

Reproducibility of dynamic contrast-enhanced MRI in human muscle and tumours: comparison of quantitative and semi-quantitative analysis

Susan M. Galbraith,¹ Martin A. Lodge,² N. Jane Taylor,² Gordon J. S. Rustin,^{3*} Søren Bentzen,¹ J. James Stirling² and Anwar R. Padhani²

¹Gray Cancer Institute, Mount Vernon Hospital, Middlesex HA6 2JR, UK

²Paul Strickland Scanner Centre, Mount Vernon Hospital, Middlesex HA6 2JR, UK

³Department of Medical Oncology, Mount Vernon Hospital, Middlesex HA6 2JR, UK

Received 18 January 2001; Revised 4 July 2001; Accepted 24 July 2001

ABSTRACT: The purpose of this study was to determine the reproducibility of dynamic contrast-enhanced (DCE)-MRI and compare quantitative kinetic parameters with semi-quantitative methods, and whole region-of-interest (ROI) with pixel analysis. Twenty-one patients with a range of tumour types underwent paired MRI examinations within a week, of which 16 pairs were evaluable. A proton density-weighted image was obtained prior to a dynamic series of 30 T_1 -weighted spoiled gradient echo images every 11.9 s with an intravenous bolus of gadopentetate dimeglumine given after the third baseline data point. Identical ROIs around the whole tumour and in skeletal muscle were drawn by the same observer on each pair of examinations and used for the reproducibility analysis. Semi-quantitative parameters, gradient, enhancement and AUC (area under the curve) were derived from tissue enhancement curves. Quantitative parameters (K^{trans} , k_{ep} , v_e) were obtained by the application of the Tofts' model. Analysis was performed on data averaged across the whole ROI and on the median value from individual pixels within the ROI. No parameter showed a significant change between examinations. For all parameters except K^{trans} , the variability was not dependent on the parameter value, so the absolute values for the size of changes needed for significance should be used for future reference rather than percentages. The size of change needed for significance in a group of 16 in tumours for K^{trans} , k_{ep} and v_e was -14 to $+16\%$, ± 0.20 ml/ml/min (15%) and ± 1.9 ml/ml (6%), respectively (pixel analysis), and -16 to $+19\%$, ± 0.23 ml/ml/min (16%) and ± 1.9 ml/ml (6%) (whole ROI analysis). For a single tumour, changes greater than -45 to $+83\%$, ± 0.78 ml/ml/min (60%) and ± 7.6 ml/ml (24%), respectively, would be significant (pixel analysis). For gradient, enhancement and AUC the size of change needed for significance in tumours was ± 0.24 (17%), ± 0.05 (6%) and ± 0.06 (8%), respectively for a group of 16 (pixel analysis), and ± 0.96 (68%), ± 0.20 (25%) and ± 0.22 (32%) for individuals. In muscle, the size of change needed for significance in a group of 16 for K^{trans} , k_{ep} and v_e was -30 to $+44\%$, ± 0.81 ml/ml/min (61%) and ± 1.7 ml/ml (13%). For gradient, enhancement and AUC it was ± 0.09 (20%), ± 0.02 (8%) and ± 0.03 (12%). v_e , enhancement and AUC are highly reproducible DCE-MRI parameters. K^{trans} , k_{ep} and gradient have greater variability, with larger changes in individuals required to be statistically significant, but are nevertheless sufficiently reproducible to detect changes greater than 14–17% in a cohort of 16 patients. Pixel analyses slightly improve reproducibility estimates and retain information about spatial heterogeneity. Reproducibility studies are recommended when treatment effects are being monitored. Copyright © 2002 John Wiley & Sons, Ltd.

KEYWORDS: DCE-MRI; reproducibility; quantitative analysis; semi-quantitative analysis

*Correspondence to: G. J. S. Rustin, Department of Medical Oncology, Mount Vernon Hospital, Northwood, Middlesex HA6 2JR, UK.

Email: rustin@mtvern.co.uk

Contract/grant sponsor: National Lottery Charities Board.

Contract/grant sponsor: The Cancer Research Campaign.

Contract/grant sponsor: Oxigene Inc.

Abbreviations used: CA4P, combretastatin A4 phosphate; DCE-MRI, dynamic contrast-enhanced magnetic resonance imaging; DMXAA, 5,6-dimethylxanthenone acetic acid.

INTRODUCTION

The technique of dynamic contrast enhanced MRI (DCE-MRI) is becoming increasingly widespread, for improving the accuracy of diagnostic imaging,^{1–3} but also in research into aspects of tissue microcirculation^{4–7} and assessment of microvascular changes following treatment.^{8–13} The development of novel therapeutic agents in

the field of cancer medicine, such as those that act on the tumour vasculature, has increased the importance of standardizing this powerful tool so that accurate assessment of a drug's activity can be made early in its clinical development.¹⁴ Tofts *et al.* have recently set out a consensus of opinion about the standard quantities and symbols that should be adopted when using this technique to obtain fully quantitative parameters which relate to tissue perfusion, microvessel permeability and surface area.¹⁵ However, many groups still use semi-quantitative indices, such as the initial gradient of the signal intensity time curve in a T_1 -weighted series of images. Relative changes in such semi-quantitative parameters are indirectly related to changes in the physiological end-point of interest such as perfusion.^{3,12,16,17}

To date there has been very little published on the reproducibility of DCE-MRI between sessions on different days.¹⁸ This is clearly an important aspect particularly when assessing the magnitude of effects seen following treatment. This paper describes the reproducibility of this technique in 21 patients who had two examinations within a week with no interval treatment. Semi-quantitative parameters, gradient, AUC (area under the signal intensity–time curve for the first 90 s) and enhancement were compared with the quantitative parameters, K^{trans} ('transfer constant' for the transfer of contrast medium from the vessel into the extracellular extravascular space, EES), k_{ep} (the 'rate constant' for the transfer of contrast medium back from the EES into the vessel) and v_e (the contrast medium leakage space). In addition, parameter estimates calculated on a whole tumour region of interest (ROI) or by individual pixel analysis were compared.

METHODS

Patients and MR imaging

Patients participating in the Cancer Research Campaign (CRC) phase I trials of the vascular targeting agents dimethylxanthone acetic acid (DMXAA) and combretastatin A4 phosphate (CA4P) had serial DCE-MRI scans, including a pair of pre-treatment scans performed within 1 week of each other. All patients participating in this study gave written informed consent and the trial protocols were approved by our institutional committee on clinical research and ethics. The MRI studies were performed on a 1.5 T System, Magnetom Symphony (Siemens Medical Systems, Erlangen, Germany) using a body coil. At each scanning session, diagnostic images required for disease assessment were first obtained. A marker lesion (>2 cm in size) was chosen for the DCE-MRI. Care was taken when repositioning the patient in exactly the same position on the subsequent visit in order to obtain the same anatomical slice location. Three to five slices were chosen up to 8 cm apart with one slice

through the centre of the marker lesion and muscle. Proton density-weighted spoiled gradient echo FLASH (fast low angle shot) images ($TR = 350$ ms, $TE = 9.8$ ms, flip angle 20°) were then acquired at the same slice positions to enable the calculation of tissue gadopentatate dimeglumine (Gd-DTPA) concentration.¹⁹ A dynamic series of 30 T_1 -weighted FLASH images was then acquired for the same slice positions, with three images prior to a manual bolus intravenous injection of 0.1 mmol/kg Gd-DTPA, given over 10–12 s using a standardized injection protocol. Images were acquired consecutively with no time gaps. Each set of images took 11.9 s to acquire, and the whole sequence took 6 min. The imaging parameters for the T_1 -weighted FLASH sequence were TE 9–10 ms, TR 80 ms, 70° flip angle and 10 mm slice width. System gain factors were maintained between acquisition of the proton density and T_1 -weighted dynamic series of images.

