ) and non-secretory myeloma (n=3). The mean percentage of plasma infiltrated cells in those with a quantifiable bone marrow biopsy prior to treatment (n=31) was 44% (range 0-100%), mean serum paraprotein (n=27) was 21.3 g/L (range 0-62 g/L) and mean serum free light chains (n=22) were κ: 446mg/L

(range:0-2480mg/L) and λ: 584mg/L (range 0-5000mg/L).

21 of 26 patients who underwent the post treatment MRI scan were classified as responders (2 complete response, 5 very good partial response, 14 partial response) and 5 as non-responders (3 stable disease, 2 progressive disease), based on change in serum paraproteins (n=20), serum free light chains (n=3) and bone marrow biopsy (n=3).

Observer scores

Although pre treatment WB-DWI raw scores were significantly different between observers ($p<0.001$), the assessment of response was similar between observers, with no significant difference in the percentage change in score following treatment either on a per patient basis ($p=0.49$), or when assessed by body region ($p=0.23$) (Table 2). There was concordance in classification of response between observers in 24 of 26 patients if imaging response was defined by a ≥ 30% reduction in score; in 2 patients both observers considered that a reduction in score had occurred, but did not meet the ≥ 30% threshold for 1 observer. When a ≥10% reduction in score was used to define imaging response, there was complete agreement between observers; both observers correctly identified response in 22 patients (18 responders, 4 non-responders), whilst 4 patients were incorrectly classified (1 false positive, 3 false negative). The sensitivity, specificity, PPV and NPV of the WB-DWI to detect treatment response in comparison to reference standard laboratory assessments were therefore 86% (18/21) and 80% (4/5), 95% (18/19) and 57% (4/7) for both observers. Given a prevalence of response of 80%, adjusted PPV (Bayes

formula) was 94.5%, NPV 58.8%. Subjectively, there was no difference in DW appearances between those receiving GCSF, compared with those who were not and only 1 of the patients falsely assessed as a non-responder by image scoring had received GCSF.

ADC Analysis

Marrow ADC values in the cohort 1b patients (mean±SD 859.4±117.7 $x10^{-6}mm^2/s$), were similar to the pre treatment ADC values in the cohort 2 patients (mean±SD 813.5±114 $x10^{-6}mm^2/s$), with no significant difference in any ADC metric. ADC values in the normal volunteers were lower than in the patients (mean±SD 734.9±53.6 $x10^{-6}mm^2/s$), with median ADC being significantly different from the 2 patient groups (Table 3).

Repeatability measurements

ADC measurement was found to be highly repeatable in segmented marrow volumes; mean coefficient of variation was 3.8% in the volunteer cohort 1a (r=80.1 $x10^{-6}mm^2/s$, r%=10.8, group CI=28.3 $x10^{-6}mm^2/s$) (Figure 1a) and 2.8% in the patient cohort 1b (r=71.1 $x10^{-6}mm^2/s$, r%=8.3, group CI=26.9 $x10^{-6}mm^2/s$) (Figure 1b). A paired t-test between the 2 observations indicated no significant difference between subject measurements for either cohort 1a or 1b (p=0.41, p=0.67 respectively). The measurement was also highly consistent, with intraclass correlation coefficients (ICC) of 0.981 (95% confidence interval (CI): 0.907-0.996) and 0.991 (95% CI: 0.956-0.998) for intra- and inter-observer comparisons respectively.

Response to Treatment

In one (responding) patient quantitative analysis was not possible due to lack of visible bone marrow at both time points, precluding successful segmentation. Mean ADC increased in all but one of 20 responding patients (mean group increase 19.8%±21.5%), whilst ADC decreased in all 5 patients who did not respond to treatment (mean decrease 3.2%±2.2%) (Figure 2a). In 2 patients with progressive disease, review of the pre-treatment segmentations confirmed that regions of marrow that encompassed new areas of disease post treatment had been included in the original segmentation. The differences in percentage change in ADC metrics between those who responded to treatment and those who did not is given in Table 4 and illustrated in Figures 3-5. A paired t-test indicated that ADC differences pre and post treatment were significant (p= 0.004). Given these differences, the number of subjects required to detect a statistically significant change following treatment with 85% power and α=0.05 is 24 (for the same power and α, it would require 326 patients to detect the differences in the 2 baseline measurements noted in cohort 1b).

