

SCIENTIFIC REPORTS



OPEN

Considering tumour volume for motion corrected DWI of colorectal liver metastases increases sensitivity of ADC to detect treatment-induced changes

Ryan Pathak^{1,2}, Jingduo Tian¹, Neil A. Thacker¹, David M. Morris^{1,4}, Hossein Ragheb¹, Charles Saunders³, Mark Saunders² & Alan Jackson¹ 

ADC is a potential post treatment imaging biomarker in colorectal liver metastasis however measurements are affected by respiratory motion. This is compounded by increased statistical uncertainty in ADC measurement with decreasing tumour volume. In this prospective study we applied a retrospective motion correction method to improve the image quality of 15 tumour data sets from 11 patients. We compared repeatability of ADC measurements corrected for motion artefact against non-motion corrected acquisition of the same data set. We then applied an error model that estimated the uncertainty in ADC repeatability measurements therefore taking into consideration tumour volume. Test-retest differences in ADC for each tumour, was scaled to their estimated measurement uncertainty, and 95% confidence limits were calculated, with a null hypothesis that there is no difference between the model distribution and the data. An early post treatment scan (within 7 days of starting treatment) was acquired for 12 tumours from 8 patients. When accounting for both motion artefact and statistical uncertainty due to tumour volumes, the threshold for detecting significant post treatment changes for an individual tumour in this data set, reduced from 30.3% to 1.7% (95% limits of agreement). Applying these constraints, a significant change in ADC (5th and 20th percentiles of the ADC histogram) was observed in 5 patients post treatment. For smaller studies, motion correcting data for small tumour volumes increased statistical efficiency to detect post treatment changes in ADC. Lower percentiles may be more sensitive than mean ADC for colorectal metastases.

The apparent diffusion coefficient (ADC) is calculated from diffusion-weighted magnetic resonance imaging (DWI)¹. Selected tissue volumes are sensitised to free water diffusion using strong magnetic gradients so that signal is lost at a rate proportional to the rate of Brownian motion along the encoded direction which, in the range detected by conventional clinical DWI sequences, occurs primarily in the extravascular extracellular space (EES)². ADC will therefore be affected by the size and configuration of the EES but is also affected by other factors such as the local macromolecular environment, the presence of necrotic or cystic areas or fibrosis. In oncological applications increases in ADC in response to treatment has been taken to represent decrease in cell density or loss of the diffusion restriction by cell membranes as cells die or apoptose³. ADC offers a potential early response biomarker for clinical trials or personalised therapeutic regimes⁴, however measured changes must be reliably attributed to therapeutic response, not to measurement error or noise⁵⁻⁷. Accurate ADC measurements would increase confidence in post treatment response and could be combined with other potential markers e.g. lactate dehydrogenase enzyme (LDH) levels^{8,9}.

Previous studies in the liver have found post-treatment mean ADC changes in the range of 10 to 30%^{10,11}. It is important to avoid misinterpretation of post treatment mean ADC changes by calculating repeatability/

¹University of Manchester, Wolfson Molecular Imaging Centre, Manchester, UK. ²The Christie Hospital, NHS Foundation Trust, Manchester, UK. ³University of St Andrews, St Andrews, Scotland, UK. ⁴MRC Centre for Inflammation Research, University of Edinburgh, Edinburgh, UK. Correspondence and requests for materials should be addressed to R.P. (email: ryan.pathak@alumni.manchester.ac.uk)

| Acquisition parameters for protocol A and protocol B |
|---|
| B values of 100, 200, 400, 600 s/mm ² (3 orthogonal gradient directions) |
| 6 signal averages per image |
| TR 8000, TE 88 |
| Single shot echo-planar sequence (SS-EPI) |
| SENSE parallel imaging |
| Spectral attenuated inversion recovery (SPAIR) fat suppression |
| 5 mm axial slice thickness, 20 slices with no inter-slice gap |
| FOV of 384 × 384 |
| Bandwidth 1400–1800 Hz per pixel |
| Pixel size of 1.5 × 1.5 mm |
| Acquisition matrix 128 × 128 |
| Pixel size of 1.5 × 1.5 mm (therefore voxel volume 11.25 mm ³) |

Table 1. We acquired A and B data separately, however protocol A can be synthesized from the raw B data.

reproducibility^{12,13} and establishing either study-specific baseline thresholds or appropriate estimates where test-retest measurements are not possible¹⁴. Accurate estimation of other metrics from whole tumour 3D histograms could also increase observed post treatment changes. Examples where histogram analysis has been applied include the brain^{15,16}, peritoneum¹⁷ and the liver¹⁸. In a study of glioma, the 5th percentile was best for differentiating between high and low grade tumour¹⁶. The 25th percentile was most sensitive to post treatment ADC changes in peritoneal tumours¹⁷. In a study comparing whole liver ADC with and without colorectal metastatic tumours, the 5th percentile was significantly lower for the diseased group¹⁸.

ADC is calculated from multiple DWI acquisitions that assume perfect spatial registration between images therefore significant misregistration from motion will affect ADC accuracy. Respiratory triggering or use of navigator echo techniques can mitigate motion effects, improving image quality in terms of SNR, while maintaining stable ADC values, when compared to breath-hold sequences^{19,20} and free breathing acquisitions²¹. There is however conflicting evidence with other studies showing no advantage to navigator triggering²² with decrease in reproducibility and ADC stability compared to free breathing^{23,24}.

As shown in our previous work, uncertainty in the accurate estimation of ADC due to statistical measurement errors also adversely affects the ability to reliably detect change²⁵. The smaller the sample and wider the distribution of ADC voxel values, the greater the statistical uncertainty around the mean ADC estimate. Instability in repeated measures from smaller volumes has been observed with increased coefficient of variance (CoV)¹⁴, and size dependent improved reproducibility with whole tumour volumes²⁶. The CoV for ADC is a group statistic (ratio of the standard deviation and mean ADC for the study group) that allows comparison of ADC reproducibility between studies and requires large cohorts to infer significant differences when comparing matched study groups, rather than for individual tumour ADC changes.