Data analysis

Images were transferred to a Sun workstation (Sparc 10, Sun Microsystems, Mountain View, CA), and analysed using Analyze[™] software (Mayo Foundation, Rochester, MN). Using information from anatomic T_1 - or T_2 -weighted images and post-contrast T_1 images, ROIs were carefully drawn around the tumours, including the whole tumour where possible, but excluding pulsatility artefacts from blood vessels, and susceptibility artefacts from adjacent bowel. ROIs were also drawn for areas of skeletal muscle (usually paraspinous muscle). The same observer placed identical ROIs for each pair of images. The observer was not blinded to the placement of the ROI on the first image when positioning the ROI for the second image.

The dynamic image data were analysed using both quantitative and semi-quantitative approaches. Semi-quantitative analysis was performed directly with the T_1 -weighted images but quantitative analysis required conversion of the MR signal intensities to Gd-DTPA concentrations.¹⁹ This was performed by converting signal intensities to T_1 relaxation time values using the proton density images in conjunction with data from a calibration experiment that involved phantoms with known T_1 relaxation time values. Gadolinium concentration $C_t(t)$ was then inferred from the tissue T_1 using the equation

$$C_t(t) = (1/T_1(t) - 1/T_{1_0})/r_1$$

where T_{1_0} is the tissue T_1 without contrast and r_1 is the longitudinal relaxivity of protons *in vivo* due to Gd-DTPA (taken to be 4.5 L/S/mmol at 1.5 T).²⁰ Although there may be some error in the T_1 determination, if the proton density of the phantoms differs from that in the tissue, this would be consistent between the paired MRI

examinations. Measurements of T_1 values of tissues *in vivo* using this protocol are consistent with T_1 values at 1.5 T in the literature.²¹

The ROIs were subsequently applied to the dynamic image data in two ways. In the first method (ROI analysis), the mean of the image data within the ROI was calculated for each of the dynamic images, resulting in a single time-intensity data set. In the second method (pixel analysis), time-intensity data were obtained for each pixel within the ROI. Each individual time-intensity data set was separately analysed as described below and the results were presented as parametric images. Although the tumour tissue response was frequently heterogeneous, a single global value for the entire region was obtained by taking the median of all the individual pixel parameters. The median rather than the mean was used as the distributions of the parameters were skewed. In muscle, tissue enhancement analysis was only done on a whole ROI basis.

A standard compartmental model²² was used to describe the arterial influx of Gd-DTPA into EES and its venous efflux. Using this model, the time course of contrast agent concentration in tissue can be described by the following equation:

$$C_t(t) = K^{\text{trans}} \times [C_p(t) \otimes \exp(-k_{\text{ep}} \times t)]$$

$C_t(t)$ represents the tracer concentration in tissue at time t ; $C_p(t)$ represents the tracer concentration in arterial blood plasma at time t ; K^{trans} represents the transfer constant for transport from plasma to EES; k_{ep} represents the rate constant for transport from the EES to plasma; and \otimes denotes convolution. Following the work of Tofts and Kermode^{15,23} the input function $C_p(t)$ was approximated by a bi-exponential function that was scaled according to the administered dose of Gd-DTPA, D (mmol/kg body weight):

$$\begin{aligned} C_p(t) &= D[a_1 \exp(-m_1 t) + a_2 \exp(-m_2 t)] \\ a_1 &= 3.99 \text{ kg/l}, \quad a_2 = 4.78 \text{ kg/l}, \\ m_1 &= 0.144 \text{ min}^{-1} \quad \text{and} \quad m_2 = 0.0111 \text{ min}^{-1} \end{aligned}$$

This model was fitted to the dynamic MR concentration data using nonlinear least squares estimation to obtain values for the two free parameters K^{trans} and k_{ep} . A third parameter v_e , which was interpreted as the volume of EES per unit volume of tissue, or 'leakage space' was obtained by dividing K^{trans} by k_{ep} .

The time of arrival of the bolus was calculated from the concentration-time curve for the whole ROI. This was done using the following steps:

1. Straight lines joining the first 8 pairs of adjacent points were extrapolated back to the time axis (point 1 and point 2, point 2 and point 3, etc.).
2. The points at which the lines intersected the time axis were calculated.

3. Intercepts that arose from lines with negative gradients were discarded.
4. Intercepts that were greater than either of the time points that defined the line were discarded.
5. The greatest of the remaining intercepts was selected as the onset time. The resulting onset time was then applied to the separate analysis of all the individual pixels within the ROI. Different pixels within a large heterogeneous tumour may have slightly different onset times so using the mean for the whole tumour is an approximation, but is justified because individual pixel concentration-time data are noisy. Including the onset time as an additional parameter in the model increases the number of free parameters and would correspondingly increase the variability of the parameter estimates. An alternative was to estimate the average onset time by fitting the mean concentration-time curve to a model that included the time of bolus arrival as a parameter in the fit. This was rejected because large heterogeneous tumours sometimes resulted in onset times that were a few seconds away from the value expected by visual inspection of the data.

Goodness-of-fit criteria. Pixels that resulted in poor fits were excluded from subsequent analysis. The criteria for acceptance were:

1. The following goodness-of-fit parameter was calculated:

$$\frac{1}{N} \sum_{i=0}^{N-1} \frac{|C_i - M_i|}{M_i}$$

where C_i is the measured concentration data at the i th time point, M_i is the corresponding point on the fitted curve and N is the number of discrete measurements. This parameter reflected the average difference between the measured data and the fitted curve and pixels with values less than 0.5 were found to indicate an acceptable fit.

2. $0.0 < K^{\text{trans}} < 5.0$ (ml/ml/min)
3. $0.0 < v_e < 1.0$.

In a tissue with highly permeable blood vessels, the rate at which Gd-DTPA enters the EES is limited by the tissue perfusion, and in this situation K^{trans} is equivalent to the blood plasma flow per unit volume of tissue.¹⁵ In a tissue such as brain with a tight blood-brain barrier and high perfusion, the rate at which Gd-DTPA enters the EES is limited by vessel permeability, and in this situation, K^{trans} is equivalent to the permeability-surface area product. In malignant tumours, vessels are generally more permeable than normal tissues, but the permeability is heterogeneous across the tumour.^{24,25} Tumour K^{trans}

therefore reflects a combination of permeability–surface area product and tissue perfusion.

In addition to these quantitative parameters, the maximum gradient, G , of the signal intensity–time curve and the enhancement E were calculated:

$$E = \left[\frac{S_t - S_0}{S_0} \right]_{\max} \quad G = \left[\frac{d[(S_t - S_0)/S_0]}{dt} \right]_{\max}$$

where S_t is the signal intensity at time t , S_0 is the baseline signal intensity and t is time. The initial AUC was calculated as the area under the signal intensity time curve for the first 90 s after bolus arrival, t_0 .