The t-statistic was negative in 19 of 20 responding patients and positive in all 5 non-responding patients, with a significant difference between responders and non-responders (mean t= -85.2 and 16.8 respectively, U=96, p=0.002) (Figure 2b).

ROC analysis of ADC change over the segmented volume indicated that a 3.3% increase in ADC correctly identified response to treatment with a 90% sensitivity and 100% specificity ($A_z = 0.950$). However, if repeatability in a positive direction is considered (8%), sensitivity decreased to 70% (specificity 100%).

There was a significant negative correlation between percentage change in mean ADC and percentage change in laboratory marker between pre and post treatment time points ($r = -0.614$, $p=0.001$), and a significant positive correlation between the t-statistic and the change in laboratory marker ($r = 0.699$, $p<0.001$).

Discussion

Our study demonstrates that WB-DWI is a repeatable, quantifiable technique which can differentiate responders from non-responders in myeloma bone disease with high sensitivity and specificity. Change in bone marrow signal intensity with treatment in myeloma has been previously explored. Bannas et al [27] demonstrated 63.6% (7/11) sensitivity and 86.4% (19/22) specificity compared to standard laboratory methods with conventional WB T1W and STIR techniques after stem cell transplant, which was considered inadequate for assessing response. The visual scoring assessment of DW images in this study improved on this, with a sensitivity and specificity of 86% (18/21) and 80% (4/5) respectively, but criteria for imaging response are not universally accepted. Bannas et al [27] defined imaging response categories on the basis of changes in size or number of lesions, whilst Hillengass et al [28] assessed both the number of focal lesions and the degree of diffuse infiltration. We therefore attempted to incorporate these elements into the evaluation of subjective scores of whole body MIPS, additionally using the axial b900 s/mm^2 images and ADC maps if initial assessment indicated that disease was diffuse. The variability in observer scores pre-treatment was likely to be part of the learning curve of the interpretation of these images; in future viewing the MIP images together with the source data may further improve inter-reader variability. The choice

of a 30% change in score to indicate response was arbitrary and based on the use of a ≥ 30% decrease in the sum of diameters of target lesions on RECIST criteria to define response [29].

Mean marrow ADC of the whole skeleton increased in responders but not in non-responders, likely as a result of tumor cell lysis. Previous attempts at quantification have classified response based on assessment of target lesions [6,30] and we have shown an accuracy consistent with these studies, but with the added advantage of assessing the entire skeleton. In the one responding patient incorrectly classified by ADC analysis, however, the segmentation was difficult due to low disease burden (2-5% affected cells on BMT) resulting in little visible disease on the DWI, although the qualitative analyses correctly identified this patient as a responder. A quantifiable increase in ADC with treatment has been shown in a variety of tumors [14] but a particular difficulty in bone marrow is the presence or return of stromal fat which accompanies response and counteracts the rise in ADC. Thus the direction of ADC change in our study is consistent with other studies but the degree of change is smaller [6,30]. It is likely that these smaller ADC increases can be largely explained by the timing at which the post treatment assessment was made; Messiou et al [16] found an increase in ADC in responders at 4-6 weeks, which was followed by a decrease at 20 weeks, likely reflecting early necrosis/oedema, followed by a later return of normal marrow fat, thereby reducing the overall increase in ADC. In addition, the volumetric segmentation method used in this study resulted in the inclusion of some normal marrow with the diseased marrow, so reducing the overall ADC change seen. There was also a heterogeneous range of treatments administered to the patients, which varied in their mechanism of action. Thus the

