Reliability of the apparent diffusion coefficient assessment of cervical lymph nodes in phantoms and healthy volunteers

Poster No.: C-0591
Congress: ECR 2018
Type: Scientific Exhibit
Authors: A. Iannessi¹, H. Beaumont², B. MOREAU¹, C. Hoog¹, C. Klifa²;
¹Nice/FR, ²Valbonne/FR
Keywords: Oncology, Molecular imaging, Lymph nodes, MR-Diffusion/Perfusion, Experimental investigations, Instrumentation, Normal variants
DOI: 10.1594/ecr2018/C-0591

Any information contained in this pdf file is automatically generated from digital material submitted to EPOS by third parties in the form of scientific presentations. References to any names, marks, products, or services of third parties or hypertext links to third-party sites or information are provided solely as a convenience to you and do not in any way constitute or imply ECR’s endorsement, sponsorship or recommendation of the third party, information, product or service. ECR is not responsible for the content of these pages and does not make any representations regarding the content or accuracy of material in this file.

As per copyright regulations, any unauthorised use of the material or parts thereof as well as commercial reproduction or multiple distribution by any traditional or electronically based reproduction/publication method is strictly prohibited.

You agree to defend, indemnify, and hold ECR harmless from and against any and all claims, damages, costs, and expenses, including attorneys’ fees, arising from or related to your use of these pages.

Please note: Links to movies, ppt slideshows and any other multimedia files are not available in the pdf version of presentations.

www.myESR.org
Aims and objectives

Apparent Diffusion Coefficient (ADC) can quantify the level of free water diffusion restricted by an increase in tissue cellularity. Applications of ADC in cancer imaging has motivated intensive research: ADC is now one of the main QIBs derived from Diffusion MRI.

Several studies have documented the incremental value of ADC assessment as a complement or substitute to standard sequences for the detection of malignant tumors [1], the degree of malignancy [2][3] or to evaluate response to treatment [4][5][6]. Since lymph node involvement is pivotal in oncologic imaging [7], ADC has been tested for its detection of malignant adenomegalies [8][9]. Results are discordant [10] [11].

Previous literature comprises heterogeneous studies protocols and results [12]. Several sequential unitary processes are necessary to output an ADC assessment, the lack of reliability of any of these unitary processes is likely to degrade the final ADC assessment. It is therefore particularly relevant to study if ADC qualifies as a quantitative biomarker.

The main objective of this study was to evaluate the variability of ADC measurements in vitro on a phantom, and in vivo on cervical lymph nodes. The secondary objective was to understand and quantify ADC measurement errors, in view of correcting them in future studies.
Methods and materials

We first tested QIBA metrics for quality control (QC) of ADC image quality, and then performed a reliability analysis of ADC measurements. Finally we measured ADC values of cervical lymph nodes in healthy volunteers.

This prospective study was conducted at the Centre Antoine Lacassagne, cancer center in Nice, France, between March and November 2016. We used a GE MRI scanner 1-5T MR450W and ADW Volume Share 5.4.6 software to process images (GE Healthcare).

I. Quality control test

We used a DIN 6858-1 PET-CT phantom (PTW) consisting in a cylindrical Plexiglas body filled with a mixture of ice and water. Three smaller cylinders were inserted into the body, one of them was filled with water at 0° C (Figure 1, left side).

Homogeneity of temperature inside the cylinder was thermometer-controlled according to the process defined into QIBA profile to achieve thermal equilibrium (>1 hour) over the entire MRI exam period. For each b values, four successive acquisitions spaced in time from more than 12min were performed allowing retrospective checks.

The diffusion protocol was: 3 directions DW SS-EPI with b=0, 100, 600, 800 s/mm², TR=9451ms, TE=80ms, Number of average = 2, FOV 320*320mm, contiguous slice thickness of 4mm, encoding frequency axis R/L.

Four successive acquisitions were made for each b value, the phantom symmetry axis was laser-centered to the magnetic field positioning the 0°C water cylinder at the center of the scanner. Acquisitions of the phantom were performed horizontally (x-axis) and vertically (y-axis). We measured circular regions of interest (ROI) of 2.5 cm diameter and composed of 123 voxels (Figure 2). Mean ADC and standard deviation (SD) were computed.