Statistical analysis

For each patient, the difference between the measurements of a parameter at each scan, d , was calculated. The distribution of d was tested for normality using the Shapiro–Wilk test. To establish if the size of d was dependent on the parameter value, Kendall's tau for correlation of the absolute value of d against the mean value for the two scans of each parameter was calculated. Values of p obtained from the above statistical tests were adjusted for multiple comparisons using Bonferroni's method.²⁶ If Kendall's tau indicated a significant correlation of the absolute value with the mean then the data were logarithmically transformed (base 10). The Shapiro–Wilk test and Kendall's tau were recalculated on the transformed data to retest for normality, and dependence of the difference on the mean. Wilcoxon's signed ranks test was used to compare distributions of ordinal data between the paired examinations. The following statistical measures of reproducibility were then obtained from a one-way analysis of variance (ANOVA) on the original ordinal data or transformed data as appropriate:

1. The mean squared difference dsd was calculated from

$$dsd = \sqrt{\frac{\sum d^2}{n}}$$

This can be used to calculate the 95% confidence interval for change which might occur in a group of n patients from:

$$CI = \pm \frac{1.96 \times dsd}{\sqrt{n}}$$

Any change in a group of n greater than this value would be significant at the 5% level. For transformed data this confidence interval can be expressed as a percentage of the mean by:

$$\%CI = \frac{100 \times \text{antilog}(\log_{10} \text{mean} \pm CI)}{\text{mean}}$$

This will produce a confidence interval that is not symmetric about the mean.

2. The within-patient standard deviation (wSD) was obtained by taking the square root of the within-patient mean square value in the ANOVA table. This can also be derived from the dsd as

$$wSD = \frac{dsd}{\sqrt{2}}$$

3. The within patient coefficient of variation (wCV) was derived by dividing the wSD by the overall mean for each parameter. For transformed data the wCV was obtained by:

$$wCV = \text{antilog}(wSD) - 1$$

When this value is large (>0.5) then the estimation of wCV is unreliable.²⁷

4. The repeatability of a parameter, r , was calculated as 2.77 multiplied by wSD.²⁸ The difference between two measurements for the same subject will be less than this figure for 95% of pairs of observations. This can be expressed as both an absolute value on the original scale, appropriate for parameters where there is no correlation of d with individual means and on the log scale for parameters with such a correlation. For transformed data it can be expressed as a percentage of the mean by:

$$\%r = \frac{100 \times \text{antilog}(\log_{10} \text{mean} \pm r)}{\text{mean}}$$

This will again give an asymmetric 95% confidence interval around the mean.

5. The ratio of the between-patient variance to the within-patient variance was derived for each parameter and each method of analysis, and tested for a significant difference in these variances. A parameter with a large variance in the patient population tested, but a small variance within individual patients would have a high value of this ratio.

Statistical analysis was performed using JMP statistics software package (SAS Institute Inc., USA).

RESULTS

The patient characteristics, tumour type and size are given in Table 1. The range of tumour types was typical for patients in phase I trials. Twenty-one patients had two pre-treatment examinations within a week. Slice repositioning between the two examinations was inadequate for one patient. Three patients had unreliable injections of contrast agent on one of the two examinations due to fracturing of glass syringe containing Gd-DTPA during injection (one patient) or leakage of Gd-DTPA from injection site (two patients). One further patient was excluded because of a large number of pixels in the

Table 1. Patient characteristics

Patient age	Patient sex	Tumour type	Tumour site	Tumour size (cm)
55	M	Renal cell carcinoma	Liver	14 × 12
61	M	Renal cell carcinoma	Kidney	12 × 10
44	F	Renal cell carcinoma	Renal bed	5 × 5
60	F	Renal cell carcinoma	Kidney	13 × 13
42	M	Renal cell carcinoma	Para-aortic lymph node	4.5 × 3
70	F	Ovarian serous cystadenocarcinoma	Para-aortic node	6 × 9
57	F	Ovarian serous cystadenocarcinoma	Inguinal lymph node	6 × 5.5
50	F	Ovarian serous cystadenocarcinoma	Pelvis	13 × 11
41	F	Peritoneal carcinoma	Pelvis	6 × 6
58	F	Peritoneal carcinoma	Pelvis	3 × 2
56	F	Leiomyosarcoma	Adrenal	8 × 8.5
59	F	Leiomyosarcoma	Para-aortic lymph node	8 × 7
68	M	Leiomyosarcoma	Abdomen	21 × 18
63	F	Leiomyosarcoma	Pelvis	17 × 9
48	M	Leiomyosarcoma	Chest	18 × 16
69	F	Leiomyosarcoma	Pelvis	17 × 14
62	F	Leiomyosarcoma	Chest	17 × 19
54	F	Neuroendocrine thymic carcinoma	Anterior chest wall	11 × 6
48	F	Adrenocortical carcinoma	Liver	5 × 3.5
56	F	Colonic adenocarcinoma	Liver	8 × 5
46	F	Breast adenocarcinoma	Pelvis	9 × 7

tumour ROI were classed as modelling failures (20%). The shape of the Gd-DTPA concentration–time curve for these pixels suggested a significant vascular contribution. The estimation of parameters was therefore inaccurate, as the model used assumes no contribution of intravascular contrast. Thus, 16 patients formed our study cohort.

Figure 1 illustrates the mean values of tumour and muscle parameters from the two examinations for the individual patients, calculated using the mean of the median values obtained from pixel analysis in tumour and from the whole ROI analysis in muscle. The mean value of tumour K^{trans} for all patients calculated from the Gd-DTPA–time curve was 0.43 ml/ml/min (range 0.07–1.03). This is comparable with literature values for human tumours.^{7,29} The mean tumour k_{ep} was 1.31 ml/ml/min (range 0.44–2.70). The mean v_e was 32.3 ml/ml (range 12.9–62.4%). These values are also in the published range for human tumours.

In Fig. 2 the absolute value of the difference between the scans (d) is plotted against the mean of the median value for the two scans for each parameter using the pixel analysis in tumour for each patient. The mean difference and the 95% confidence intervals for the mean difference are also shown, which ranged from $\pm 7\%$ for enhancement and v_e to -14 to $+16\%$ for K^{trans} . No parameter had a mean difference significantly different from zero. The distribution of d for each parameter was not significantly different from normal, and only K^{trans} had a significant dependence of d on the mean value (Kendall's tau, $p = 0.002$). After logarithmic transformation of K^{trans} values, there was no longer a significant dependence of d on the mean and Kendall's tau was not significant ($p = 0.72$). The repeatability limits are also shown.

The mean value for each parameter, the wSD, wCV and ratio of between-patient variance to within-patient variance calculated in tumours by pixel analysis are listed in Table 2(a), and the same indices for the tumour ROI analysis are given in Table 2(b). All parameters had significantly smaller within-patient variance than between-patient variance, and the value of this ratio, together with the wCV, enables comparison of the reproducibility of each parameter. Enhancement and v_e are the most reproducible parameters, and have the highest variance ratios, indicating relatively small variability within individual patients compared with the large variability in these parameters across the 16 patients in our cohort. When calculated either on a pixel-by-pixel or whole ROI basis, the wCVs were around 9% for enhancement and v_e in tumours. For K^{trans} , k_{ep} and gradient the results are more variable, with wCVs of 21–29%. AUC has intermediate reproducibility, with a wCV of 12%.