PPV and NPV calculated are pertinent to this patient population and may well differ if the prevalence of response were different. The use of the t-statistic, to capture changes in the shape and position of ADC histograms potentially increases confidence in determining response to treatment. The changes in the higher centile values in responding patients suggests that these voxels are likely to represent tumor infiltrated regions where oedema occurs following treatment while the lower centile values that show little change may represent areas of marrow fat [15]. For future studies, consideration needs to be given to the optimum time point for assessing treatment response. The use of the fat/water imaging to calculate fat fraction maps would also potentially provide further understanding of longitudinal ADC changes as disease responds to treatment.

WB repeatability estimates of ADC marrow values have not been well established in the literature. In our study, we found these to be highly repeatable in a cohort of myeloma patients. They were also comparable with similar estimates in a cohort of healthy volunteers, indicating that repeatability estimates derived from volunteer studies can be applied to patient studies using this segmentation method. Also, though the repeatability study was a pilot, the data obtained is comparable to data from other soft tissue tumors [23,25,26] and indicates that the size of ADC change in responders to treatment was approximately twice that of the repeatability of the measurement.

Correlations between percentage change in laboratory markers, ADC and the t statistic demonstrated in our study were stronger than those previously published for DWI [6] or conventional MRI [27,28]. ADC values of focal myeloma lesions have also been inversely correlated with serum paraprotein concentrations [31], and a

significant positive correlation demonstrated between ADC and bone marrow cellularity [7].

Limitations of our method were that the skull remains less well demonstrated on DWI, partly due to very high signal on DWI of normal brain [32], but also because the most commonly encountered metallic implants were irremovable dental fixations, causing local artefacts. This was the site of greatest disparity in raw image score between observers. Similar difficulties in the skull have been documented in prostatic metastases [33]. In the quantitative ADC analysis, one case was not evaluable due to lack of visible marrow, however serum paraproteins were not measurable in 8 of 33 patients and bone marrow biopsy was insufficient in 7 patients, which argues for our imaging based method. The effect of GCSF on images, however, remains a confounding factor. Although administration of GCSF may cause increased signal intensity on DWI [32], no subjective differences in DWI appearances were seen in this study in those who received GCSF, but additional analysis was not undertaken because the doses and schedules for GCSF were extremely variable between patients in relation to the scan time-points. Further exploration of the degree, onset and duration of DW appearances in relation to GCSF would be helpful.

In conclusion, our study shows the potential of WB-DWI as a biomarker of treatment response in myeloma. Simple analysis of image appearances by experienced observers was able to correctly identify response to treatment with a high sensitivity. ADC analysis was found to be highly repeatable and also provided good sensitivity for assessing response. Ongoing work is underway to establish consensus on analysis methods.

Acknowledgments: Dr. P. Gall & Dr. T Feiweier at Siemens Healthcare for provision of the OncoTreat software and WIP DWI sequence, Dr. M. Blackledge at Institute of Cancer Research for optimisation of shimming techniques used in the acquisition and Mrs K Thomas at Royal Marsden Hospital for statistical advice.

References

1. International Myeloma Working Group. Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: a report of the International Myeloma Working Group. British Journal of Haematology. 2003; 121(5): 749-57

2. Dimopoulos M, Terpos E, Comenzo Rl, Tosi P, Beksac M, Sezer O et al. International myeloma working group consensus statement and guidelines regarding the current role of imaging techniques in the diagnosis and monitoring of multiple myeloma. Leukemia. 2009; 23: 1545-56

3. Dimopoulos M, Kyle R, Fermand JP, Rajkumar SV, San Miguel J, Chanan-Khan A et al. Consensus recommendations for standard investigative workup: report of the International Myeloma Workshop Consensus Panel 3. Blood. 2011; 117(18): 4701-05