According to equations in Table 1, we computed: the measurement Repeatability (R), estimated by the Coefficient of Variation (CV_R) and the Repeatability Coefficient (RC_R), the accuracy (ADC Bias estimate), ADC noise estimate and b-value dependency.

The signal-to-noise ratio (SNR) was computed using formula F (shown in Table 1) and involved computing the "Temporal Noise Image" from the diffusion mapping at b = 0, with a 2 cm circular ROI.

Results were compared to QIBA 's references values [13].
In addition, we analyzed the planar spatial correlation of ADC measures in shifting ROIs along the x and y axis. The ADC reference value was measured at the image center using formula C (see Table 1). We used circular ROIs of 2.2 cm diameter and 2cm shifts from the center either to the right (x-axis) or to the bottom (y-axis) of the image.

II. Measurement variability

1. SPHERE phantom study

A second phantom was used (NEMA NU2-2012 (PTW)), named SPHERE Phantom (Figure 1, right side). The SPHERE phantom embedded 6 different spheres (diameters 10, 13, 17, 22, 28 and 37mm), filled with room temperature water.

We simulated clinical conditions in using the cervical level of the routine whole-body MRI i.e axial DW SS EPI with b=50 and b=1000s/mm², TR=10384ms, TE set to minimum (around 70ms for all scans), Number of averages=2, Parallel imaging factor=2, FOV=400*400mm, contiguous 5mm slice thickness, encoding frequency axis R/L. The phantom was laser centered, equidistant from all spheres. Four acquisitions were made at 1 day intervals. All values were averaged over 4 days.

ADC measures were obtained from spherical Volumes of Interest (VOIs) centered on spheres Figure 3.

The relative ADC error was computed for each sphere size, considering that the reference ADC value was from the 37 mm sphere. We analyzed the correlation between VOI size and precision of measurements in computing the CVR. Additional analysis documented the measurement error, first in measuring bias, second in computing the CVR through several concentric VOIs of decreasing size in the largest sphere, according to Table 1 (Formula A). Then partial volume effect was quantified by calculating the relative error within a VOI with a diameter equal to 80% the diameter of a sphere compared to a VOI of identical size within the largest sphere. Mean and SD of ADC values were computed for all VOIs size.

2. In vivo study

Informed consent were obtained from 13 healthy volunteers. Exclusion criteria were: chronic disease, history or ongoing symptoms of infection like fever, cough, rhinorrhea, dysphagia, and odynophagia, history of cervical surgery, claustrophobia, and all usual contraindications for MRI. Demographic status and smoking habits were recorded for the 13 volunteers. Volunteers were scanned using the same machine as phantom study. The acquisition was performed with a neck phased-array coils and the volunteer was instructed to breath normally.
Technical settings of diffusion sequence for volunteers were identical to those of the SPHERE phantom.

Two readers assessed ADC values of lymph nodes: a senior radiologist with more than 6 years of experience in cancer imaging, and a junior radiologist.

Lymph nodes volume were manually segmented on the b1000 scan, and the graphic was exported to the ADC map (Figure 4). At least 4 lymph nodes were selected per volunteer, including the largest. VOIs were segmented in delineating hyper intense diffusion areas on b1000 scans while excluding lymph nodes hilum. Each node was segmented twice by each observer using the same acquisition with an interval of 7 to 60 days (mean 41 days) [14]. Mean and SD ADC values were recorded.

Inter and intra-observer agreements were calculated according to Bland Altman's method using the R CRAN software. Bias and Limit of Agreement (LoA) were computed. Inter and intra-observer differences in segmenting lymph node's volume and in ADC values were analyzed using the sum of Wilcoxon ranks for paired values test. A result of p value < 0.05 was considered significant.
Images for this section:

**Fig. 1:** Phantoms used in the study. Left: ICEWATER phantom filled with 0° C water (DIN 6858-1, PTW, Freiburg, Germany). Right: SPHERE phantom at room temperature featuring spheres of various sizes between 10 and 37 mm diameters (NEMA NU2-2012, PTW, Freiburg, Germany).