As there is no significant correlation of d with the initial parameter value except for K^{trans} , the percentage variation will be less for tumours with higher initial parameter values, and more for those with a low value. In these cases the statistical terms expressed as absolute values are more useful determinants of measurement variability, such as wSD or repeatability²⁷ and should be used for future reference rather than the percentage values which are also quoted to aid comparison between parameters for this study. In an individual patient changes greater than the repeatability value would be significant. For example, a change in v_e greater than 7.6% in an individual tumour would be statistically significant [Fig. 2(a)]. For K^{trans} the repeatability 0.26

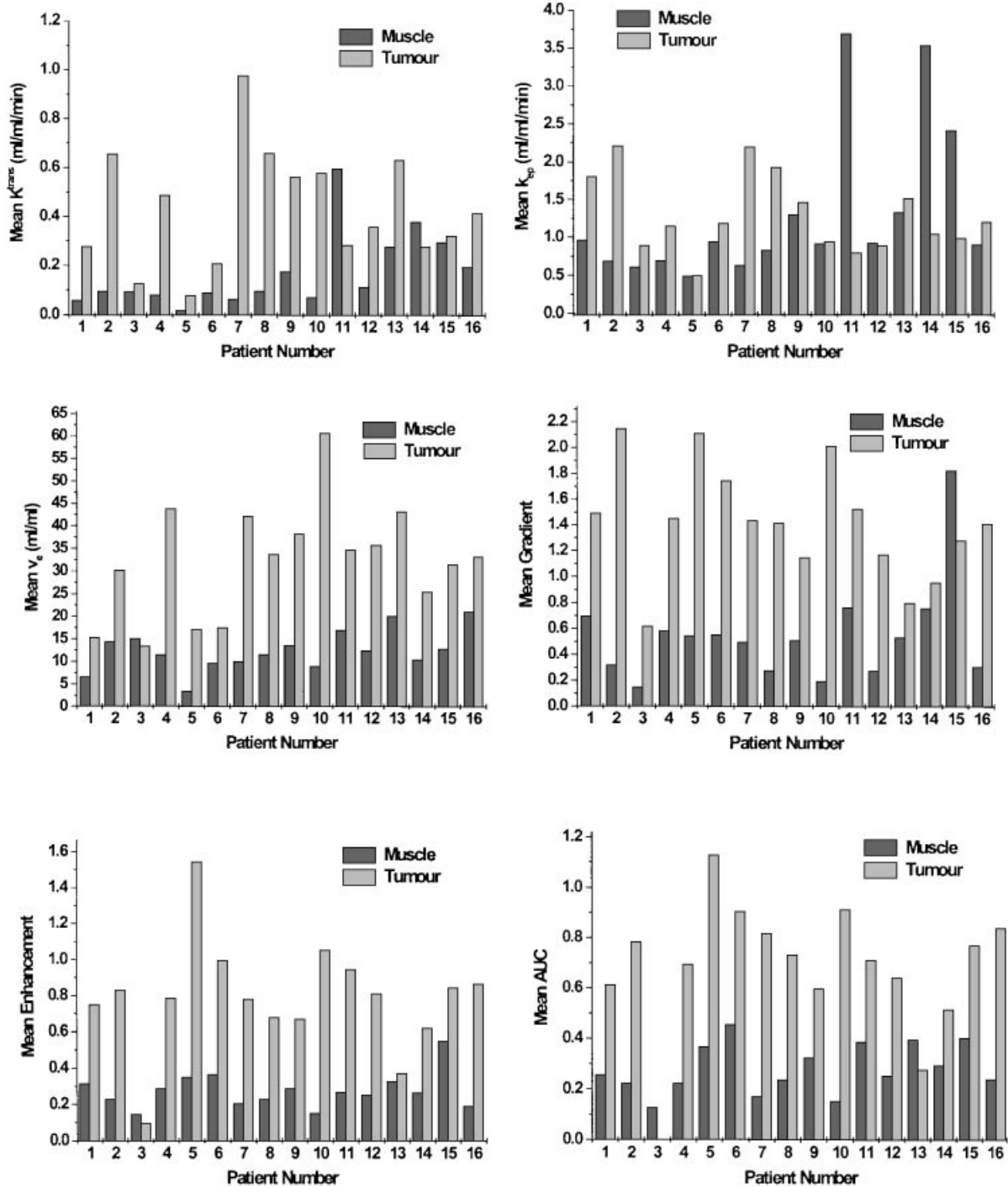


Figure 1. Mean parameter values in tumour and muscle for the paired examinations for each patient. For tumour ROIs, the mean of the median values from the ROI on the first and second examination was used

on the logarithmic scale can be expressed as a percentage of the mean (-45 to $+83\%$). Any reduction in an individual greater than 45% or increase greater than 83% would be statistically significant. The size of change that would be required for statistical significance in a group of 16 patients is given by the 95% CIs around the mean difference. For example the 95% CIs in tumours for K^{trans} , k_{ep} and v_e were -14 to $+16\%$, ± 0.2 ml/ml/min

(15%) and ± 1.9 ml/ml (6%), respectively (pixel analysis)

The results were similar for pixel and for whole ROI analysis, although the values for the statistical measures of reproducibility were generally slightly smaller for pixel analysis than for whole ROI analysis. For example the 95% CIs for K^{trans} are -14 to $+16\%$ for pixel analysis and -16 to $+19\%$ for whole ROI analysis. Calculation of