We hypothesised that correcting for motion artefact and accounting for statistical measurement error, factors previously identified that negatively affect accuracy in ADC, would increase our confidence in attributing a post treatment ADC change as due to biological differences rather than measurement error or noise.

Materials and Methods

Patients. This single site prospective study was compliant with and approved by the NHS Health Research Authority Research Ethics Committee, United Kingdom, following approval from local Research & Development administrations at The Christie Hospital NHS Trust, Manchester. Formal written informed consent was recorded for each volunteer that participated. Volunteers were recruited from the colorectal oncology clinic and imaged consecutively, as they presented.

Inclusion criteria included; Histological primary colorectal carcinoma, radiological liver metastasis (at least one, minimum volume 1 cm³), no ongoing treatment. Exclusion criteria were contraindications to MRI or ongoing treatment. Patients were scanned on two separate occasions within 1–7 days prior to any new treatment commencing. Patients, who were medically fit and did not withdraw consent, were re-scanned within 1 week after commencing chemotherapy (1st or 2nd line regime).

Image acquisition. Images were acquired on a Philips Achieva 1.5 Tesla scanner. DWI (twice refocused diffusion-encoding scheme) were acquired in two slightly different ways on the same patients, each method labeled A and B (See Table 1 for acquisition parameters). The differences between methods are outlined as follows; for A, 18 images (6 repeats in 3 orthogonal directions) were acquired at each slice position and averaged by the scanner software to form the final axial slice image composite (1 of 20 slices). In B, for each slice, 6 repeated acquisitions in 3 orthogonal directions were acquired individually and transferred to a standalone workstation for retrospective post processed motion correction. Although A data could be synthesized from the raw B data we chose to acquire A as a separate data set to avoid correlated errors in the analysis.

Motion correction. A detailed description of the motion correction method has been published open-access²⁷. In brief, a local-rigid alignment (LRA) method designed specifically for use with DWI data was used. A reference slice was chosen (b-100) and split into 4 quadrants with the quadrant containing the tumour

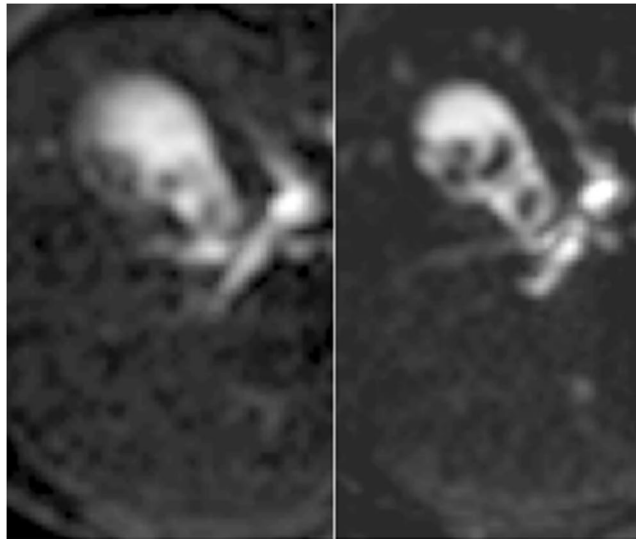


Figure 1. A comparison of part of the liver and gallbladder of a patient with b -100 mm^2/s^2 images acquired using standard protocol A (Left) versus acquired and post-processed motion corrected protocol B (Right). Post motion-correction, gallstones become clearly delineated within the high signal bile.

used to match adjacent slices. Limits in the degree of allowed movement were set to 3 pixels in x and y directions and 2 slices in the z directions (above and below), based on our observations of typical respiratory motion in liver data. There was little or no observable rotation (less than a pixel at the edge of an ROI). For each reference slice, 2 slices above and 2 below were used for matching. An existing matching algorithm was used for the reference slice against target slice based upon conventional statistics to optimise the cost function (Supplementary Information Appendix 1). The reference slice had 6 repeated acquisitions; therefore, in total there were 30 potential slices to match from when you include the 2 slices above and below for each repeat. The 6 most closely matched from the 30 potential slices were selected, and this process was repeated for each of the 3 orthogonal gradient directions. These 18 slices were then combined to form the composite image for that axial slice position. This was repeated for each axial slice acquired through the liver (20 slices). As the process required 2 slices above and below for each repeated acquisition in a given gradient direction, the top and bottom 2 slice positions could not be used for motion correction. An example of the effects of this method of LRA is given in Fig. 1 (note that the gallstones within the gallbladder become sharper and easier to delineate).

Image analysis and lesion definition. ADC values were estimated by mono-exponential fitting of 4 b -value images (100, 200, 400, 600 s/mm^2) corrected for high b -value SNR bias²⁸ (Supplementary Information Appendix 2). Manual whole tumour ROIs (largest and second largest where available, greater than 1 cm^3) were delineated from averaged b -100 image slices. The first and last slices through the tumour were excluded to minimise partial volume effects. The process was performed independently for A and B datasets. A manual delineation method was chosen as, in our experience, automated or semi-automated ROI selection methods are less robust in the liver, specifically due to physiological motion and low SNR (Fig. 2). In order to maximise the range of ROI sizes available for the error model (see below), single slice ROIs were also defined within the delineated 3D tumour volume.

Statistical error model for uncertainty in ADC estimation. A statistical measurement error model estimating ADC uncertainty has been described fully in a previous publication within this journal and is available open access²⁵. Briefly, the estimate of a mean or percentile to accurately describe a histogram is dependent on the sample size of the distribution (equivalent to tumour volume in this case). The wider the distribution the larger the error in accurately estimating a given metric and conversely, the larger the sample size and narrower the distribution width, the more precise the estimation will be. Motion artefact, SNR, tumour heterogeneity, and tumour boundary mismatches between test-retest volumes would also be expected to affect distribution width. All of these variables, as well as those we have not considered to affect the statistical accuracy, contribute to measurement error and are accounted for within the error model in terms of the ADC distribution width of an individual tumour.

The difference in ADC histograms between a single slice from a large tumour and small tumour are given as an example in Fig. 2. The larger ROI has a more bell-curve distribution, with a narrower proportional standard deviation, whereas the smaller tumour has a skewed distribution with a much larger proportional standard deviation.