4. Bird JM, Owen RG, D'Sa S, Snowden JA, Pratt G, Ashcroft J et al. Guidelines for the diagnosis and management of multiple myeloma 2011. British Journal of Haematology. 2011; 154: 32-75

5. Durie BGM, Harousseau JL, Miguel JS, Blade J, Barlogie B, Andersin K et al. International uniform response criteria for multiple myeloma. Leukemia. 2006; 20: 1467-73

6. Horger M, Weisel K, Horger W, Mroue A, Fenchel M, Lichy M. Whole-Body Diffusion-Weighted MRI With Apparent Diffusion Coefficient Mapping for Early Response Monitoring in Multiple Myeloma: Preliminary Results. American Journal Roentgenology. 2011; 196: 790-95

7. Hillengass J, Bauerle T, Bartl R, Andrulis M, Mcclanahan F, Laun FB, et al. Diffusion-weighted imaging for non-invasive and quantitative monitoring of bone marrow infiltration in patients with monoclonal plasma cell disease: a comparative study with histology. British Journal of Haematology. 2011;153: 721-8

8. Horger M, Weisel K, Bares R, Ernemann U, Claussen CD, Lichy M et al. Modern imaging techniques during therapy in patients with multiple myeloma. Acta Radiologica. 2011; 52: 881-8

9. Tan E, Weiss BM, Mena E, Korde N, Choyke PL, Landgren O. Current and future imaging modalities for multiple myeloma and its precursor states. Leukaemia & Lymphoma. 2011; 52(9): 1630-40

10. Terpos E, Moulopoulos LA, Dimopoulos MA. Advances in Imaging and the Management of Myeloma Bone Disease. Journal of Clinical Oncology. 2011; 29: 1907–15

11. Hanrahan CJ, Christensen CR, Grim JR. Current Concepts in the Evaluation of Multiple Myeloma with MR Imaging and FDG PET/CT. RadioGraphics. 2010; 30:127–42

12. Delorme S, Baur-Melnyk A. Imaging in multiple myeloma. European Journal of Radiology. 2009; 70:

401-8

13. Rajkumar SV, Harousseau JL, Durie B, Anderson KC, Dimopoulos M, Kyle R et al. Consensus recommendations for the uniform reporting of clinical trials: report of the International Myeloma Workshop Consensus Panel 1. Blood. 2011; 117(18): 4691-95

14. Bonekamp S, Corona-Villalobos CP, Kamel IR. Oncologic applications of diffusion-weighted MRI in the body. Journal of Magnetic Resonance Imaging. 2012; 35(2): 257-79

15. Messiou C, Collins DJ, Morgan VA, DeSouza NM. Optimising diffusion weighted MRI for imaging metastatic and myeloma bone disease and assessing reproducibility. European Radiology. 2011; 21: 1713-18

16. Messiou CM, Giles SL, Collins DJ, Davies FE, Morgan GM, Dines S et al. Assessing Response of Myeloma Bone Disease with Diffusion Weighted MRI. British Journal of Radiology. 2012; 85(1020): 1198-203

17. Gertz MA. Current status of stem cell mobilization. British Journal of Haematology. 2010; 150(6): 647-62

18. Collins DJ, Blackledge M. Techniques and Optimisation. In: Koh DM & Thoeny HC, editors. Diffusion-Weighted MR Imaging: Applications in the Body. Berlin: Springer; 2010. p19-32

19. Koh DM, Blackledge M, Padhani AR, Takahara T, Kwee TC, Leach MO et al. Whole-body diffusion-weighted MRI: tips, tricks and pitfalls. American Journal of Roentgenology. 2012; 199(2): 252-62

20. Fletcher, Robert H. Fletcher ; Suzanne W. Clinical epidemiology : the essentials (4th ed.). Baltimore, Md.: Lippincott Williams & Wilkins. 2005.