© PTW, Freiburg, Germany

**Fig. 2:** Imaging of ICEWATER phantom at b0-b100 (top) and b0-b800 (bottom). From left to right: Diffusion mapping, axial view of ADC mapping and coronal view of ADC mapping. Red circular ROIs are set at center of the ice water cylinder.
### Table 1: Definition of quality control metrics according to QIBA DW-MRI profile

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td>( CV_R = 100% \times \frac{\sigma_R}{\mu_R} )</td>
</tr>
</tbody>
</table>
| | \( CV_R \): Coefficient of variation (\%)  
| | \( \sigma_R \): standard deviation (mm²/s) of each measurements  
| | \( \mu_R \): mean of each measurements ADC means (mm²/s) |
| **B** | \( RC_R = 2.77 \times \sigma_R \) |
| | \( RC_R \): repeatability coefficient (mm²/s) |
| **C** | ADC bias estimate = \( \frac{\mu - DC_{True}}{\sigma} \)  
| | \( \%bias = 100\% \times \frac{\mu - DC_{True}}{DC_{True}} \)  
| | \( DC_{True} \): ADC=1,1.10^3 mm²/s in 0°C water |
| **D** | ADC noise Estimate = \( 100\% \times \frac{\sigma}{\mu} \) |
| | \( \sigma \): standard deviation of ADC values within the ROI (mm²/s)  
| | \( \mu \): mean ADC (mm²/s) within the ROI |
| **E** | ADC b value dependence = 100\% \( \frac{ADC_{b_{min,b2}} - ADC_{b_{min,b1}}}{ADC_{b_{min,b1}}} \)  
| | \( b_{min} = b_0 \)  
| | \( b_1 = b_{600} \)  
| | \( b_2 = b_{800} \) |
| **F** | SNR_{ndyn} = \( \frac{\text{Spatial mean pixel value on Signal Image}}{\text{Spatial mean pixel value on Temporal Noise Image}} \) |
| | \( \text{SNR}_{ndyn} \): Signal to Noise Ratio |

**Fig. 3:** Measurements on the SPHERE phantom  
Left: Spherical VOIs of decreasing sizes centered on the largest sphere.  
Right: Spherical VOIs centered on sphere of various sizes. VOIs diameters are set to 80% of physical sphere’s diameters.
Fig. 4: Measurements of cervical lymphnodes - Imaging of a healthy volunteers cervical lymph node. a): Diffusion mapping at $b = 50$. b): mapping at $b = 1000$ on which the Volume of Interest (VOI) is contoured before being exported to other series. c) and d): Mapping of the Apparent Diffusion Coefficient. In red, the VOI is determined by operator 1 (a, b and c) and by operator 2 (d.)

© Sciences, MEDIAN Technologies - Valbonne/FR
Results

I. Quality control test

Table 2 displays QIBA's recommended limit values for: repeatability, accuracy, precision and b-value dependency.

We found that SNR computed from diffusion scans at b = 0 was 17:1, lower than the recommended limit value (50:1).

Variations of ADC measures relative to spatial positions are summarized in Figure 5.

We found no significant correlation between ADC values and lateral shifts, but a significant correlation with vertical shifts (Pearson correlation coefficients respectively r = 0.25, 95%CI [-0.45, 0.76] and r = 0.95 95%CI [0.46; 1]).

We also found an increasing bias when shifting measurements from the scan isocenter (Table 3). The maximum 10% error threshold recommended by QIBA was exceeded for ADC values measured at least 6 cm distant from the isocenter (Table 3).

II. Reliability analysis

1. SPHERE phantom study

Firstly, our analysis showed that when VOIs are set within spheres of decreasing size, relative error and measurements variability of ADC measurements increased (Table 4). Secondly, we found no significant mean ADC difference for VOIs of decreasing sizes set within the largest sphere. In that case, we found less than 2% error between largest and smallest VOIs.

Correlation and variability analysis of ADC measurements with VOI size seemed to indicate a significant partial volume effect. Partial volume effect was visually confirmed on images.

2. In vivo study

Thirteen volunteers were included in the in vivo study. Age ranged 22 to 50 (mean 32.4 years), and gender ratio (M/F) was 38.5%. Two volunteers were active smokers or recent
ex-smokers (15.4%). Overall, fifty-four cervical lymph nodes were selected for analysis mainly on carotid-jugular sites, with a mean volume of 1 cm$^3$ (Appendix 1).