The error model we have applied to this data was originally fitted to ADC calculated from quality assured test-retest tumour volumes (i.e. minimal visible motion artefact) using an error propagation method, to estimate measurement uncertainty for $\Delta\text{ADC}\%$ (the percentage change in ADC between baselines).

The suitability of the error model (S) to describe the distribution of this data set with (B) and without (A) motion correction was tested using Chi-squared (χ^2) goodness of fit (see below). Where the data was of sufficient

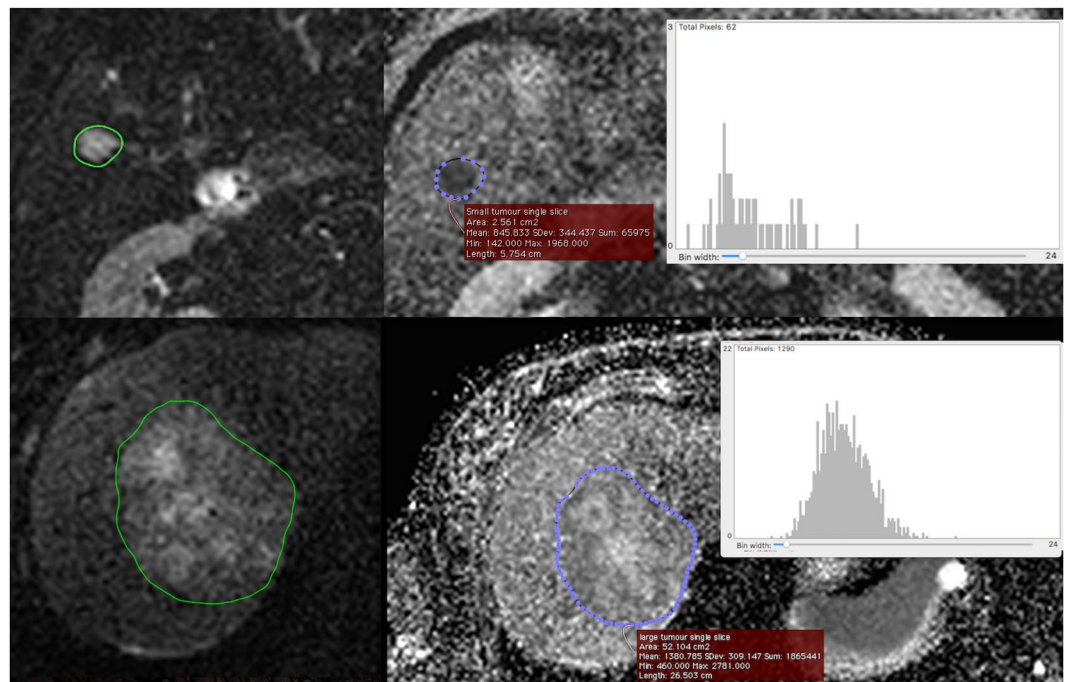


Figure 2. Tumour regions were delineated on b-100 mm²/s² diffusion images (left). Histogram analysis was then performed for the corresponding parametric ADC map (right) for whole tumour volumes (single slice represented here). A comparison of ADC distributions for a small tumour (top right) to a large tumour (bottom right) is given as an example. The Y-axis indicates frequency (number of voxels), and the X-axis indicates increasing ADC values ($\times 10^{-5}$ mm²/s). The mean ADC for that ROI and the standard deviation is given in the red box.

quality to describe the inverse relationship between tumour volume and statistical measurement uncertainty²⁵, the error model could be appropriately applied in order to standardise/scale individual tumours to their level of measurement uncertainty.

Sample size. Tumour response to chemotherapy agents that may have a variety of mechanisms of action can be heterogeneous due to micro-environmental or genetic factors, as well as geographical variations, i.e. spatial heterogeneity^{29–31}. In this study, in order to reach the target sample size of 15, each lesion was treated as an independent entity for repeatability analysis and in the assessment of early post treatment change.

Repeatability statistics. Histogram analysis of repeatability and Δ ADC% LoA for individual tumours, were calculated with a 5% level of significance. The group coefficient of variance (CoV) was also calculated (for A, B and S) to compare this data set with other published studies. CoV is calculated as the ratio (%) of the standard deviation of the group (difference between test and retest absolute ADC) and the average absolute ADC measurement (10^{-5} mm²/s). Tumour volumes were also compared at each visit (A vs. B for test and A vs. B for retest) and between visits (volume repeatability) in order to assess the stability of volume delineation. The definitions and formulas required for calculating 95% LoA and CoV can be found in the reference by Winfield *et al.*¹⁴. All repeatability statistical formulas and subsequent calculations were constructed with analysis performed using Microsoft Excel 2011 (OS X).

This study compared repeatability of Δ ADC% histogram metrics between a post-acquisition processing method of motion correction and a non-motion corrected acquisition for the same tumour data. Where the data was deemed suitable (Chi-squared (χ^2) goodness of fit) for application of an error model that takes into consideration tumour volume, individual tumour Δ ADC% was scaled to the estimated level of measurement uncertainty in the measurement. We have assigned this process and the outcomes as “method S” for “standardisation/scaling”.

The χ^2 goodness to fit tests the independence of two distributions, in this case the estimated uncertainty in individual tumour Δ ADC% values for this study data set (A and B) against the distribution of uncertainty for the original error model. If the two are found to be independent then the error model parameters cannot be used to standardise to the level of uncertainty in Δ ADC% measurements, implying that there are other factors dominating over tumour size (e.g. motion) and influencing the variability between test and retest measurements.