21. Armitage P, Berry G. Statistical Methods in Medical Research. 3rd Ed. London: Blackwell Science Ltd; 1994

22. Bland JM, Altman DG. Measuring agreement in method comparison studies. Statistical Methods in Medical Research. 1999, 8(2): 135-60

23. Corona-Villalobos CP, Pan L, Halappa VG, Bonekamp S, Lorenz CH, Eng J et al. Agreement and reproducibility of apparent diffusion coefficient measurements of dual-b-value and multi-b-value diffusion-weighted magnetic resonance imaging at 1.5 Tesla in phantom and in soft tissues of the abdomen. Journal of Computer Assisted Tomography. 2013, 37(1): 46-51

24. Kim SY, Lee SS, Park B, Kim N, Kim JK, Park SH et al. Reproducibility of measurement of apparent diffusion coefficients of malignant hepatic tumors: effect of DWI techniques and calculation methods. Journal of Magnetic Resonance Imaging. 2012, 36(5): 1131-8

25. Kyriazi S, Collins DJ, Messiou C, Pennert K, Davidson RL, Giles SL et al. Metastatic ovarian and primary peritoneal cancer: assessing chemotherapy response with diffusion-weighted MR imaging – value of histogram analysis of apparent diffusion coefficients. Radiology. 2011, 261(1): 182-92

26. Koh DM, Blackledge M, Collins DJ, Padhani AR, Wallace T, Wilton B et al. Reproducibility and changes in the apparent diffusion coefficients of solid tumours treated with combrestatin A4 phosphate and bevacizumab in a two-centre phase 1 clinical trial. European Radiology. 2009, 19(11): 2728-38

27. Bannas P, Hentschel HB, Bley TA, Treszl A, Eulenberg C, Derlin T et al. Diagnostic Performance of whole-body MRI for the detection of persistent or relapsing disease in multiple myeloma after stem cell transplantation. European Radiology. 2012; 22(9): 2007-12

28. Hillengass J, Ayyaz S, Kilk K, Weber MA, Hielscher T, Shah R et al. Changes in magnetic resonance imaging before and after autologous stem cell transplantation correlate with response and survival in multiple myeloma disease. Haematologica. 2012; 97(11): 1757-60

29. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). European Journal of Cancer. 2009; 45: 228-47

30. Fenchel M, Konaktchieva M, Weisel K, Kraus S, Claussen Cd, Horger M. Response Assessment in Patients with Multiple Myeloma during Antiangiogenic Therapy using Arterial Spin Labelling and Diffusion-Weighted Imaging: A Feasibility Study. Academic Radiology. 2010; 17: 1326-33

31. Sommer G, Klarhofer M, Lenz C, Scheffler K, Bongartz G, Winter L. Signal characteristics of focal bone marrow lesions in patients with multiple myeloma using whole body T1w-TSE, T2w-STIR and diffusion-weighted imaging with background suppression. European Radiology. 2011; 21: 857 – 862

32. Padhani AR, Koh DM, Collins DJ. Whole-Body Diffusion-weighted MR Imaging in Cancer: Current Status and Research Directions. Radiology. 2011; 261(3): 700-18

33. Nemeth AJ, Henson JW, Mullins ME, Gonzalez RG, Schaefer PW. Improved detection of skull metastasis with diffusion-weighted MR imaging. American Journal of NeuroRadiology. 2007; 28(6): 1088-92

Table 1: Summary of scoring system used to score image appearances

No. of lesions	Score	Max size (mm) of a single lesion	Score
Diffuse disease	4	Diffuse	4
>10	3	>20	3
>5	2	10-20	2
1 – 4	1	<10	1
0	0		

Note: In patients with diffuse disease a generalised reduction in signal intensity of any region following treatment was recorded by dividing scores by a factor of 2.