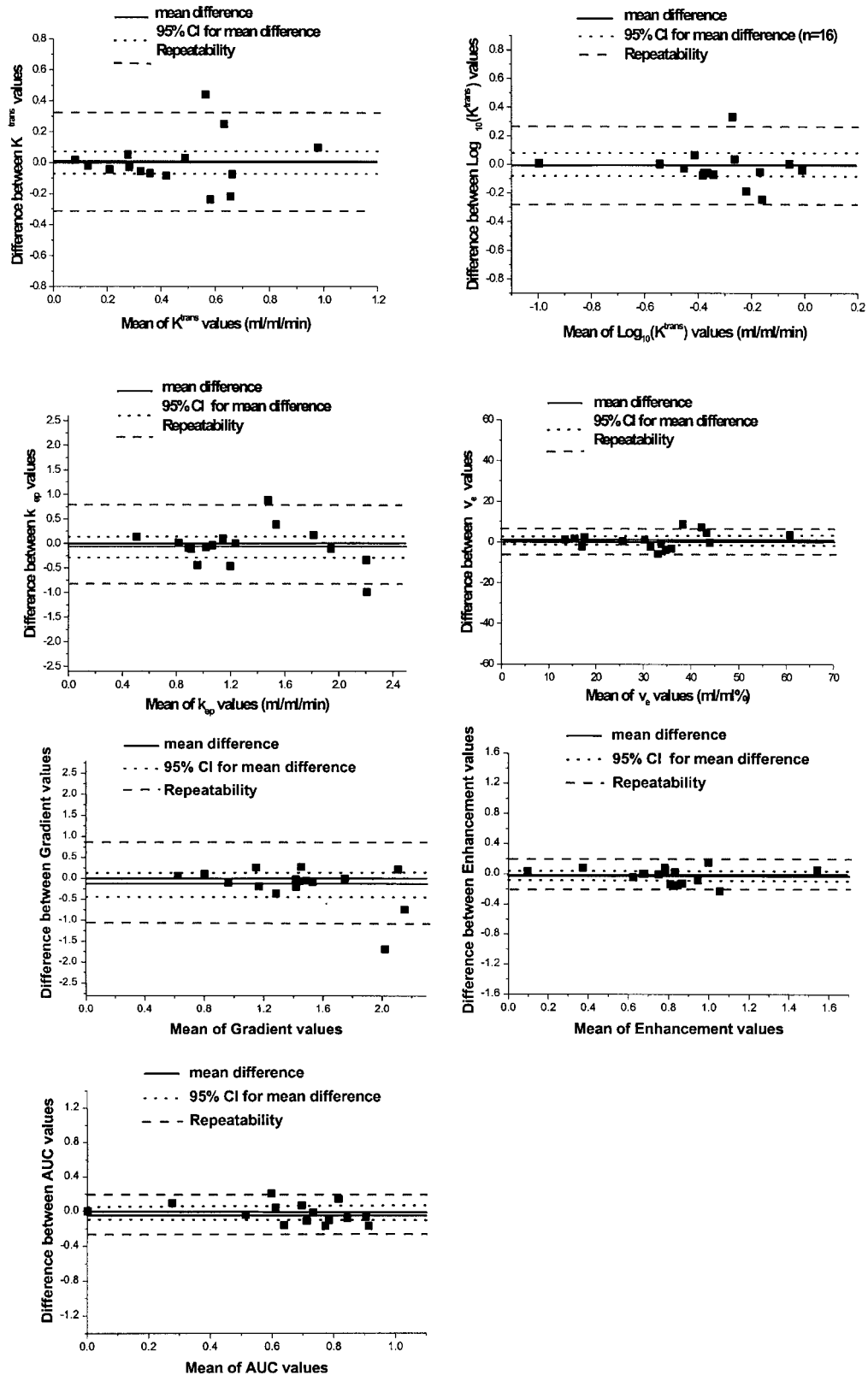


Figure 2. Difference between median parameter values plotted against mean of median parameter values for tumour in each patient (pixel-by-pixel analysis). To enable comparison across parameters, the y-axis scale was determined as \pm twice the overall mean parameter value [Table 2(a)]. The mean difference between scans is shown with the 95% confidence interval for a group of 16 (CI). The repeatability [$2.77 \times \text{wSD}$ (within-patient standard deviation)] is also shown and represents the level of change that would be significant in an individual. For K^{trans} the original untransformed data and the log-transformed data are shown

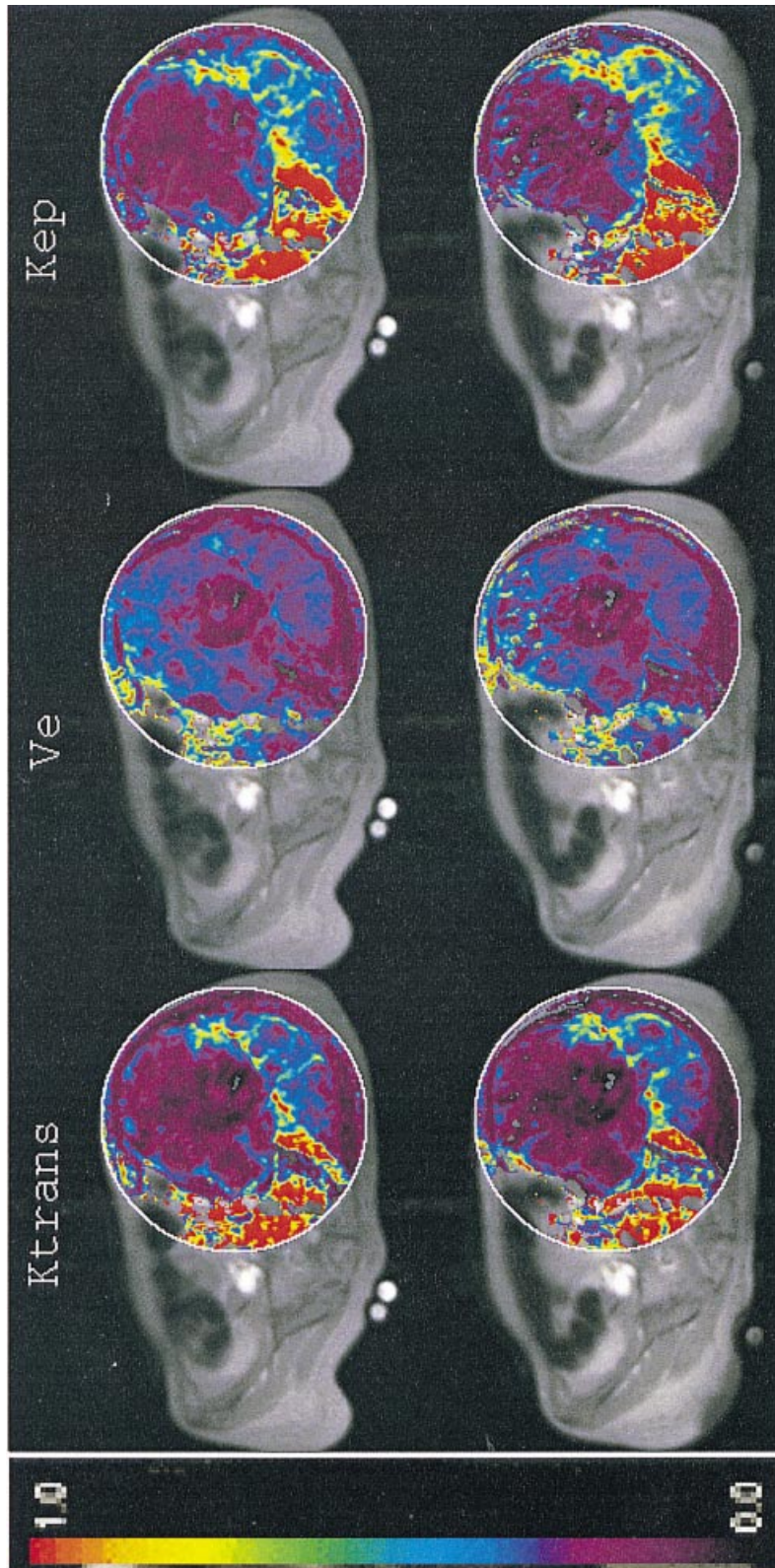


Plate 1. Paired pre-treatment scans from patient 15. These axial images through the pelvic area demonstrate a large leiomyosarcoma metastatic to the left iliac bone. A circular ROI is shown within which a K^{trans} parameter map is displayed. The colour scale is displayed on the left in ml/ml/min. These images demonstrate excellent spatial reproducibility within the tumour and surrounding normal tissues

Table 2. Calculated DCE-MRI kinetic parameters and their reproducibility for tumour: (a) pixel analysis; and (b) whole ROI analysis

