Tina Memo No. 2016-016
Internal, update of 2013-001 submitted to MRM, Green Open Access from 11/11/16.

Computing Similarity Measures on Multi-Dimensional DCE-MRI Variables for Hepatic Colorectal Metastases.

Scott V. Notley, Neil A. Thacker, Laura Horsley, Ross Little, Yvonne Watson, Saifee Mullamitha, Gordon C. Jayson, Alan Jackson.

Last updated
20 / 10 / 2016

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Computing Similarity Measures on Multi-Dimensional DCE-MRI Variables for Hepatic Colorectal Metastases.

Abstract

Purpose: Recent genetic profiling studies have highlighted inter- and intra-patient heterogeneity, however, knowledge is limited in liver metastases. DCE-MRI variables show a clear dependency between signal noise variance and mean value. Appropriate transformation of these variables allows robust fusion into a statistically meaningful similarity metric. The introduced methods are applied to asses inter and intra-subject variations for a repeated measures dataset consisting of DCE-MRI parameters measured on hepatic colorectal cancer.

Methods: A maximum-likelihood optimisation method is introduced to transform derived variables. Four DCE-MRI parameters were extracted from 102 selected tumour regions. A statistical distance metric was calculated for intra- and inter- subject tumour differences. Statistical errors for these quantities were estimated using bootstrap resampling

Results: Transformation of the variables to the homoscedastic space more closely matches the assumptions of standard summary statistics. Fusion into a single chi-squared statistic shows that inter subject variation in hepatic tumours is measurable and significantly greater than intra-subject variation.

Conclusion: Appropriate data transforms for DCE-MRI derived parameters ensure that interpretation of statistical measures are credible. They allow principled combination of data and produce an improvement in statistical sensitivity compared to conventional approaches. In this work they have improved the ability to assess biological differences between hepatic colorectal tumour metastases.

Key Words: Dynamic Contrast Enhanced Magnetic Resonance Imaging; Similarity Measures; Multi-Dimensional; Homoscedastic; Transforms

Introduction

Knowledge of the gene mutations that drive cancer has led to the development of a large number of mechanism-based therapeutics (MBT). However, there is a clear need to improve trial design to limit patient exposure to ineffective drugs and to accelerate the decision making for new agents. Imaging biomarkers are particularly attractive as they allow interrogation of the whole tumour, repeated measurements over time and support studies of inter- and intra-tumoural heterogeneity (1,2,3,4,5,6).

Dynamic contrast enhanced MRI (DCE-MRI) has been widely used in clinical trials of new agents and allows estimation of a number of parametric variables describing the tumour vascular microenvironment (2). Previous studies have described correlations between DCE-MRI characteristics and tumour grade and histological subtype (7,8). These are typically weak associations, inadequate for effective patient/tumour stratification even though statistical differences between tumour types may be seen in group comparison studies. There is a clear need to optimise the phenotypic information that we can extract from imaging data to improve the specificity and power of clinical trials using imaging biomarkers and, ideally, to provide sufficient statistical power to support personalised therapeutic decisions in individual patients. One common approach is the development of image acquisition and analysis strategies to improve the biological specificity and reliability of individual imaging derived parameters. Another is to ensure that the information content of this multi-parametric data is fully leveraged to support decision-making by the use of efficient statistical analytical approaches.

In this study we use DCE-MRI in hepatic metastatic colorectal cancer as a model for developing and testing optimised statistical approaches for the detection of phenotypic variation between patients and individual tumours. The liver is a preferential site for metastases in colorectal cancer, multiple lesions are common and metastatic disease is a common target in therapeutic trials of novel agents. Genetic heterogeneity has been described between primary tumours and metastases (9) and there is evidence that phenotypic and genetic heterogeneity also exists between metastatic lesions within the same patient (9). Goasguen and colleagues (10) reported significant variation in treatment response in 64% of tumour fragments derived from different metastases within a single patient. This was also associated with considerable inter-metastatic heterogeneity in levels of gene expression. This is supported by retrospective analysis of the CAIRO and CAIRO II trials (11) which demonstrated mixed response to therapy in 36% of patients with multiple metastases, associated with a decreased median survival of 23.7 months compared with 36 months in patients with homogeneous response.

We show that DCE-MRI derived parameters typically show dependency between the variance of the signal noise and the mean value of the variable. The use of non-linear transformations to generate variables with homogeneous noise has been described previously(12,13,14,15,16,17,18). This transforms the variables into a homoscedastic space where measurement error is independent of parameter distribution.These transformation methods are not widely recognised in medical data analysis where the common approach is to scale variables by their respective ranges or an estimated standard deviation based on the raw variables (19,20,21,22). Such an approach does not make any attempt to model the error characteristics of the individual parameters and does not facilitate a conventional statistical difference test. However, appropriate transformation of variables can aid with the estimation of summary statistics (17) and allows combination of non-commensurate variables and robust identification of outliers via calculation of chi-squares.

In this work we estimate the noise dependency from repeated measures using Bland-Altman plots (23). A maximum-likelihood based method is presented that optimises a function, transforming the noise on variables into a homoscedastic space. This allows a multi-dimensional approach, based upon the construction of a chi-squared statistic, supporting the fusion of information from multiple parameters. The method is evaluated on tumour data sets in the analysis of inter and intra-subject variation. Detecting such differences is more challenging than the more common task of detecting differences between normal tissue and pathology, but in line with the use of pharmacokinetic data for the assessment of tumour heterogeneity (4). The evaluation is restricted to varying degrees to demonstrate the general validity of the approach and avoids drawing false conclusions due to statistical biases in the data.

We hypothesise that: 1) the use of appropriate modelling of the parameter dependent characteristics of measurement error will allow transformation of parametric variables to more closely match the assumptions of standard statistical tests and: 2) that this will result in an increase in statistical power over the commonly used standard statistical analysis approaches.