Observations of Δ ADC% as a marker for early response to treatment. The 95% LoA for Δ ADC% between test and retest acquisitions was used to determine the threshold that would be required for an observed post treatment response to reach statistical significance for this cohort. Post treatment ADC was compared to the retest rather than test ADC values in order to minimise any potential biological changes prior to treatment and determine whether any of the three approaches could observe a post treatment Δ ADC% response with statistical

| Lesion | Test volume | | Retest volume | | Volume repeatability | | Test ADC | | Retest ADC | |
|--------|-------------|-------|---------------|-------|----------------------|-----------------|----------|-------|--------------|--------------|
| | A | B | A | B | Δ VOL% A | Δ VOL% B | A | B | A | B |
| 1 | 19.8 | 21.8 | 21.8 | 22.6 | 1.9 | 0.8 | 111.9 | 102.6 | 105.5 | 105.6 |
| 2 | 88.3 | 84.3 | 86.4 | 86.3 | 1.9 | 2.0 | 189.9 | 189.0 | 181.3 | 182.0 |
| 3 | 105.1 | 104.6 | 114.7 | 112.7 | 9.6 | 8.0 | 115.5 | 114.4 | 114.3 | 111.0 |
| 4 | 4.4 | 5.1 | 6.1 | 6.0 | 1.7 | 0.9 | 95.7 | 120.2 | 105.3 | 113.8 |
| 5 | 13.6 | 13.2 | 15.1 | 13.6 | 1.5 | 0.4 | 202.2 | 187.4 | 173.7 | 193.5 |
| 6 | 53.7 | 47.5 | 52.2 | 43.5 | 1.5 | 3.9 | 128.0 | 122.1 | 125.3 | 119.8 |
| 7 | 7.1 | 7.4 | 8.0 | 7.7 | 0.9 | 0.3 | 107.1 | 107.3 | 114.8 | 113.0 |
| 8 | 3.5 | 4.0 | 2.6 | 3.5 | 0.9 | 0.5 | 108.9 | 108.2 | 106.0 | 108.1 |
| 9 | 90.2 | 86.3 | 88.3 | 83.4 | 1.9 | 2.9 | 122.6 | 115.9 | 110.0 | 107.0 |
| 10 | 1.6 | 1.2 | 1.1 | 1.5 | 0.4 | 0.3 | 67.9 | 78.9 | 94.0 | 78.3 |
| 11 | 115.7 | 114.6 | 124.4 | 123.5 | 8.7 | 8.9 | 130.0 | 120.1 | 125.3 | 122.9 |
| 12 | 8.7 | 8.5 | 10.0 | 8.3 | 1.3 | 0.2 | 116.5 | 109.7 | 95.7 | 97.0 |
| 13 | 7.9 | 8.4 | 4.6 | 7.8 | 3.3 | 0.6 | 121.9 | 123.5 | 79.8 | 121.1 |
| 14 | 5.4 | 5.0 | 5.4 | 4.8 | 0.0 | 0.2 | 120.2 | 124.0 | 127.9 | 124.9 |
| 15 | 9.3 | 8.9 | 9.5 | 8.6 | 0.2 | 0.3 | 126.4 | 114.2 | 122.4 | 116.0 |
| CoV | 4.5% | | 5.6% | | 7.2% | 6.8% | 5.6% | | 7.7% | |

Table 2. Lesion volumes and ADC values for protocol A and B. Volume delineation (cm^3) and calculated mean ADC (mm^2/s) was similar between standard (A) and motion corrected (B) methods for both the test and retest baseline acquisitions of the same tumour (CoV of 4.5% and 5.6% for volume, 5.6% and 7.7% for ADC). Volumes were repeatable for both A and B with 7.2% and 6.8% percentage change in volume between test and retest (Δ VOL%). Average mean ADC was $121 \times 10^{-5} \text{mm}^2/\text{s}$ for A and $122 \times 10^{-5} \text{mm}^2/\text{s}$ after motion correction. The bold values in the “Retest” column indicate the pre-treatment tumours used for post treatment response.

significance. The pre-treatment ADC values (re test) for patients with a post treatment acquisition are highlighted in bold (Table 3).

Observations of Δ ADC% against post treatment LDH trends. Serial serum LDH (U/L) was taken at two weekly intervals as part of routine care and used as a biomarker of disease response. The percentage change in LDH (Δ LDH%) from pre treatment to 3 months post treatment was calculated to observe any association with significant changes in ADC. A Pearson Correlation Coefficient was calculated to observe any significant correlation between Δ ADC% and Δ LDH%.

Results

11 patients (1 female) were recruited for test-retest imaging (July 2015 to May 2016). The average age was 68 (range 58 to 84). 8 of the 11 participants consented to, and were able to tolerate, early post treatment imaging. Chemotherapy regimens were either 1st or 2nd line treatments: combinations of oxaliplatin, irinotecan, fluorouracil, folinic acid, capecitabine, bevacizumab, panitumumab. The average time between acquisitions was four days. The average number of days between initiating chemotherapy and post treatment imaging was 5.5 days (range 2 to 11 days). The average number of days between the retest and post-treatment acquisition was 12 days (range 3 to 20 days).

Average mean ADC for 15 delineated tumour ROIs was $121 \times 10^{-5} \text{mm}^2/\text{s}$ (A) and $122 \times 10^{-5} \text{mm}^2/\text{s}$ (B) which is in line with previous measurements of colorectal liver metastases. Absolute values of mean ADC were consistent before and after motion correction. CoV between A and B was 5.6% (test group) and 7.7% (retest group) (Table 2). Retrospective motion correction did not therefore adversely affect absolute mean ADC values compared to the ROIs delineated from A. CoV for tumour volumes (test A vs. test B, and retest A vs. retest B) were low (<5.6%). Motion correction (B) therefore did not affect volume delineation when compared to A (Table 2).

Applying the statistical error model for estimation of measurement uncertainty. Figure 3 displays the distribution of statistical measurement uncertainty estimated for each Δ ADC% (all defined ROIs), comparing the standard method A and motion corrected method B from this single site data set to the original data set previously published that was used to fit the model (quality assured non-motion affected data). The overall shape of the distribution was consistent, showing an inverse relationship with the number of voxels within an ROI.

The suitability of the model to describe the present data set of non-motion corrected and motion corrected ROIs, was assessed using χ^2 distributions testing. Table 3 outlines the calculated the χ^2 statistic (critical chi square value) for each histogram metric. The distribution of uncertainty estimates for Δ ADC% measurements using A was different to the model distribution, for all histogram metrics. Eliminating the major contribution of motion (B), improved the fit for almost all metrics, with no difference between the distribution for uncertainty estimates of Δ ADC% measurements and the model. As expected, motion was the dominant process that contributes to poor repeatability.