Table 2: Mean scores assigned to the pre and post treatment WB-DWI images for all patients, by body area for each observer (Obs)

	WB-DWI Score Pre Treatment		WB-DWI Score Post Treatment		WB-DWI Score % reduction post treatment		P value *
Region	**Obs 1**	**Obs 2**	**Obs 1**	**Obs 2**	**Obs 1**	**Obs 2**	
Skull	127	41	75	19	41	54	0.85
C Spine	150	121	96	74	36	39	0.44
D Spine	174	169	107	111	39	34	0.76
L Spine	159	143	98	81	38	43	0.87
Pelvis	156	136	96	83	38	39	0.40
Ribs	181	162	124	109	31	33	0.37
Long Bones	139	131	87	82	37	37	0.59
Total	1086	903	683	559	37	38	0.23

* Paired t test

Table 3: Comparison of Histogram Metrics (mean and centile values of ADC) between the 3 cohorts studied

Histogram metrics ADC x10^{-6}mm^2/s	**Healthy Volunteers Cohort 1a (n=8)** (mean ±SD)	**Patient Volunteers Cohort 1b (n=7)** (mean ±SD)	**Patients Cohort 2 (n=25)** (mean ±SD)	**P value * (1a vs 2)**	**P value * (1b vs 2)**
Mean	734.9 ± 53.6	859.4 ± 117.7	813.5 ± 114.1	0.061	0.324
10th Centile	533.0 ± 28.1	551.1 ± 82.9	537.0 ± 99.3	0.726	0.420
25th Centile	586.4 ± 28.0	656.0 ± 109.0	624.8 ± 95.5	0.550	0.191
50th Centile (median)	657.4 ± 36.2	793.0 ± 143.4	733.4 ± 103.3	**0.028**	0.207
75th Centile	778.6 ± 69.9	981.6 ± 173.5	911.9 ± 148.0	**0.017**	0.302
90th Centile	1041.5 ± 178.7	1262.1 ± 152.4	1207.5 ± 247.2	0.067	0.395
Skew	2.7 ± 0.5	2.0 ± 0.9	2.0 ± 0.8	**0.013**	0.824
Kurtosis	10.5 ± 5.0	9.2 ± 9.3	7.8 ± 5.8	0.081	1.000

* Mann Whitney U test

Table 4: Comparison of percentage change in Histogram Metrics (mean and centile values of ADC) between responding and non-responding patients post treatment

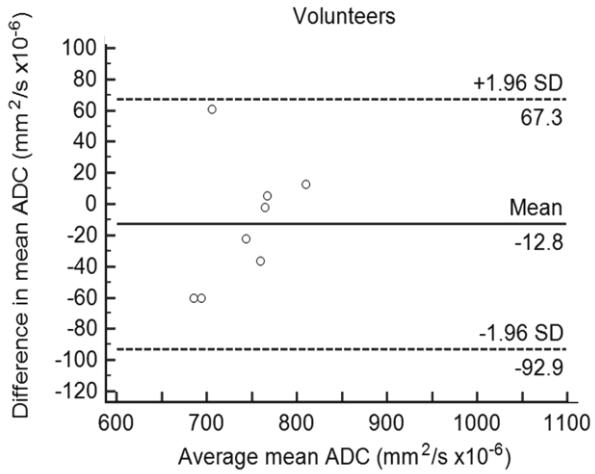
Percentage change in ADC metric post treatment	Responding patients (n=20)	Non-responders (n=5)	P value *
Mean	19.8	-3.2	**0.002**
10th Percentile	4.7	-0.1	0.049
25th Percentile	14.3	-0.9	0.558
50th Percentile (median)	19.6	-2.0	**0.002**
75th Percentile	24.5	-4.4	**0.002**
90th Percentile	27.2	-6.5	**0.002**
Skew	-6.4	33.9	0.067
Kurtosis	8.4	59.8	0.118

* Mann Whitney U test

Figure 1:

Bland-Altman plots demonstrating repeatability of ADC measurement in (a) Cohort 1a: healthy volunteers and (b) Cohort 1b: patient volunteers. ADC values in cohort 1a are lower than in 1b, however the mean difference between the 2 measurements in each subject is not significantly different from zero, indicating no systemic bias, and does not increase with increasing value of ADC.