Theory

Optimal Data Transforms

With repeatability data the sample noise process may be recovered by a simple subtraction of the matched measurement pairs. The dependency of the noise on the measurement value may be visualised by plotting the differences of the repeat measurements as a function of the average of the average i.e. Bland-Altman plots. We wish to be able to apply a general function to the data such that the resulting errors are homogeneous (independent of the variable value) and Gaussian. Formally we seek a general function $f(v_i;\theta)$, where v_i is the i-th sample of variable v and θ is a function parameter, such that $f(v_i;\theta) - f(r_i;\theta)$ may be approximated to a parameter independent Gaussian noise process; where r_i is the i-th repeat measurement.

The log-likelihood for a Gaussian distributed noise process is given by:

$$-\log L = \sum_{i=1}^{I} \Big[\frac{(f(v_i;\theta) - f(r_i;\theta))^2}{2\sigma_f^2} + \log(\sqrt{2\pi}\sigma_f) - \log\Big(\Big|\frac{\delta f(v_i;\theta)}{\delta I} + \frac{\delta f(r_i;\theta)}{\delta T}\Big|\Big) - \log(\Delta(v_i - r_i))\Big] \quad (1)$$

A full derivation of this equation is given in the appendix. This function may be minimised with respect to θ deriving the function $f(x)$ that best approximates homogeneous Gaussian noise. Although any continuous and differential function may be applied, for the following sections the function, $f(x)$, was defined as a power law of the form:

$$f(x) = x^{\theta} \quad (2)$$

We have not sought to find an analytic solution to this equation 1, given the small number of free parameters we have adopted a pragmatic approach to demonstrate the utility, with the minimum being found via an exhaustive search limited to a specified parameter range.

Methods

Patient Selection

The patients included in this study were undergoing imaging with Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) as a part of a clinical trial running at our institution and had given written informed consent to participate in the study. The study had ethical approval and was carried out in accordance with standards of GCP. Patients were eligible providing they were over eighteen years of age with biopsy confirmed metastatic colorectal cancer, without previous therapy for metastatic disease and disease measuring 3cm. All patients had undergone 2 baseline DCE-MRI scans, median 4 days (range 2-7 days) prior to treatment. It is this data which has been used in the current work to investigate biological variation within tumour tissues. To help control for tumour microenvironment, only patients with liver metastases were included in our analyses. This resulted in a sample of 29 subjects with numbers of metastases varying between 2 and 6, giving 102 tumours in total.

MRI Data Acquisition and Analysis

Data were acquired on a 1.5T Philips Intera system. The baseline T1 measurement consisted of 3 axial spoiled Fast Field Echo (gradient echo) volumes with flip angles 2, 10, 20 degrees, respectively and 4 signal averages. The dynamic series was acquired using the scanner whole body coil (Q body coil) for transmission and reception. The dynamic series consisted of 75 consecutively-acquired axial volumes with a flip angle of 20 degrees, 1 signal average, and a temporal resolution of 4.97 s. All studies maintained the same number of slices (25), field of view (375 mm 375 mm), matrix size (128 128), TR (4.0 ms), and TE (0.82 ms) for the baseline T1 measurement images and the dynamic series itself. Slice thickness was 4 mm for small target lesions or 8 mm for larger lesions, giving superior-inferior coverage of 100 mm or 200 mm, respectively. Gadoterate Meglumine (Dotarem) was injected intravenously (IV) by power injector at the time of the sixth dynamic acquisition at 0.2ml/kg, followed by a 20ml saline flush at a set rate of 3ml/sec. This was followed by acquisition of a post contrast T1-weighted image.

VOIs were delineated by an experienced radiographer on co-registered high resolution T1- and T2-weighted images. Whole TV was measured for each lesion. An arterial input function was measured where possible; in circumstances where this was not appropriate, a population derived input function was used (24). Analysis was performed using in-house software (Manchester Dynamic Modelling) and the extended Tofts and Kermode pharmacokinetic

model(25) was used to calculate the fractional volume of the extravascular extracellular space (v_e). The model free measurement, initial area under the gadolinium contrast curve at 60s (IAUC60) was calculated and voxels from tumour VOIs were included in the analysis if they demonstrated uptake of contrast, this was defined as an initial IAUC in the first 60 seconds (IAUC60)>0 mmol/s.

Median values of the measured parameters K^{trans}, v_p and v_e were extracted from distributions obtained from the 102 selected tumour regions. The enhancing fraction (E_{frac}) was also measured and then redefined as $E'_{frac} = 100(1 - E_{frac})$. For each tumour the process was repeated to generated a repeated measures dataset.

The data was first analysed using the standard method of scaling each variable by its standard deviation. A matrix of the measured variables, $\mathbf{V}$, is defined as:

$$\mathbf{V} = \begin{pmatrix} k^{trans}(1) & v_p(1) & v_e(1) & E'_{frac}(1) \\ \vdots & \vdots & \vdots & \vdots \\ k^{trans}(N) & v_p(N) & v_e(N) & E'_{frac}(N) \end{pmatrix} \tag{3}$$

where N is the number of tumours in the dataset ($N = 102$). A matrix $\mathbf{R}$ was similarly defined for the repeat measurements.

A distance, D_{stan}, between tumours was then defined as:

$$D^{j,k}_{stan} = \sum_{i=1}^{4} \frac{(\mathbf{V}_{j,i} - \mathbf{R}_{k,i})^2}{\sigma^2_{\mathbf{V}_{*,i}}} \tag{4}$$

where $D^{j,k}_{stan}$ is a distance between the j-th and the k-th tumours in the data set and $\sigma_{\mathbf{V}_{*,i}}$ is the standard deviation of the i-th variable (i-th column of $\mathbf{V}$) as measured.