After accounting for motion, the relationship between tumour size and statistical uncertainty in the estimate of Δ ADC% could be quantified more precisely. For the 95th percentile, B failed to sufficiently correct for motion, and therefore the model could not be applied to estimate statistical measurement uncertainty. In contrast, B successfully corrected for motion enough that the 5th and 20th percentiles (theoretically the population of voxels with

| Protocol | Histogram | CCV | DF | P-Value | Null hypothesis |
|----------|-----------------------------|-------|----|----------|-----------------|
| A | 5 th percentile | 82.86 | 14 | <0.00001 | Rejected |
| A | 20 th percentile | 69.51 | 14 | <0.00001 | Rejected |
| A | Median | 56.58 | 14 | <0.00001 | Rejected |
| A | Mean | 63.41 | 14 | <0.00001 | Rejected |
| A | 95 th percentile | 95.73 | 14 | <0.00001 | Rejected |
| B | 5 th percentile | 19.5 | 14 | 0.145 | Accepted |
| B | 20 th percentile | 14.36 | 14 | 0.423 | Accepted |
| B | Median | 13.59 | 14 | 0.481 | Accepted |
| B | Mean | 12.82 | 14 | 0.541 | Accepted |
| B | 95 th percentile | 25.39 | 14 | 0.031 | Rejected |

Table 3. Suitability of the statistical error model. The goodness of fit between the distribution of estimates of uncertainty for each $\Delta\text{ADC}\%$ measurement and the statistical error model was assessed using χ^2 distributions testing (non-motion corrected A vs. motion corrected B). If the null hypothesis of no significant difference between distributions was accepted, then $\Delta\text{ADC}\%$ for each tumour could be safely standardised for statistical measurement uncertainty. The critical χ^2 value (CCV) and degrees of freedom (DF) are displayed.

| Histogram | Method A | Method B | Method S |
|-----------------------------|----------|----------|----------|
| Mean | 30.3% | 8.7% | 1.7% |
| Median | 30.6% | 9.1% | 1.8% |
| 5 th percentile | 37.5% | 14.8% | 2.2% |
| 20 th percentile | 31.4% | 11.9% | 1.8% |
| 95 th percentile | 32.4% | 10.7% | — |

Table 4. Comparison of the 95% LoA for $\Delta\text{ADC}\%$ between methods. The 95% limits of agreement (LoA) are used to determine a statistically significant ($p < 0.05$) percentage change in ADC ($\Delta\text{ADC}\%$). For the 95th percentile, although motion correction improved the threshold for a significant change, the accuracy of any $\Delta\text{ADC}\%$ measurement could not be quantified, as the uncertainty model could not be applied (see Table 3).

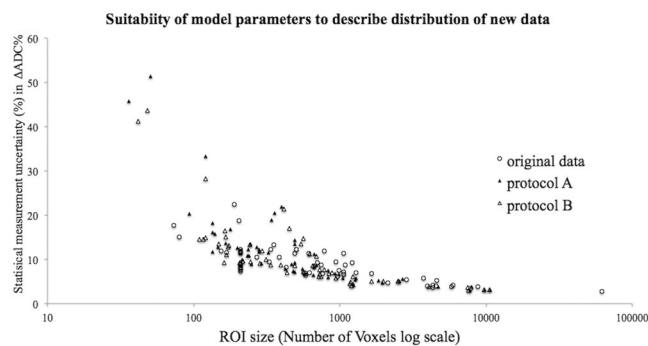


Figure 3. The distribution of uncertainty for both non-motion corrected A (solid triangle) and motion corrected B (triangle) follows a similar inverse relationship with ROI size as previously published quality assured data (no visible motion artefact) (circle), however A data is more scattered.

the highest diffusion restriction and tumour density) could be corrected for statistical measurement uncertainty using the error model (S).

Group CoV for A (test-retest) was 9.8% (median ADC), compared to 3.2% for B (test retest).

Observations of $\Delta\text{ADC}\%$ as a marker for early response to treatment. ADC histograms were defined for 12 tumours in 8 patients who underwent post treatment imaging. Comparing A, B and then applying the error model (S) for each histogram metric, the 95% LoA was used to determine a statistically significant threshold for changes in $\Delta\text{ADC}\%$ (Table 4). After correcting for motion and for statistical uncertainty in the estimated $\Delta\text{ADC}\%$ measurement, the lower ADC percentiles were the most sensitive to change, and these changes are highlighted for each post treatment tumour in Fig. 4. The 5th and 20th percentile $\Delta\text{ADC}\%$ was statistically significant in 6/12 tumours, compared to 5/12 for the median and 4/12 for the mean. No post treatment tumour $\Delta\text{ADC}\%$ value for any histogram metric was significantly less than the lower 95% LoA.

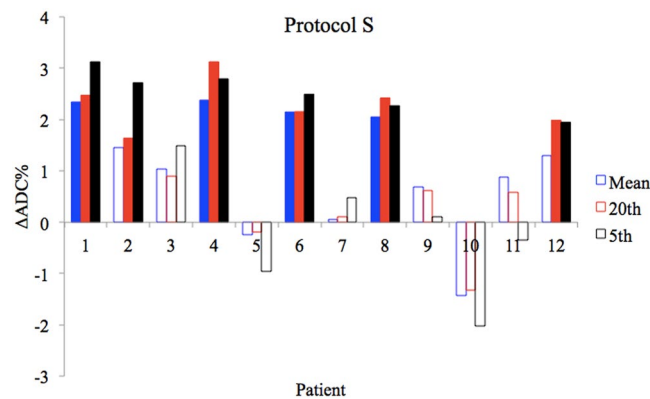


Figure 4. Tumours with statistically significant observed post treatment $\Delta\text{ADC}\%$ are highlighted (solid bars). For mean ADC significant $\Delta\text{ADC}\%$ was observed in 4/12 tumours when motion and statistical error were accounted for (protocol S). For the 20th and 5th percentiles significant $\Delta\text{ADC}\%$ was observed in 6/12 tumours. For the 5th percentile using significant $\Delta\text{ADC}\%$ was observed in 6/12 tumours.