A).



B).

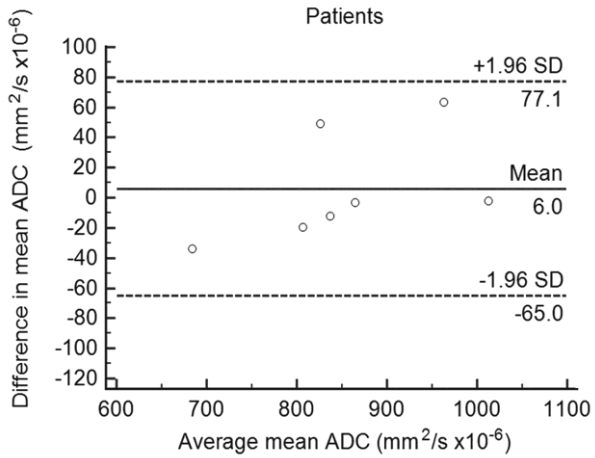
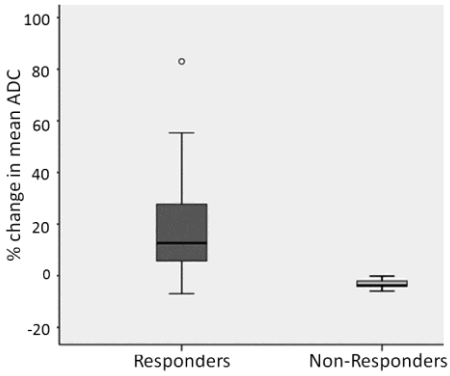


Figure 2:

Box and whisker plots comparing percentage change in (a) mean ADC and (b) in the t-statistic in 20 responding and 5 non-responding patients. In responders, there is a substantially greater percentage change in ADC and a lower t-statistic than in non-responders, with a wider distribution of values in keeping with a range of changes seen in this responding group.

A).



B).

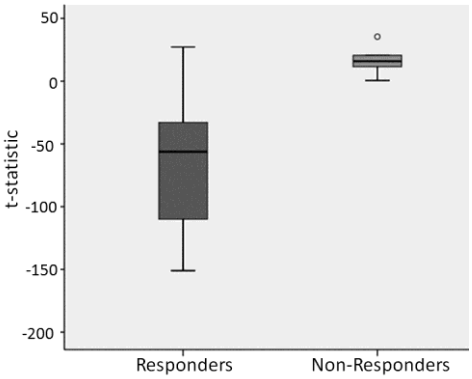
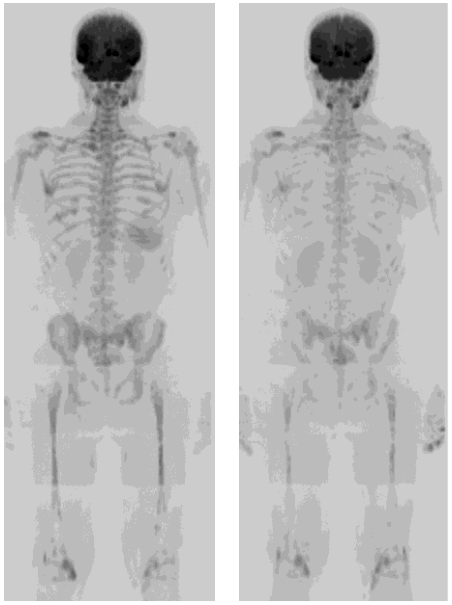


Figure 3:

65 year old female patient with non-secretory myeloma achieving a very good partial response to treatment. WB-DW (b=900s/mm^2) inverted grey scale images (a) before and (b) after 3 cycles of a Bortezomib based combination treatment regime show reduction in restricted diffusion within the marrow in all body regions. Corresponding ADC histograms (c) before and after treatment show displacement of the post treatment histogram to the right, indicating decreased cellularity.