Optimal data transforms, $f(\cdot)$, were estimated and applied to each raw variable to transform the data to the homoscedastic space. The *noise* level for the transformed data was then estimated by subtraction of the repeat measurements from the respective initial measurement and taking the standard deviation. A chi-squared variable for 4 degrees of freedom for the measured difference between tumours i and j was then defined as the sum of the squares of the difference between changes in each derived DCE MRI variable i divided by its reproducibility variance.

$$\chi^2_{j,k} = \sum_{i=1}^{4} \frac{(f_i(\mathbf{V}_{j,i}) - f_i(\mathbf{R}_{k,i}))^2}{\sigma^2_{\mathbf{n}_i}} \tag{5}$$

where $f_i(\cdot)$ is the transformation function for the i-th variable and $\sigma_{\mathbf{n}_i}$ is the standard deviation of the i-th *noise* signal, $\mathbf{n}_i$, a column vector defined as $\mathbf{n}_i = f_i(\mathbf{V}_{*,i}) - f_i(\mathbf{R}_{*,i})$, where $\mathbf{V}_{*,i}$ and $\mathbf{R}_{*,i}$ are the i-the columns of the matrices $\mathbf{V}$ and $\mathbf{R}$ respectively.

Although equations 4 and 5 look similar it is important to note that in equation 5 the denominator is an estimate of the level of *homogeneous noise*. The χ^2 distribution had Poisson like noise characteristics in that the variance of the estimate scales with the value. As discussed above, for statistical efficiency the variable needs to be transformed to a space where the distribution of the distance metric does not change with the value of the metric i.e. the homoscedastic space. In this case the correct transform that gives the homoscedastic space is given by the square root transform (as an approximation to the Anscombe transform). Thus, the quantity $D^{j,k}_{tran} = \sqrt{\chi^2_{j,k}/4}$ is expected to have a mean value of 1 for data j and k which differs only due to the presence of the modelled level of measurement error ($\sigma_{\mathbf{n}}$).

The distance measures D_{stan}, χ^2_{jk}, and D_{tran} were calculated for differences between tumours of each subject, and also for differences between tumours from different subjects. Differences were always taken to the measurement from the alternative baseline study so that the estimated reproducibility was correctly incorporated. Statistical errors for these quantities were estimated using bootstrap resampling with replacement. For each resampled Monté-Carlo dataset, the scaling parameters were re-estimated. The statistical distance between group means was then calculated as:

$$D_{stat} = \sqrt{\frac{(\mu_{inter} - \mu_{intra})^2}{\sigma^2_{inter} + \sigma^2_{intra}}} \tag{6}$$

where μ_{inter} and μ_{intra} are the inter and intra group means respectively; σ_{inter} and σ_{intra} are the respective group standard deviations. For a large number of samples, this statistical distance may be interpreted as a z-score and

significance levels (p-values) were calculated from this using the standard integration of the error function.

With such a small dataset, consisting of only 29 subjects, bias may be introduced, especially in the inter-subject distances, due to combinatorial effects of some subjects having more tumours than others. To reduce this effect we defined a maximum number of tumours per subject. Distances were computed for a maximum number of tumours per subject ranging from 2 to 6. This approach was also used to gain insight into the performance gains made with smaller data sets.

Results

Figure 1 shows histograms and figure 2 shows Bland-Altman plots constructed from the repeat data show a parameter dependent measurement repeatability. As these repeat measurements were obtained from different scan sessions we can assume these estimates include all important aspects of the variation intrinsic to the process of measurement and consequently the accuracy with which we can quantify biological change.

Table 1 shows the results of the Monté-Carlo simulations and computation of statistical measures on the distance D_{stan}. The mean and standard deviation of the mean are shown for both intra- and inter-subject groups. The z-score and corresponding p-value for the distance between the two groups (equation 6) is also shown. It can be seen that for a small data set with a maximum number of tumours of 2 the standard method fails to find a significant difference between the groups. As the maximum number of tumours per patient is increased the statistical distance between the groups increases and becomes statistically significant.

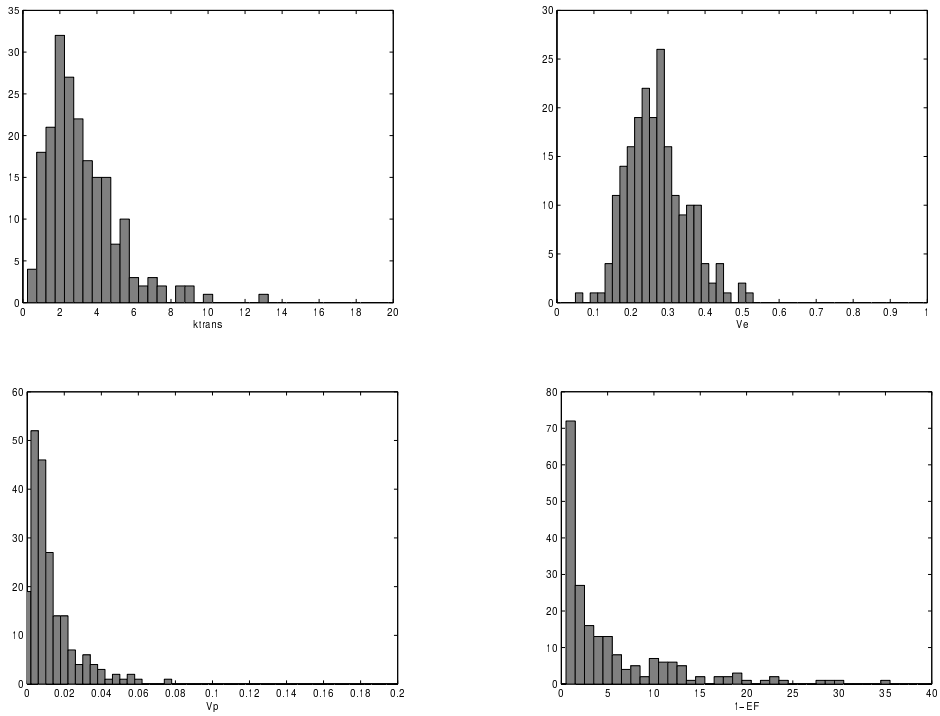

Figure 1: *The distribution of the median DCE MRI variables K_{trans}, V_e, V_p and $100(1 - E_{frac})$.*