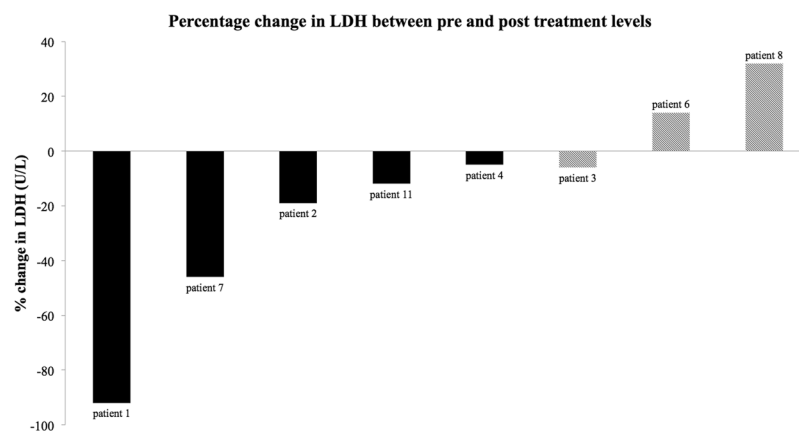


Figure 5. LDH (U/L) levels were collected as part of routine care before treatment started and then every 2 weeks thereafter. $\Delta\text{LDH}\%$ was calculated between the pre-treatment and 3 months post treatment level. The solid black bars indicate patients with tumours that demonstrated significant post treatment $\Delta\text{ADC}\%$ after accounting for motion and statistical measurement uncertainty (5th and 20th percentiles). The hatched bars indicate those tumours where there was no significant post treatment response observed for any percentile. Using method S for the 20th percentile data, a Pearson Correlation Coefficient of 0.65 was calculated, with a p value of 0.081. This was not statistically significant at the desired $p < 0.05$, however with a larger cohort, a correlation between $\Delta\text{LDH}\%$ and $\Delta\text{ADC}\%$ may be observed.

Observations of $\Delta\text{ADC}\%$ against post treatment LDH trends. Figure 5 is a comparison of $\Delta\text{LDH}\%$ (U/L) from pre-treatment levels to up to 3 months post treatment. Early $\Delta\text{ADC}\%$ was seen in 5/6 patients with reducing $\Delta\text{LDH}\%$ and 0/2 where $\Delta\text{LDH}\%$ rose. Using the 20th percentile as an example (6/12 tumours demonstrating a significant $\Delta\text{ADC}\%$) the Pearson Correlation Coefficient between protocol S $\Delta\text{ADC}\%$ and $\Delta\text{LDH}\%$ was 0.65 with a p value of 0.081, therefore not significant at the desired level ($p < 0.05$). Clearly these observations cannot be used to make any claims of correlation on such a small patient cohort, however this demonstrates the potential for combining other biological markers together with an accurate and sensitive $\Delta\text{ADC}\%$, to increase confidence of a post treatment response to therapy.

Discussion

The results of this study, comparing three alternative methods to detect post treatment changes in colorectal liver metastatic tumour ADC, demonstrates the importance of addressing misregistration caused by respiratory motion. Any method that successfully corrects reduced image quality resulting from motion should be capable of producing a similar improvement to those from this study. If motion correction strategies are not applied then strict quality assurance to exclude degraded images would be required, reducing the number of data sets that can be included in analysis. In this study, motion correction led to at least a 20% improvement of histogram metrics to detecting change (95% LoA) compared to the standard method. With this threshold, at least 4/12 tumours demonstrated early post treatment $\Delta\text{ADC}\%$, where no significant changes were seen previously using the standard approach (A).

In a recent publication of extracranial soft tissue ADC repeatability¹⁴ 141 lesions from 10 similar studies, were stratified into ROI volumes (smallest third, middle third and largest third), and it was observed that the “large” volume group had a statistically significant smaller CoV (less than 3%) compared to the other groups. This is attributed to increased sample size, reduced motion and less partial volume effect. Based on their findings, a conservative CoV of 6.5% is suggested as a threshold for future studies where repeatability of the group is not possible and tumour volumes are mixed. Using such a threshold would lead to misinterpretation of error as true biological change, for individual tumours, especially for smaller tumours. In small studies such as ours, a sample of 15 tumors does not provide the statistical power needed to observe a meaningful difference in precision (reciprocal variance) between methods A vs. B. This is partly so because many of the tumors are small, hence, yield noisy mean ADC measurements. 15 observations are insufficient to overcome sampling error of a second order statistic (ADC distribution variance).

Combining motion correction with an estimation of the level of statistical uncertainty in the accuracy of ADC estimates is directly inversely proportional to sample size/tumour volume²⁵. In this study, correcting for differences in uncertainty between tumours improved estimations of ADC repeatability to within 1.8% for 95% LoA (mean, median) (Table 4). The model cannot be used in isolation with standard protocol image acquisitions that have not been tightly quality assured, as it does not take into consideration motion effects. For this reason, a highly significant disagreement was observed between the model and all histogram metrics for A and 95th percentile for B. Only by combining both complimentary methods, the described repeatability was achievable.

Using the results for mean ADC for S as an example, significant $\Delta\text{ADC}\%$ was observed in 4/12 post treatment tumours. One of these was a different lesion to those identified by B. The tumour, which showed significant change with B, but not with S, was small and had a high uncertainty in $\Delta\text{ADC}\%$. After this uncertainty was corrected for, $\Delta\text{ADC}\%$ fell below the 95% LoA. Conversely a large tumour $\Delta\text{ADC}\%$ became statistically significant only after correcting, as there was less uncertainty in the accuracy of the measurement.

The two latter examples highlight the importance of accounting for uncertainty in the accuracy of a given measurement by scaling to statistical measurement error. Adding 150 extra voxels of data for the small tumour, would have pushed the $\Delta\text{ADC}\%$ into a statistically significant observation (for the same measured ADC and distribution width). Motion correction alone would have been insensitive to an observed $\Delta\text{ADC}\%$ for the large tumour if consideration were not given to the low level of uncertainty, i.e. increased confidence in the accuracy of the observed estimation.