Parameter	Mean	Mean difference	95% CI for mean difference	wSD	wCV	Repeatability	Variance ratio
(a) Pixel analysis							
K^{trans} (ml/ml/min)	0.43	0.00	-0.06 to +0.07 (-14 to +16%)	0.10 ⁺	24%	0.26 ^a (-45 to +83%)	20*
K_{ep} (ml/ml/min)	1.31	-0.06	±0.20 (15%)	0.28	21%	0.78	6.5*
v_e (ml/ml)	32.3	0.67	±1.9 (6%)	2.75	8.5%	7.6	42*
AUC (arbitrary units)	0.69	-0.02	±0.06 (8%)	0.08	12%	0.22	21*
Gradient (arbitrary units)	1.42	-0.16	±0.24 (17%)	0.35	24%	0.96	3.2*
Enhancement (arbitrary units)	0.79	-0.02	±0.05 (6%)	0.07	9%	0.20	38*
(b) Whole ROI analysis							
K^{trans} (ml/ml/min)	0.42	-0.03	-0.07 to 0.08 (-16 to +19%)	0.11 ⁺	29%	0.32 ^a (-50 to +100%)	31*
K_{ep} (ml/ml/min)	1.39	-0.16	±0.23 (16%)	0.33	24%	0.91	5*
v_e (ml/ml)	30.1	-0.02	±1.91 (6%)	2.75	9%	7.62	53*
AUC (arbitrary units)	0.918	-0.01	±0.08 (9%)	0.11	12%	0.31	21*
Gradient (arbitrary units)	1.25	-0.02	±0.21 (17%)	0.31	25%	0.85	3.5*
Enhancement (arbitrary units)	0.71	-0.01	±0.04 (6%)	0.06	9%	0.17	47*

CI, confidence interval; wSD, within-patient standard deviation; wCV, within-patient coefficient of variation (wSD/mean). Repeatability = $2.77 \times \text{wSD}$ and is the value of the 95% tolerance limit for the difference between two measurements on an individual. Variance ratio = ratio of between-patient variance to within-patient variance.

^a \log_{10} transformed data.

* $p < 0.05$.

individual pixel data, however, allowed the production of parametric maps, thereby demonstrating the spatial heterogeneity of the parameter across the tumour. Plate 1 shows parametric maps for K^{trans} overlaid on an anatomical MRI image of the same resolution in a patient with a large pelvic leiomyosarcoma. Visual assessment of the two images in Plate 1 reveals a close spatial reproduction of K^{trans} values, despite the marked intratumour heterogeneity.

Table 3 shows the reproducibility analysis for muscle ROIs. The mean values of all parameters are lower for muscle than for tumours, although two patients (patients 11 and 14) had a higher K^{trans} value in muscle than in tumour. For all parameters, the wCVs were higher for muscle than for tumour, although the wSDs were lower. In this tissue, the semi-quantitative parameters had

lower wCVs than the fully quantitative parameters. The estimate of wCV for muscle K^{trans} was unreliable. Again, enhancement was the most reproducible of the semi-quantitative parameters, and v_e was the most reproducible of the quantitative parameters. In muscle the size of change needed for significance in a group of 16 for K^{trans} , k_{ep} and v_e was -30 to +44%, ±0.81 ml/ml/min (61%) and ±1.7 ml/ml (13%). For gradient enhancement and AUC it was ±0.09 (20%), ±0.02 (8%) and ±0.03 (12%).

DISCUSSION

There are several statistical terms that can be used to quantitate measurement reproducibility, and to determine

Table 3. Calculated DCE-MRI kinetic parameters and their reproducibility for muscle: whole ROI analysis

Parameter	Mean	Mean difference	95% CI for mean difference	wSD	wCV	Repeatability	Variance ratio
K^{trans} (ml/ml/min)	0.15	0.07	-0.04 to 0.06 (-30 to +44%)	0.22 ⁺	NA	0.61 ^a (-75 to +308%)	5.7*
K_{ep} (ml/ml/min)	1.05	0.57	±0.81 (61%)	0.46	49%	1.28	5.6*
v_e (ml/ml)	12.7	-0.33	±1.67 (13%)	2.1	16%	5.71	10.6*
AUC (arbitrary units)	0.28	0.03	±0.03 (12%)	0.05	17%	0.13	7.2*
Gradient (arbitrary units)	0.44	0.02	±0.09 (20%)	0.13	28%	0.35	4.2*
Enhancement (arbitrary units)	0.26	0.02	±0.02 (8%)	0.03	12%	0.08	8.8*

CI, confidence interval; wSD, within-patient standard deviation; wCV, within-patient coefficient of variation. Repeatability = $2.77 \times \text{wSD}$ and is the value of the 95% limit for the difference between two measurements on an individual. Variance ratio = ratio of between-patient variance to within-patient variance.

^a \log_{10} transformed data.

* $p < 0.05$. NA, estimate of wCV from transformed data was unreliable.

measurement error for individuals or groups. A desirable parameter is one which is sensitive to wide variations in the characteristic to be studied where such variability exists in a population (large between-patient variance), but has a small variation in results obtained on different occasions in the same individual (small within patient variance), and will therefore have a large variance ratio. As shown in Fig. 1, the population of patients in this study did have a wide variation in initial DCE-MRI parameter values. The variance ratio is therefore useful for comparing reproducibility of different parameters. In this study, v_e and enhancement had the highest variance ratios in tumour and muscle. Gradient had the lowest variance ratio in both tissues. Other statistical terms are useful for describing the magnitude of spontaneous changes in a parameter. The 95% confidence interval for the mean difference in Tables 2 and 3 gives a measure of the spontaneous change to be expected in groups of a similar size to our cohort, and the formula in the statistical methods section allows calculation of such an interval for groups of different sizes. In individuals, the repeatability derived from the wSD represents the 95% confidence limit of the spontaneous change that might occur. If the absolute difference between examinations is greater for patients with larger parameter values, as for K^{trans} (determined by a significant value for Kendall's tau), then the repeatability is calculated on the logarithmic scale, or can be expressed as a percentage of the overall mean to give 95% limits for the relative change in an individual. For example, the parameter K^{trans} needs to decrease by more than 45% or increase more than 83% to be confident that this is due to treatment effect rather than spontaneous change or measurement error (Table 2). Alternatively, K^{trans} could be used in a group of 16 patients to detect mean reductions greater than 14% or increases greater than 16% and in practice, changes of smaller magnitude than this are unlikely to be of clinical significance. The leakage space v_e , or the volume of the extracellular extravascular space into which Gd-DTPA diffuses, is a more reproducible parameter, and can be used in individuals to detect changes greater than 7.6 ml/ml.