Max Tumours	Dataset Size (intra/inter)	μ_{intra}	σ_{intra}	μ_{inter}	σ_{inter}	Total σ	Stat. Dist	p-value
2	58/59	1.54	0.738	2.96	0.488	0.885	1.6	0.0548
3	75/76	1.46	0.417	2.83	0.384	0.567	2.4	0.0082
4	85/86	1.34	0.312	2.69	0.342	0.450	3.0	0.0013
5	92/93	1.27	0.272	2.71	0.319	0.420	3.4	0.0003
6	96/97	1.17	0.225	2.64	0.294	0.370	3.9	<0.0001

Table 1: *Statistical measures made on distance D_{stan}.*

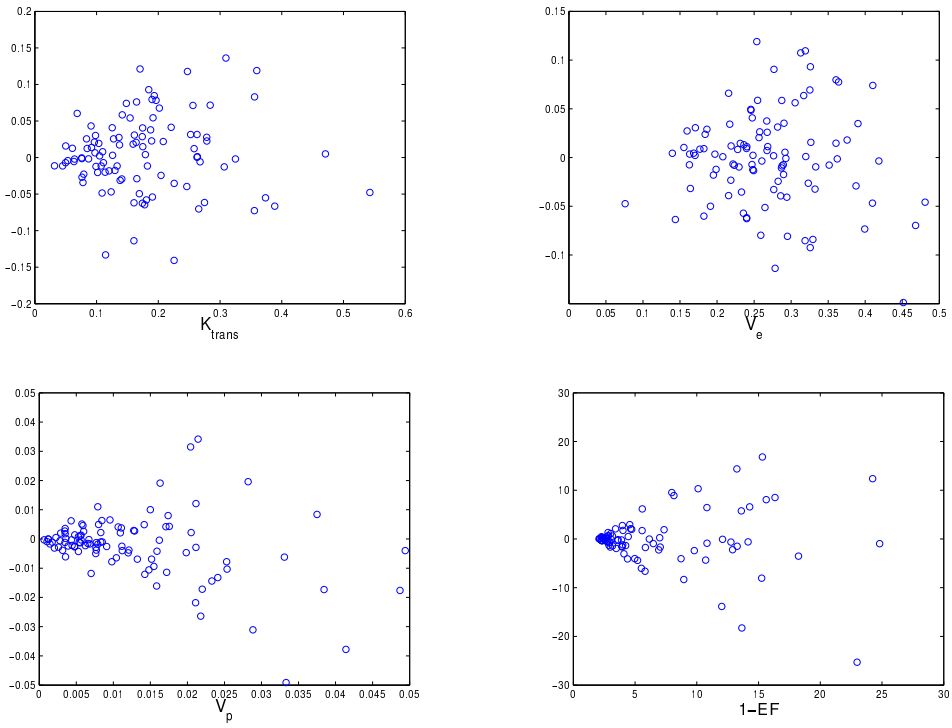

Figure 2: *Bland-Altman plots, showing reproducibility for median values of k_{trans}, V_p, V_e and E_{frac} (top left to bottom right). A clear parameter dependant accuracy is seen for all variables.*

The data transform method was applied to each measured variable and the optimal transform values shown in table 2 were found. Figure 3 show the histograms of the transformed variables and figure 4 shows the corresponding Bland-Altman plots. These figures show data distributions more closely conforming to a Gaussian and no evidence of any dependency of repeatability on parameter values.

	Transform Parameter A	
Variable	$(f(x) = x^A)$	Standard Deviation
k-trans	0.2	2.7
v_e	0.3	2.4
v_p	0.1	1.7
1-E_{frac}	0.6	1.9

Table 2: *Transform parameters for each variable*

The statistical distance, D_{tran}, constructed for differences between tumours of each subject, and also for differences between tumours from different subjects (Figure 5) gave mean values of 1.46 $\pm$ 0.14 and 2.16 $\pm$ 0.18. Differences were always taken to the measurement from the alternative baseline study so that the estimated reproducibility was correctly incorporated. Statistical errors for these quantities were estimated using bootstrap resampling with replacement. Both of these values are statistically significantly different to the null hypothesis.

Table 2 shows the standard deviation of each of the transformed variables. The distribution of the transformed and scaled measurements from all 29 subjects were found to have standard deviations between 1.7 and 2.7. As the expected value for pure noise is 1.0, this indicates that the signal pertaining to biological variation for each individual measurement is quite weak, and insufficient to allow effective separation of this tumour data.

Table 3 shows the results of the Monté-Carlo simulations and computation of statistical measures on the raw chi-squared distance χ^2 with transformed variables. In this case it may be seen that for a maximum of 2 and 3 tumours the significance of the result is increased over the raw variables; for a maximum of 2 the method is able to find a significant difference between groups. As the dataset is increased in size the results remain significant but are slightly reduced compared to the results found using D_{stan}.