A). B).



C).

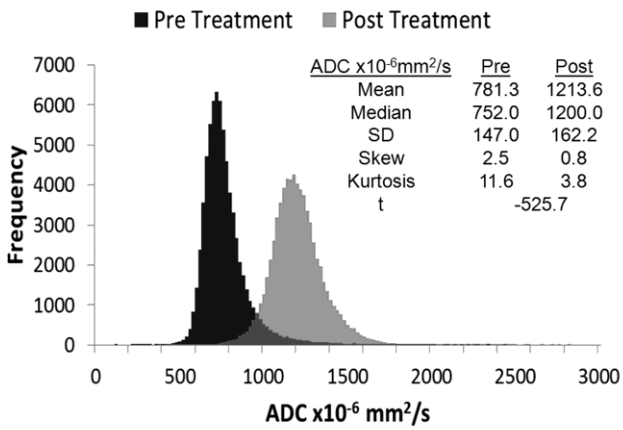
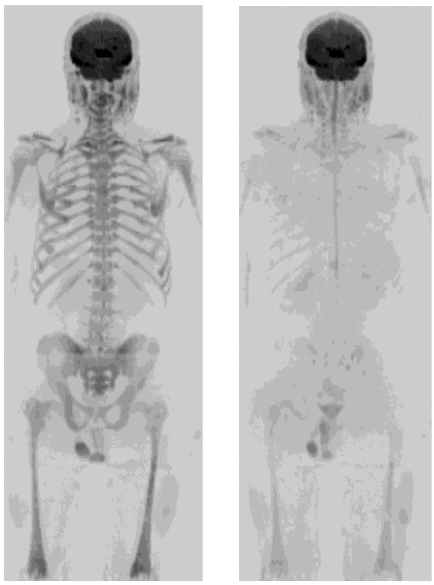


Figure 4:

64 year old male patient with IgGλ myeloma achieving a very good partial response to treatment. WB-DW (b=900s/mm^2) inverted grey scale images (a) before and (b) after 3 cycles of a Thalidomide based combination regime show reduction in restricted diffusion within the marrow in all body regions. Corresponding ADC histograms (c) before and after treatment show flattening of the histogram, with some shift to the right and an increase in low and high ADC values potentially indicating return of marrow fat, as well as decrease in tumor cellularity.

A). B).



C).

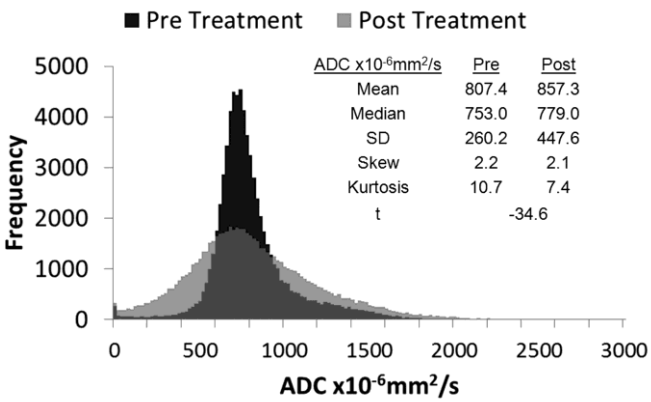


Figure 5:

72 year old female patient with IgGκ myeloma not responding to treatment. WB-DW (b=900s/mm^2) inverted grey scale images (a) before and (b) after 3 cycles of a Bortezomib based combination regime show some increase in marrow signal intensity. Corresponding ADC histograms (c) before and after treatment show little change in position or shape. At the post treatment time point, change in serum paraprotein levels indicated stable disease, but the patient later progressed. The small spikes of very low and very high ADC voxel values reflect the increased difficulty of undertaking segmentations where disease burden was lower.

A). B.)

C).