The semi-quantitative parameters, enhancement and initial AUC are reasonably reproducible, and can be used in individual patients to detect changes above 0.2 (25%) and 0.22 (33%), respectively, or mean changes greater than 7% in a group of this size. The gradient, like K^{trans} , is a more variable parameter, but in a group can be used to detect changes of more than 17%. The greater variability of gradient and K^{trans} may be because they are more sensitive to changes in the arterial input function, which will have day-to-day variation, and be affected by changes in cardiac output. Thus, using a complex pharmacokinetic model to produce fully quantitative parameters does not significantly alter the reproducibility of the technique. Where measurement of relative changes in an individual or group of patients is required, the

simpler semi-quantitative techniques are as reproducible. Nevertheless, the parameters from the pharmacokinetic model are more easily related to the physiological events in tissues, and allow comparisons between reports from different institutions. In addition, the reproducibility of a parameter is only one measure of its usefulness. It is also important to assess the sensitivity of the parameter to measuring real change in physiological events in the tissue for a given intervention. In another study we have compared the extensive changes in blood flow measured using iodoantipyrine (IAP) in rat tumours after treatment with a vascular targeting agent (combretastatin A4 phosphate) with changes in DCE-MRI parameters¹³ (and R. Maxwell, personal communication). We found that K^{trans} accurately mirrored blood flow changes, but gradient measurements markedly underestimated the size of the treatment effect, suggesting that K^{trans} is a more appropriate parameter for measuring response to this agent. This is reflected in the variance ratios for these parameters. Although the wCVs are the same (24%), the variance ratio is greater for K^{trans} than for gradient (20 vs 3.2), indicating a larger spread of K^{trans} values in this cohort of patients.

Analysing these parameters for each pixel rather than over the whole ROI does not dramatically affect the reproducibility, although generally the reproducibility was slightly better for pixel analysis. However, tumours are characteristically heterogeneous and averaging MRI kinetic parameters over the whole tumour removes valuable information about this heterogeneity. Pixel analysis enables tumour heterogeneity to be displayed as a parameter image, thus retaining the information about tumour heterogeneity without loss of reproducibility.

These results also demonstrate that the reproducibility is slightly worse for muscle than for tumours. The lower mean values of parameters in muscle explains this to some degree. Reductions in individuals greater than 75 and 79% for K^{trans} and gradient, 27 and 32% for v_e and enhancement would be significant. This tissue therefore provides a useful comparison with tumour for treatment effects, but there would need to be larger changes in MRI parameters in muscle before a significant change is observed.

Standardization of the Gd-DTPA injection by using a MR-compatible power injector and calculation of individual arterial input functions would be expected to significantly improve the reproducibility of these parameters in both tumour and muscle. Some variability will remain, due to actual changes in tissue blood flow, variations in magnet fields,³⁰⁻³² temperature and positioning of the patient. The use of the same observer to position ROIs on each pair of scans, without blinding the observer to the positioning of the ROI on the first examination, will give better reproducibility than if the second ROI positioning were completely independent of the first. The fact that four patients' data were rejected

due to problems with injection or slice positioning illustrates the learning curve associated with using dynamic MRI techniques. It is also important to realize that semi-quantitative reproducibility estimates calculated here cannot simply be applied from one centre to another, particularly when different equipment and sequences are used. This occurs because the baseline signal for any given tissue using a particular sequence will differ by the choice of imager used even if identical sequences are used. The quantification techniques used here aim to minimize the variabilities that occur due to the choice of imaging systems, magnetic field strengths, sequences and parameters and thus allow comparisons between and within patients and between different imaging centres. Furthermore, quantification techniques enable the derivation of kinetic parameters that are based on some understanding of physiological processes and so can provide insights into tumour biology.

Weber *et al.* published a similar assessment of reproducibility of metabolic measurements using FDG-PET.³³ In their paper, the 95% limits of agreement for an individual represented around 20% change in the parameters studied in tumours, which were mainly in the thorax. In the thorax, compared with regions such as the brain, the low background activity allows more accurate definition of tumour ROIs and therefore more reproducible results. The reproducibility of v_e and enhancement is similar to the PET parameters, whilst the K^{trans} and gradient are more variable. Studies assessing the variability of tumour volume determinations by CT have shown a mean coefficient of variation (CV) of 11% in repeat measurements of the volume of liver metastases,³⁴ and CVs between 16.5 and 113% for laryngeal tumours.³⁵ In brain tumours imaged by CT a change in volume of more than 20% was needed to be statistically significant.³⁶ These studies demonstrate that the reproducibility of DCE-MRI parameters compares reasonably with that of simple volume measurements. However, the good reproducibility seen in this study is dependent on several factors; lack of blinding of the observer for the second measurement, good scanning technique, care in patient positioning and reproducible injection technique. These results are not necessarily applicable to all centres, so similar assessments of reproducibility at individual centres using DCE-MRI would be recommended.

In conclusion, these results show that the DCE-MRI parameters v_e and enhancement have good reproducibility within individuals, and could be used to measure changes following treatment intervention. The parameters K^{trans} , k_{ep} and gradient are more variable but nevertheless can detect changes in tumours in a group of 16 patients of more than -14 to $+16\%$, ± 16 or $\pm 17\%$ respectively. If individual patient management decisions are being made on the results of a DCE-MRI scan, we would recommend using two pre-treatment scans to assess measurement variability before commenting on treatment effects.

Acknowledgements

We would like to thank the Cancer Research Campaign (CRC) and Oxigene Inc. who funded the DCE-MRI examinations for the patients in this study, Jane Boxall, Research Sister for her help with patient care and the patients involved in this study for their co-operation.

REFERENCES

- Hittmair K, Eckersberger F, Klepetko W, Helbich T, Herold CJ. Evaluation of solitary pulmonary nodules with dynamic contrast-enhanced MR imaging—a promising technique. *Magn. Reson. Imag.* 1995; **13**: 923–33.
- Postema S, Pattynama PM, Broker S, van der Geest RJ, van Rijswijk CS, Baptist Trimbos J. Fast dynamic contrast-enhanced colour-coded MRI in uterine cervix carcinoma: useful for tumour staging? *Clin. Radiol.* 1998; **53**: 729–34.
- Ostergaard M, Stoltenberg M, Lovgreen-Nielsen P, Volck B, Sonne-Holm S, Lorenzen I. Quantification of synovitis by MRI: correlation between dynamic and static gadolinium-enhanced magnetic resonance imaging and microscopic and macroscopic signs of synovial inflammation. *Magn. Reson. Imag.* 1998; **16**: 743–54.
- Hawighorst H, Knapstein PG, Weikel W, Knopp MV, Zuna I, Knof A, Brix G, Schaeffer U, Wilkens C, Schoenberg SO, Essig M, Vaupel P, van Kaick G. Angiogenesis of uterine cervical carcinoma: characterization by pharmacokinetic magnetic resonance parameters and histological microvessel density with correlation to lymphatic involvement. *Cancer Res.* 1997; **57**: 4777–86.
- Daldrup-Link HE, Link TM, Rummeny EJ, August C, Konemann S, Jurgens H, Heindel W. Assessing permeability alterations of the blood–bone marrow barrier due to total body irradiation: in vivo quantification with contrast enhanced magnetic resonance imaging. *Bone Marrow Transplant.* 2000; **25**: 71–8.
- Wen JG, Chen Y, Ringgaard S, Frokiaer J, Jorgensen TM, Stodkilde-Jorgensen H, Djurhuus JC. Evaluation of renal function in normal and hydronephrotic kidneys in rats using gadolinium diethylenetetramine-pentaacetic acid enhanced dynamic magnetic resonance imaging. *J. Urol.* 2000; **163**: 1264–70.
- Padhani A, Gapinski C, Macvicar D, Parker G, Suckling J, Revell P, Leach M, Dearnaley D, Husband J. Dynamic contrast enhanced MRI of prostate cancer: correlation with morphology and tumour stage, histological grade and PSA. *Clin. Radiol.* 2000; **55**: 99–109.
- Kennedy SD, Szczepaniak LS, Gibson SL, Hilf R, Foster TH, Bryant RG. Quantitative MRI of Gd-DTPA uptake in tumors: response to photodynamic therapy. *Magn. Reson. Med.* 1994; **31**: 292–301.
- Reddick W, Bhargava R, Taylor J, Meyer W, Fletcher B. Dynamic contrast-enhanced MR imaging evaluation of osteosarcoma response to neoadjuvant chemotherapy. *J. Magn. Reson. Imag.* 1995; **5**: 689–94.
- Baba Y, Furusawa M, Murakami R, Yokoyama T, Sakamoto Y, Nishimura R, Yamashita Y, Takahashi M, Ishikawa T. Role of dynamic MRI in the evaluation of head and neck cancers treated with radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.* 1997; **37**: 783–7.
- Hawighorst H, Engenhardt R, Knopp MV, Brix G, Grandy M, Essig M, Miltner P, Zuna I, Fuss M, van Kaick G. Intracranial meningiomas: time- and dose-dependent effects of irradiation on tumor microcirculation monitored by dynamic MR imaging. *Magn. Reson. Imag.* 1997; **15**: 423–32.
- Beauregard DA, Thelwall PE, Chaplin DJ, Hill SA, Adams GE, Brindle KM. Magnetic resonance imaging and spectroscopy of combretastatin A4 prodrug-induced disruption of tumour perfusion and energetic status. *Br. J. Cancer* 1998; **77**: 1761–7.
- Galbraith SM, Taylor NJ, Maxwell RJ, Lodge M, Tozer GM, Baddeley H, Wilson I, Prise VE, Rustin GJS. Combretastatin A4 Phosphate (CA4P) targets vasculature in animal and human tumours. *Br. J. Cancer* 2000; **83**(Suppl 1): 12.