Table 4 shows the results of the Monté-Carlo simulations and computation of statistical measures on the square-root transformed distance metric D_{tran}. In this case application of the correct transforms, to the raw data in the

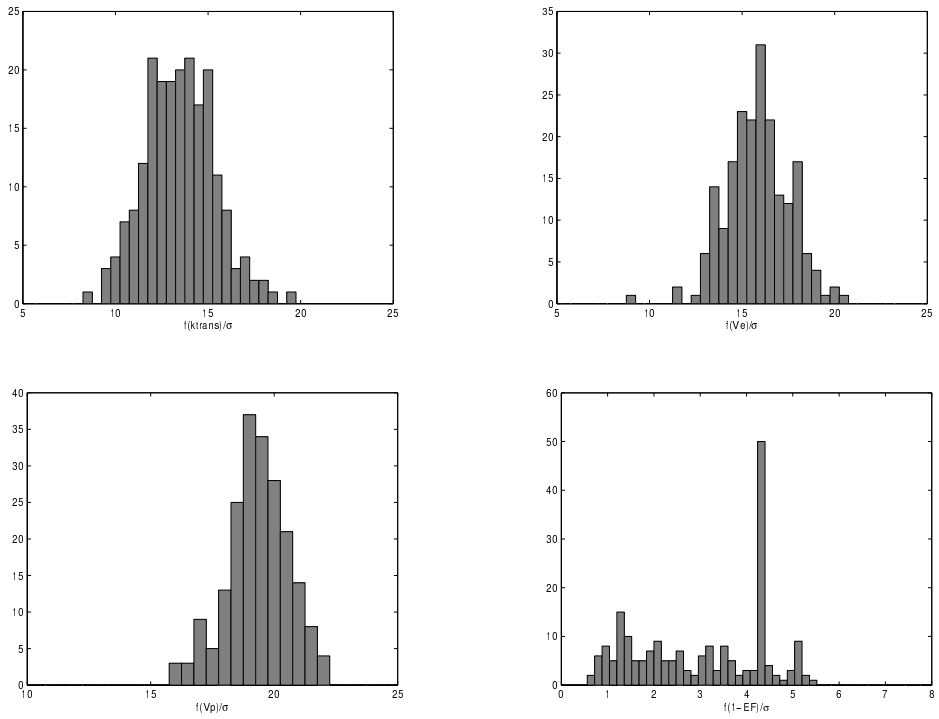

Figure 3: *Transformed variables scaled to reproducibility. Unlike the original parameter distributions (Figure 1), the spread of each variable (e.g. variance) is a measure of the associated information content. Gaussian random variables containing no signal (only noise) are expected to have a Gaussian distribution with unit variance, see Table 2.*

first instance and the square-root transform to the distance measure, results in the greatest statistical efficiency out of the three methods with significant results for all sizes of data sets.

Max Tumours	Dataset Size (intra/inter)	μ_{intra}	σ_{intra}	μ_{inter}	σ_{inter}	Total σ	Stat. Dist	p-value
2	58/59	2.51	0.696	6.38	0.960	1.858	3.3	0.0004
3	75/76	2.84	0.667	5.90	0.787	1.034	3.0	0.0013
4	85/86	2.71	0.530	5.29	0.738	0.920	2.8	0.0025
5	92/93	2.54	0.488	5.36	0.695	0.850	3.3	0.0004
6	96/97	2.54	0.461	5.39	0.721	0.856	3.4	0.0003

Table 3: *Statistical measures made on distance* χ^2.

Max Tumours	Dataset Size (intra/inter)	μ_{intra}	σ_{intra}	μ_{inter}	σ_{inter}	Total σ	Stat. Dist	p-value
2	58/59	1.41	0.193	2.27	0.173	0.258	3.3	0.0004
3	75/76	1.52	0.156	2.21	0.144	0.212	3.4	0.0003
4	85/86	1.49	0.131	2.10	0.141	0.177	3.1	0.0010
5	92/93	1.44	0.118	2.09	0.136	0.180	3.6	0.0001
6	96/97	1.43	0.112	2.11	0.127	0.170	4.0	<0.0001

Table 4: *Statistical measures made on distance* D_{tran}.

Figure 6 shows a comparison of the statistical distance found using the different distance measures as a function of the maximum number of tumours allowed in the dataset. Due the size of the Monté-Carlo datasets (1000 realisations), errors are small in comparison to the distance values and are not shown on this plot.

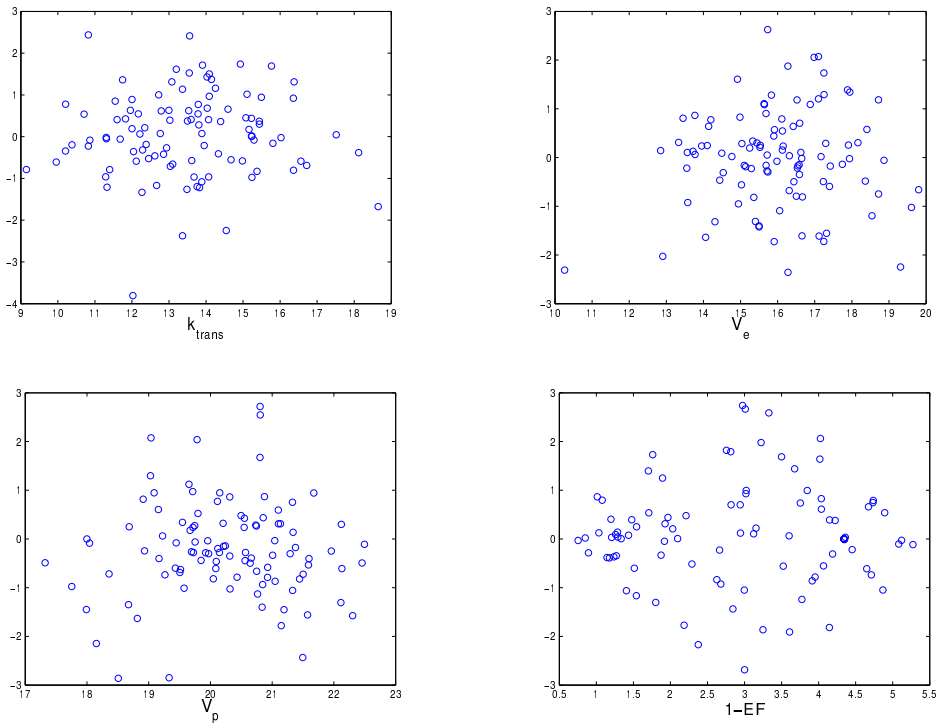

Figure 4: *Bland-Altman plots, showing reproducibility for transformed variables derived from k_{trans}, V_p, V_e and E_{frac} (top left to bottom right), scaled on the x axis to units of measured reproducibility σ. For a successful transformation the residual distributions (distribution of scatter above and below zero) should be independent of the variable. The high density value in the E_{frac} plot is due to the quantisation of this variable at 100% which causes identical values which cannot be separated by a transformation.*