When both approaches are combined (S) for the lower percentiles (20th and 5th), the number of tumours with a statistically significant observed $\Delta\text{ADC}\%$ increased from 4 to 6. Despite the relative increased 95% LoA (Table 4), these percentiles were the most sensitive to $\Delta\text{ADC}\%$ in the post treatment cohort (Fig. 4). As observed in other studies^{16–18}, theoretically this may be related to a larger shift in the ADC histogram at the lower percentiles as intra-tumoural regions with dense populations of malignant cells (higher diffusion restriction and therefore SNR) undergo death and necrosis after treatment.

Ideally, repeatability assessment within a shorter timeframe (24 hours) would have limited any variability due to biological disease progression. The clinical performance status of the recruited volunteers, given the palliative nature of disease, and timing with routine care, limited the flexibility for scanning. Variability in the timings for post treatment acquisitions may have affected the results for repeatability of $\Delta\text{ADC}\%$. An optimal evidence based time for post treatment scanning should be investigated.

ADC distribution width will be affected by precision of the delineation of tumour boundaries, which will impact test-retest repeatability. This error is quantified within the model. We acknowledge however, in the real-world clinical scenario of tumour response assessment, there will be an additional error from delineation of tumour boundaries when different observers (e.g. two different clinicians) assess pre and post treatment tumours separately.

Conclusion

In this single site study, we have demonstrated that combining retrospective motion correction with an estimation of statistical measurement uncertainty improves estimation of repeatability and increases the likelihood of observing a significant early post treatment response, with a $\Delta\text{ADC}\%$ threshold for significant change of 1.7% (mean ADC) in this cohort. Our post treatment $\Delta\text{ADC}\%$ results have shown however that the lower percentiles (20th and 5th) may be more sensitive to $\Delta\text{ADC}\%$ when using this combined method. As Supplementary Data we have provided a spreadsheet that can be expanded and populated with test retest or pre and post treatment ADC data. Provided the number of voxels, ADC and standard deviation of the histogram for each baseline is known, the statistical measurement uncertainty for $\Delta\text{ADC}\%$ is automatically calculated, together with a χ^2 goodness of fit that determines whether the model parameters are suitable to be used for a given data set.

Data Availability

The full dataset of ADC values for all defined ROIs and subsequent calculations are provided within a protected Microsoft Excel file (Supplementary Data). A table is provided (unprotected) that can be expanded and populated with test retest or pre and post treatment ADC data. Provided the number of voxels, ADC and standard deviation of the histogram for each baseline is known, the statistical measurement uncertainty for $\Delta\text{ADC}\%$ is automatically calculated, together with a χ^2 goodness of fit that determines whether the model parameters are suitable to be used for a given data set.

References

1. Le Bihan, D., Johansen-Berg, H. & Diffusion, M. R. I. at 25: exploring brain tissue structure and function. *NeuroImage* **61**, 324–341, <https://doi.org/10.1016/j.neuroimage.2011.11.006> (2012).
2. Le Bihan, D. & Lima, M. Diffusion Magnetic Resonance Imaging: What Water Tells Us about Biological Tissues. *PLoS Biology* **13**, e1002203, <https://doi.org/10.1371/journal.pbio.1002203> (2015).
3. Chenevert, T. L. *et al.* Diffusion Magnetic Resonance Imaging: an Early Surrogate Marker of Therapeutic Efficacy in Brain Tumours. *JNCI: Journal of the National Cancer Institute* **92**, 2029–2036, <https://doi.org/10.1093/jnci/92.24.2029> (2000).