14. Eckhardt S, Pluda J. Development of angiogenesis inhibitors for cancer therapy. *Invest. New Drugs* 1997; **15**: 1–3.
15. Tofts P, Brix G, Buckley D, Evelhoch J, Henderson E, Knopp M, Larsson H, Lee T, Mayr N, Parker G, Port R, Taylor J, Weisskoff R. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusible tracer: standardized quantities and symbols. *J. Magn. Reson. Imag.* 1999; **10**: 223–32.
16. Hawighorst H, Libicher M, Knopp MV, Moehler T, Kauffmann GW, Gv K. Evaluation of angiogenesis and perfusion of bone marrow lesions: role of semiquantitative and quantitative dynamic MRI. *J. Magn. Reson. Imag.* 1999; **10**: 286–94.
17. Bonnerot V, Charpentier A, Frouin F, Kalifa C, Vanel D, Di Paola R. Factor analysis of dynamic magnetic resonance imaging in predicting the response of osteosarcoma to chemotherapy. *Invest. Radiol.* 1992; **27**: 847–55.
18. Li K, Zhu X, Jayson G, Carrington B, Jones A, Lawrance J, Waterton J, Checkley D, Tessier J, Jackson A. Quantitative dynamic contrast-enhanced MRI in tumors. A reproducible technique in the head? A reproducible technique in the breast? *Proc. Intl Soc. Magn. Reson. Med.* 2000; **8**: 724.
19. Parker GJ, Suckling J, Tanner SF, Padhani AR, Revell PB, Husband JE, Leach MO. Probing tumor microvascularity by measurement, analysis and display of contrast agent uptake kinetics. *J. Magn. Reson. Imag.* 1997; **7**: 564–74.
20. Donahue KM, Burstein D, Manning WJ, Gray ML. Studies of Gd-DTPA relaxivity and proton exchange rates in tissue. *Magn. Reson. Med.* 1994; **32**: 66–76.
21. de Certaines JD, Henriksen O, Spisni A, Cortsen M, Ring PB. In vivo measurements of proton relaxation times in human brain, liver, and skeletal muscle: a multicenter MRI study. *Magn. Reson. Imag.* 1993; **11**: 841–50.
22. Kety S. Blood–tissue exchange methods. Theory of blood–tissue exchange and its application to measurement of blood flow. *Meth. Med. Res.* 1960; **8**: 223–227.
23. Tofts PS, Kermode AG. Measurement of the blood–brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts. *Magn. Reson. Med.* 1991; **17**: 357–67.
24. Deane BR, Greenwood J, Lantos PL, Pratt OE. The vasculature of experimental brain tumours. Part 4. The quantification of vascular permeability. *J. Neurol. Sci.* 1984; **65**: 59–68.
25. Jain R. Transport of molecules across tumor vasculature. *Cancer Metastasis Rev.* 1987; **6**: 559–93.
26. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *Br. Med. J.* 1995; **310**: 170.
27. Bland J, Altman D. Measurement error proportional to the mean. *Br. Med. J.* 1996; **313**: 106.
28. Bland J, Altman D. Measurement error. *Br. Med. J.* 1996; **313**: 744.
29. Hunter GJ, Hamberg LM, Choi N, Jain RK, McCloud T, Fischman AJ. Dynamic T1-weighted magnetic resonance imaging and positron emission tomography in patients with lung cancer: correlating vascular physiology with glucose metabolism. *Clin. Cancer Res.* 1998; **4**: 949–55.
30. Henriksen O, de Certaines JD, Spisni A, Cortsen M, Muller RN, Ring PB. In vivo field dependence of proton relaxation times in human brain, liver and skeletal muscle: a multicenter study. *Magn. Reson. Imag.* 1993; **11**: 851–6.
31. Brateman L, Jennings LW, Nunnally RL, Vaughan JT. Evaluations of magnetic resonance imaging parameters with simple phantoms. *Med. Phys.* 1986; **13**: 441–8.
32. Barker GJ, Simmons A, Arridge SR, Tofts PS. A simple method for investigating the effects of non-uniformity of radiofrequency transmission and radiofrequency reception in MRI. *Br. J. Radiol.* 1998; **71**: 59–67.
33. Weber WA, Ziegler SI, Thodtmann R, Hanauske AR, Schwaiger M. Reproducibility of metabolic measurements in malignant tumors using FDG PET. *J. Nucl. Med.* 1999; **40**: 1771–1777.
34. Van Hoe L, Van Cutsem E, Vergote I, Baert AL, Bellon E, Dupont P, Marchal G. Size quantification of liver metastases in patients undergoing cancer treatment: reproducibility of one-, two-, and three-dimensional measurements determined with spiral CT. *Radiology* 1997; **202**: 671–5.
35. Hermans R, Feron M, Bellon E, Dupont P, Van den Bogaert W, Baert AL. Laryngeal tumor volume measurements determined with CT: a study on intra- and interobserver variability. *Int. J. Radiat. Oncol. Biol. Phys.* 1998; **40**: 553–7.
36. Mahaley MSJ, Gillespie GY, Hammett R. Computerized tomography brain scan tumor volume determinations. Sensitivity as an objective criterion of response to therapy. *J. Neurosurg.* 1990; **72**: 872–8.