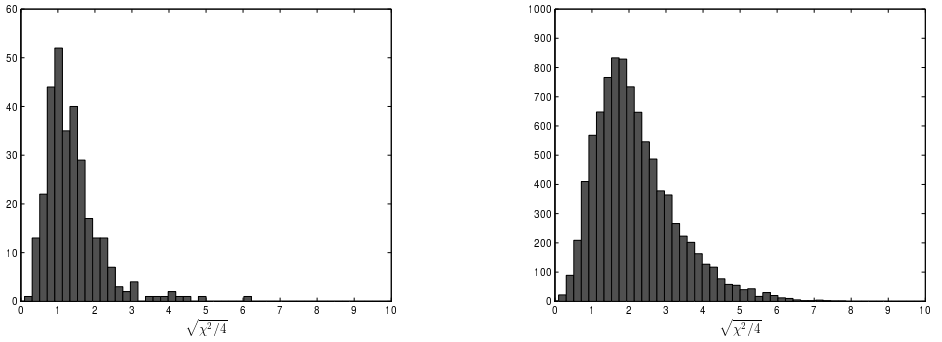

Figure 5: *The distributions of statistical distances, D_{tran}, for within subjects (left) and between subject (right) tumours. Computed using a chi squared statistic based on the transformed and scaled DCE MRI summary variables of median K trans,v_e,v_p and (1-E_{frac}).*

Discussion

The growing evidence that significant biological variation exists within and between metastatic deposits implies that heterogeneity of tumour response to different therapies might be observed if appropriate biomarkers can be developed. Repeated or multiple tissue biopsies are clearly impractical giving rise to an increasing need for alternate non-invasive approaches. Imaging biomarkers (IB) provide a potential solution offering unique advantages over soluble or tissue-based biomarkers. Ideally, IB could be used to identify biological / genetic variations to support enrichment of clinical trial data and provide predictive information to guide therapy. However there remain substantial technical problems associated with the use of IBs in this context. Identification of biological variability

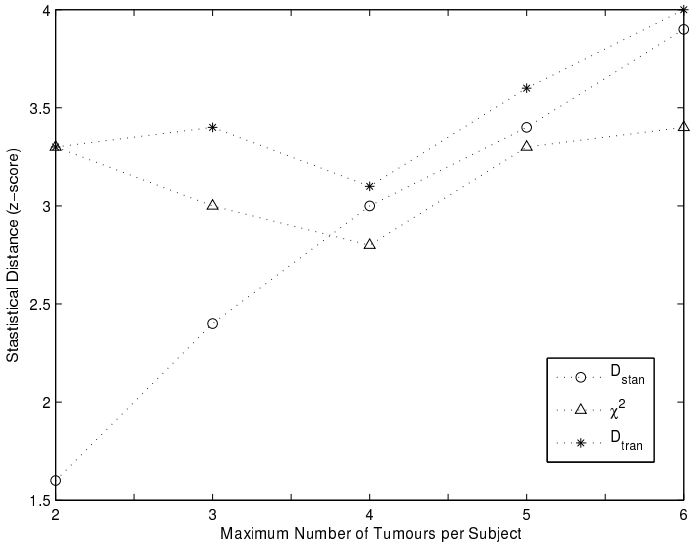

Figure 6:

within tumours requires the calculation of reliable and robust IBs from each voxel in the tumour. In practice such pixel-by-pixel mapping is associated with significant errors related to physiological movement and measurement bias induced by biological variation within the tissue. For example, the accuracy of measurements of blood volume varies systematically with the measured value (26) and the error models associated with many IB demonstrate similar but more complex behaviour (27).

In this work we have presented a statistically motivated approach to comparing multi-dimensional intra-subject tumour measurements to inter-subject measurements from DCE-MRI. The intra-subject results show a high level of similarity between tumours. This has implications when using multiple tumours from individual subjects to boost summary statistics that assume independence. Thus, in this work, results limited to 2 metastasis per subject are the most reliable. With this limit enforced the method is able to detect significant differences between the two groups showing more heterogeneity between subjects than within. A repeated measurements data set composed of 4 MR derived IB in hepatic liver colorectal metastases was collected. The measurements are non-commensurate in that they have differing scales and dynamic ranges. Further to this, Bland-Altman plots show that the variables have parameter dependent noise characteristics invalidating the simple use of combined data in their raw state.

For variables with known parametric distributions, variance stabilisation techniques to transform variables to a space where the noise characteristics are homogeneous have been used since the late 1940s (12). The work of Box and Cox (13) describes a method for optimally determining transforms for variables of arbitrary distribution. In this work, the variables under consideration have been sampled as a repeated measures data set and this give direct access to the noise distributions. A maximum-likelihood optimisation of a power law transform, based on the repeated measures noise, is presented.

A power law is determined that transforms variables to a space where the noise distribution is homogeneous. During optimisation of the power law, the variance of the fitted Gaussian is also maximally stabilised across the parameter range. The histograms of the transformed variables are shown to have become symmetrical and more Gaussian in nature; an indication that the noise is no longer parameter dependent. Crucially it can be seen from Bland-Altman plots that parameter dependency has been removed, allowing meaningful statistical measures (such as chi squares) to be computed.