4. Sinkus, R., Van Beers, B. E., Vilgrain, V., DeSouza, N. & Waterton, J. C. Apparent diffusion coefficient from magnetic resonance imaging as a biomarker in oncology drug development. *European journal of cancer* **48**, 425–431, <https://doi.org/10.1016/j.ejca.2011.11.034> (2012).
5. Deckers, F. *et al.* Apparent diffusion coefficient measurements as very early predictive markers of response to chemotherapy in hepatic metastasis: a preliminary investigation of reproducibility and diagnostic value. *Journal of magnetic resonance imaging: JMRI* **40**, 448–456, <https://doi.org/10.1002/jmri.24359> (2014).
6. deSouza, N. M. *et al.* Implementing diffusion-weighted MRI for body imaging in prospective multicentre trials: current considerations and future perspectives. *European radiology*, <https://doi.org/10.1007/s00330-017-4972-z> (2017).
7. Hoang, J. K. *et al.* Diffusion-weighted imaging for head and neck squamous cell carcinoma: quantifying repeatability to understand early treatment-induced change. *AJR. American journal of roentgenology* **203**, 1104–1108, <https://doi.org/10.2214/AJR.14.12838> (2014).
8. Li, G. *et al.* The prognostic value of lactate dehydrogenase levels in colorectal cancer: a meta-analysis. *BMC Cancer* **16**, 249, <https://doi.org/10.1186/s12885-016-2276-3> (2016).
9. Marmorino, F. *et al.* Serum LDH predicts benefit from bevacizumab beyond progression in metastatic colorectal cancer. *British Journal Of Cancer* **116**, 318, <https://doi.org/10.1038/bjc.2016.413> (2017).
10. Cui, Y., Zhang, X. P., Sun, Y. S., Tang, L. & Shen, L. Apparent diffusion coefficient: potential imaging biomarker for prediction and early detection of response to chemotherapy in hepatic metastases. *Radiology* **248**, 894–900, <https://doi.org/10.1148/radiol.2483071407> (2008).
11. Koh, D. M. *et al.* Predicting response of colorectal hepatic metastasis: value of pretreatment apparent diffusion coefficients. *AJR. American journal of roentgenology* **188**, 1001–1008, <https://doi.org/10.2214/AJR.06.0601> (2007).
12. Raunig, D. L. *et al.* Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment. *Statistical Methods in Medical Research* **24**, 27–67, <https://doi.org/10.1177/0962280214537344> (2015).
13. Sullivan, D. C. *et al.* Metrology Standards for Quantitative Imaging Biomarkers. *Radiology* **277**, 813–825, <https://doi.org/10.1148/radiol.2015142202> (2015).
14. Winfield, J. M. *et al.* Extracranial Soft-Tissue Tumours: Repeatability of Apparent Diffusion Coefficient Estimates from Diffusion-weighted MR Imaging. *Radiology* **284**, 88–99, <https://doi.org/10.1148/radiol.2017161965> (2017).
15. Pope, W. B. *et al.* Recurrent Glioblastoma Multiforme: ADC Histogram Analysis Predicts Response to Bevacizumab Treatment. *Radiology* **252**, 182–189, <https://doi.org/10.1148/radiol.2521081534> (2009).
16. Kang, Y. *et al.* Gliomas: Histogram analysis of apparent diffusion coefficient maps with standard- or high-b-value diffusion-weighted MR imaging—correlation with tumour grade. *Radiology* **261**, 882–890, <https://doi.org/10.1148/radiol.11110686> (2011).
17. Kyriazi, S. *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* **261**, 182–192, <https://doi.org/10.1148/radiol.11110577> (2011).
18. Lambregts, D. M. *et al.* Whole-liver diffusion-weighted MRI histogram analysis: effect of the presence of colorectal hepatic metastases on the remaining liver parenchyma. *European journal of gastroenterology & hepatology* **27**, 399–404, <https://doi.org/10.1097/meg.0000000000000316> (2015).
19. Kandpal, H., Sharma, R., Madhusudhan, K. S. & Kapoor, K. S. Respiratory-triggered versus breath-hold diffusion-weighted MRI of liver lesions: comparison of image quality and apparent diffusion coefficient values. *AJR. American journal of roentgenology* **192**, 915–922, <https://doi.org/10.2214/AJR.08.1260> (2009).
20. Taouli, B. *et al.* Diffusion-weighted imaging of the liver: comparison of navigator triggered and breathhold acquisitions. *Journal of magnetic resonance imaging: JMRI* **30**, 561–568, <https://doi.org/10.1002/jmri.21876> (2009).
21. Nasu, K., Kuroki, Y., Sekiguchi, R. & Nawano, S. The effect of simultaneous use of respiratory triggering in diffusion-weighted imaging of the liver. *Magnetic resonance in medical sciences: MRMS: an official journal of Japan Society of Magnetic Resonance in Medicine* **5**, 129–136 (2006).
22. Jerome, N. P. *et al.* Comparison of free-breathing with navigator-controlled acquisition regimes in abdominal diffusion-weighted magnetic resonance images: Effect on ADC and IVIM statistics. *Journal of Magnetic Resonance Imaging* **39**, 235–240, <https://doi.org/10.1002/jmri.24140> (2014).
23. Kwee, T. C., Takahara, T., Koh, D. M., Nieuvelstein, R. A. & Luijten, P. R. Comparison and reproducibility of ADC measurements in breathhold, respiratory triggered, and free-breathing diffusion-weighted MR imaging of the liver. *Journal of magnetic resonance imaging: JMRI* **28**, 1141–1148, <https://doi.org/10.1002/jmri.21569> (2008).
24. Chen, X. *et al.* Liver diffusion-weighted MR imaging: reproducibility comparison of ADC measurements obtained with multiple breath-hold, free-breathing, respiratory-triggered, and navigator-triggered techniques. *Radiology* **271**, 113–125, <https://doi.org/10.1148/radiol.13131572> (2014).
25. Pathak, R. *et al.* A data-driven statistical model that estimates measurement uncertainty improves interpretation of ADC reproducibility: a multi-site study of liver metastases. *Sci Rep* **7**, 14084, <https://doi.org/10.1038/s41598-017-14625-0> (2017).
26. Lambregts, D. M. *et al.* Tumour ADC measurements in rectal cancer: effect of ROI methods on ADC values and interobserver variability. *European radiology* **21**, 2567–2574, <https://doi.org/10.1007/s00330-011-2220-5> (2011).
27. Ragheb, H. *et al.* The Accuracy of ADC Measurements in Liver Is Improved by a Tailored and Computationally Efficient Local-Rigid Registration Algorithm. *PloS one* **10**, e0132554, <https://doi.org/10.1371/journal.pone.0132554> (2015).
28. Gudbjartsson, H. & Patz, S. The Rician distribution of noisy MRI data. *Magnetic resonance in medicine* **34**, 910–914 (1995).
29. Asselin, M. C., O'Connor, J. P., Boellaard, R., Thacker, N. A. & Jackson, A. Quantifying heterogeneity in human tumours using MRI and PET. *European journal of cancer* **48**, 447–455, <https://doi.org/10.1016/j.ejca.2011.12.025> (2012).
30. O'Connor, J. P. *et al.* Imaging intratumour heterogeneity: role in therapy response, resistance, and clinical outcome. *Clinical cancer research: an official journal of the American Association for. Cancer Research* **21**, 249–257, <https://doi.org/10.1158/1078-0432.CCR-14-0990> (2015).
31. Tourell, M. C. *et al.* The distribution of the apparent diffusion coefficient as an indicator of the response to chemotherapeutics in ovarian tumour xenografts. *Sci Rep* **7**, 42905, <https://doi.org/10.1038/srep42905> (2017).

Acknowledgements

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking (www.imi.europa.eu) under grant agreement number 115151, resources of which are composed of financial contribution from the European Unions Seventh Framework Programme (FP7/2007–2013) and EFPIA companies in-kind contribution. There was, however, no financial or in-kind contribution from EFPIA companies to the research specifically described in this paper. Funding was also provided by the CRUK grant for The Cambridge and Manchester Cancer Imaging Centre (C8742/A18097). The funders of the research leading to these results had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

We are submitting original research with data that has not yet been published in any journal. We declare that all named authors have read the manuscript and have agreed to submit in its present form. All named authors have made a sufficient contribution to the work. R.P. has been involved with the recruitment, data acquisition, analysis and writing of the first and subsequent drafts. J.T. and H.R. has been involved with the data analysis and development of the statistical error model. N.T. has been heavily involved in the design and supervision of the statistical error model and data analysis. D.M. was responsible for design and implementation of the standardized MRI protocol for the overall project. C.S. was involved in clinical data collection and analysis including the LDH data. M.S. was the clinical P.I., responsible for patient recruitment, as well as having input into the writing process. A.J. was the overall project lead and edited second and subsequent drafts of the manuscript and provided overall supervision and invaluable guidance.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-40565-y>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019