Homogeneous noise more closely matches the assumptions made in statistical tests based on analysis of variance. Summary statistics, such as mean and variance, generated from the analysis are now a truer reflection of the properties found in the data, with data dependent biases minimised. Furthermore, by meeting these assumptions, the statistical efficiency of the tests is increased (13).

Scaling by the homogeneous noise level further transforms the variables to a space that is meaningful, as the variance of the variable is now related to information content. The standard deviation of the scaled distributions is an estimate of the signal-to-noise ratio since a scaled variable consisting of pure noise would be expected to have a standard deviation of 1. Thus, the transformation of the variables has allowed them to be robustly combined in

a single distance metric that operates over a Euclidean space. Any *significant* changes in this distance may now be more accurately attributed to measurable biological changes. It should be noted that, in this case, it is more meaningful to estimate a standard deviation since the variables are more consistent with a symmetrical Gaussian distribution. For the raw variable the distributions are heavily long tailed and skewed.

This approach may seem closely related to the common practice of scaling variables by the sample standard deviation or max/min range. However, the approach makes no attempt to understand the noise characteristics. In general, any statistics measured in this space, which make assumptions of homogeneous noise, will be unstable.

The work presented is motivated by the need for a statistical analysis that can better summarise what may be confidently inferred from the data. Following the methods outlined in this paper, improvements in statistical efficiency ensure we are extracting the optimal descriptive information from the available data. Evidence for this is shown in figure 6, and by comparison of tables 1, 3 and 4, which show that the square root transformed distances computed with the transformed variables show the greatest separation for all level of maximum tumour cut. Of note is the fact that, even with the careful statistical approach adopted in this work, the ability to detect homogeneity between groups requires a cohort. From the given dataset of 59 tumours, with an inter-subject distance of 3.3 standard deviations (table 4), we may infer that approximately 23 metastasis would be required to identify between group levels of heterogeneity within a single subject at the 95% confidence level.

In conclusion we have demonstrated the development of appropriate data transforms for DCE-MRI derived parameters. To ensure that interpretation of statistical measures are credible, transforms of this nature should be applied as standard procedure. The characteristics of the transformed variables allow principled combination of data from multiple IB to characterise individual tumour deposits and produces a significant improvement in statistical sensitivity when compared to conventional approaches. Using this approach we have demonstrated highly significant differences between metastases across subjects and are in accordance with recent studies describing heterogeneity in genetic makeup and therapeutic response.

Appendix: Log-Likelihood Derivation

For scientific validity the probability of a given event must be independent of arbitrary experimental choices. Thus, $P(v_i - r_i) = P(f(v_i;\theta) - f(r_i;\theta))$ for any continuous and differentiable monotonic function $f(x)$. This may be expressed in terms of probability densities as:

$$P(v_i - r_i) = p(v_i - r_i)\Delta(v_i - r_i) = p(f(v_i;\theta) - f(r_i;\theta))\Delta(f(v_i;\theta) - f(r_i;\theta)) \tag{7}$$

in the limit of $\Delta(v_i - r_i)$ tending to zero. Thus:

$$p(v_i - r_i) = p(f(v_i;\theta) - f(r_i;\theta))\frac{\delta(f(v_i;\theta) - f(r_i;\theta))}{\delta(v_i - r_i)}\Delta(v_i - r_i) \tag{8}$$

The Gaussian log-likelihood function, i.e. the likelihood of the original noise process being Gaussian, is given by:

$$-\log L = \sum_{i=1}^{I}\Big[\frac{(v_i - r_i)^2}{2\sigma^2} + \log(\sqrt{2\pi}\sigma) - \log(\Delta(v_i - r_i))\Big] \tag{9}$$

where $\sigma = \sigma_{v_i - r_i}$ (the standard deviation of the noise). Applying the function f results in:

$$-\log L = \sum_{i=1}^{I}\Big[\frac{(f(v_i;\theta) - f(r_i;\theta))^2}{2\sigma_f^2} + \log(\sqrt{2\pi}\sigma_f) - \log\Big(\Big|\frac{\delta(f(v_i;\theta) - f(r_i;\theta))}{\delta(v_i - r_i)}\Big|\Big) - \log(\Delta(v_i - r_i))\Big] \tag{10}$$

where $\sigma_f = \sigma_{f(v_i;\theta) - f(r_i;\theta)}$. Simplifying gives the final form of the Gaussian log-likelihood function for the transformed data as:

$$-\log L = \sum_{i=1}^{I}\Big[\frac{(f(v_i;\theta) - f(r_i;\theta))^2}{2\sigma_f^2} + \log(\sqrt{2\pi}\sigma_f) - \log\Big(\Big|\frac{\delta f(v_i;\theta)}{\delta I} + \frac{\delta f(r_i;\theta)}{\delta T}\Big|\Big) - \log(\Delta(v_i - r_i))\Big] \tag{11}$$

Acknowledgements

This work was funded by an investigator-led research grant from F. Hoffmann-LaRocheLtd, the Manchester Experimental Cancer Medicine Centre and is part of the work of the CRUK EPSRC Cancer Imaging Centre in Cambridge and Manchester (Grant C8742/A18097). The funding source had no part in the collection, analysis or the interpretation of data, in the writing or the decision to publish this manuscript.

The authors would also like to acknowledge the work of Dr Mark Saunders in the recruitment and referral of patients included in this study.

References

1. H. C. Canuto, M. I. Kettunen, A. S. Krishnan, A. A. Neves, M. de Backer, D-E. Hu, M. P. Hobson, and K. M. Brindle. Characterization of image heterogeneity using 2d minkowski functionals increases the sensitivity of detection of a targeted mri contrast agent. Magn. Reson. Med. 2009; 61(5):1218-1224.

2. C. J. Rose, S. Mills, J. P. B. OConnor, G. A. Buonaccorsi, C. Roberts, B. Whitcher Y. Watson, G. Jayson, A. Jackson, and G. J. M. Parker. Quantifying spatial heterogeneity in dynamic contrast-enhanced mri parameter maps. Magn. Reson. Med. 2009; 62(2):488-499.

3. R. Fisher, L. Pusztai, and C. Swanton. Cancer heterogeneity: implications for targeted ther- apeutics. Br. J. Cancer 2013; 108(3):479-485.

4. M-C Asselin, J. P. B. OConnor, R. Boellaard, N. A. Thacker, and A. Jackson. Quantifying Heterogeneity in Human Tumours using MRI and PET. European Journal of Cancer 2012; 48: 447-455.

5. K-L Li, L. J. Wilmes, R. G. Henry, M. G. Pallavicini, J. W. Park, D. D. Hu-Lowe, T. M. McShane, D. R. Shalinsky, Y-J Fu, R. C. Brasch, and N. M. Hylton. Heterogeneity in the angiogenic response of a bt474 human breast cancer to a novel vascular endothelial growth factor-receptor tyrosine kinase inhibitor: assessment by voxel analysis of dynamic contrast- enhanced mri. J. Magn. Reson. Imaging 2005; 22(4):511-519.

6. L. Alic, M. van Vliet, C. F. van Dijke, A. M. M. Eggermont, J. F. Veenland, and W. J. Niessen. Heterogeneity in dce-mri parametric maps: a biomarker for treatment response? Phys. Med. Biol. 2011; 56(6):1601-1616.

7. L. Martincich, V. Deantoni, I. Bertotto, S. Redana, F. Kubatzki, I. Sarotto, V. Rossi, M. Liotti, R. Ponzone, M. Aglietta, D. Regge, and F. Montemurro. Correlations between diffusion- weighted imaging and breast cancer biomarkers. Eur Radiol. 2012; 22(7):1519-1528.

8. J.H. Youk, E. J. Son, J. Chung, J. A. Kim, and E. K. Kim. Triple-negative invasive breast cancer on dynamic contrast-enhanced and diffusion-weighted mr imaging: comparison with other breast cancer subtypes. Eur Radiol. 2012; 22(8):1724-1734.

9. S. Jones, W-D Chen, G. Parmigiani, F. Diehl, N. Beerenwinkel, T. Antal, A. Traulsen, M. A. Nowak, C. Siegel, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, J. Willis, and S. D. Markowitz. Comparative lesion sequencing provides insights into tumor evolution. Proc. Natl. Acad. Sci. U.S.A. 2008; 105(11):4283-4289.

10. N. Goasguen, C. de Chaisemartin, A. Brouquet, C. Julie, G. P. Prevost, P. Laurent-Puig, and C. Penna. Evidence of heterogeneity within colorectal liver metastases for allelic losses, mrna level expression and in vitro response to chemotherapeutic agents. Int. J. Cancer 2009; 127(5): 1028-1037.

11. C.S. van Kessel, M. Samim, M. Koopman, M.A.A.J. van den Bosch, I.H.M. Borel Rinkes, C.J.A. Punt, and R. van Hillegersberg. Radiological heterogeneity in response to chemotherapy is associated with poor survival in patients with colorectal liver metastases. Eur. J. Cancer 2013; 49(11):2486-2493.

12. M. S. Bartlett. The use of transformations. Biometrics 1947; 3(1):39-52.

13. G. E. P. Box and D. R. Cox. An analysis of transformations. J. of The Royal Statistical Society. Series B (Methodological) 1964; 26(2):211-252.

14. N. R. Draper and D. R. Cox. On distributions and their transformation to normality. J. of The Royal Statistical Society. Series B (Methodological) 1969; 31(3):472-476.

15. H. M. Parsons, C. Ludwig, U. L. Günther, and M. R. Viant. Improved classification accuracy in 1- and 2-dimensional nmr metabolomics data using the variance stabilising generalised logarithm transform. BMC Bioinformatics 2007; 8:234-50.

16. A. Foi. Noise estimation and removal in mr imaging: The variance-stabilization approach. 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro 2011; pp.18091814.

17. J. M. Bland and D. G. Altman. Transforming data. BMJ 1996; 312:770.

18. J. M. Bland and D. G. Altman. Measurement error proportional to the mean. BMJ 1996; 313:106.

19. P. J. García-Laencina, J-L Sancho-Gómez, A. R. Figueiras-Vidal, and M. Verleysen. K nearest neighbours with mutual information for simulataneous classification and missing data impu- tation. Neurocomputing 2009; 72:1483-1493.

20. J. Gertheiss and G. Tutz. Feature selection and weighting by nearest neighbour ensembles. Chemometrics and Intelligent Laboratory Systems 2009; 99:30-38.

21. H. Altinçay. Improving the k-nearest neighbour rule:using geometrical neighbourhoods and manifold-based metrics. Expert Systems 2011; 28(4):391-405.

22. B. D. Ripley. Pattern Recognition and Neural Networks. Cambridge University Press 1996; ISBN 0 521 46086 7.

23. J. M. Bland and D. G. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet 1986;327:307-310.

24. G. J. Parker, A. Jackson, J. C. Waterton, and D. L. Buckley. Automated arterial input function extraction for t1-weighted dce-mri. in Proceedings of the 11th Annual Meeting of ISMRM 2003.

25. P. S. Tofts. Modeling tracer kinetics in dynamic gd-dtpa mr imaging. J. Magn. Reson. Imaging 1997; 7(1):91-101.

26. K. L. Li, X.P. Zhu, and A. Jackson. Parametric mapping of scaled fitting error in dynamic susceptibility contrast enhanced mr perfusion imaging. Br. J. Radiol 2000; 73(869):570-481.

27. K. L. Li and A. Jackson. New hybrid technique for accurate and reproducible quantitation of dynamic contrast-enhanced mri data. Magn. Reson. Med. 2004; 50(6):1286-1295.