The mean value of measured ADC was $0.87 \times 10^{-3}$ mm$^2$/s (0.66 to 1.28 $10^{-3}$ mm$^2$/s, SD was 0.12 $10^{-3}$ mm$^2$/s). We found a significant difference between the average ADC values measured by readers 1 and 2 (0.84 $10^{-3}$ and 0.90 $10^{-3}$ mm$^2$/s, respectively, $p < 0.0001$).

The inter-reader analysis showed a relative bias of -5.5%, LoA was [-18.8%; 7.7%]). The absolute bias was $0.045 \times 10^{-3}$ mm$^2$/s, LoA was [-0.146; 0.056]).

We found a significant difference of average segmented volumes between reader 1 and 2 (respectively 1.18 +/- 0.94 cm$^3$ and 1.92 +/- 1.23 cm$^3$, $p <0.0001$). There was a low correlation between measurement differences in terms of average ADC and volume segmentation ($R^2 = 0.37$) by the two observers.

Intra-observer analysis showed, respectively for reader 1 and 2, a relative bias = 0.6%; LoA=[-9.2%; 10.4%] and relative bias = 0.5%; LoA=[-8.8%; 7.7%].

Using the Beaumont et al. method [15] and based on our intra-observer reproducibility parameters, we can estimate that on longitudinal studies under strict reproducible conditions (same patient, same reader), a meaningful relative change of ADC value should be outside the range [-13%; + 15%].
Table 2: Quality control after imaging the ICEWATER phantom - Outcome of the quality control after imaging the ICEWATER phantom. The test was done with different b values as b0-b100, b0-b600 and b0-b800. Tests not meeting QIBA quality claims are displayed in bold.

© Sciences, MEDIAN Technologies - Valbonne/FR
Fig. 5: Spatial correlation of ADC - Top view: ADC changes according to the horizontal distance from magnetic center. Horizontal axis: distance in cm. Vertical axis: ADC value in mm2/s. Bottom view: ADC changes according to the vertical distance from magnetic center. Horizontal axis: Distance in cm. Vertical axis: ADC value in mm2/s.

© Sciences, MEDIAN Technologies - Valbonne/FR
**Table 3:** Spatial variations of ADC measurements - Spatial variations of ADC measurements with respect to a reference VOI at center of the magnetic field. Top rows: ADC measurements are shifted horizontally. Bottom rows: ADC measurements are shifted to the bottom. We found that ADC measurement did not change significantly when shifted right (Pearson coefficient=0.25). In opposite, ADC values increased when measurements were shifted vertically (Pearson coefficient=0.95)

© Sciences, MEDIAN Technologies - Valbonne/FR

<table>
<thead>
<tr>
<th>True sphere’s diameter (mm)</th>
<th>37</th>
<th>28</th>
<th>22</th>
<th>17</th>
<th>13</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Diameter of measured VOI’s (mm)</em></td>
<td>30</td>
<td>22</td>
<td>18</td>
<td>14</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Mean ADC over 4 days (10^{-3} mm²/s)</td>
<td>2.05</td>
<td>2.07</td>
<td>2.13</td>
<td>2.15</td>
<td>2.31</td>
<td>1.97</td>
</tr>
<tr>
<td>Relative error of ADC measurements (%)</td>
<td>0 (ref)</td>
<td>0.98</td>
<td>3.9</td>
<td>4.76</td>
<td>12.8</td>
<td>-3.9</td>
</tr>
<tr>
<td>Coefficient of variation (%)</td>
<td>0.4</td>
<td>0.39</td>
<td>1.01</td>
<td>2.1</td>
<td>2.33</td>
<td>2.84</td>
</tr>
</tbody>
</table>

**Table 4:** Scaling effect of ADC measurements - Repeated measurements of SPHERE Phantom images were performed over 4 days. Spheres of different sizes were measured. Top row: True size of spheres of interest. Second row: Diameter of VOIs centered on spheres of interest. Third row: For each sphere, measurements have been repeated four time over four days. Mean ADC values have been computed. Fourth row: Relative error (%) with respect to the VOI set into the largest sphere. Bottom row: Coefficient of variation (%) of repeated measurements over four days. To be noted: Regarding right side column, unlike for other measurements, true size sphere and VOI have same size (10mm) because of the sampling limit (8 Voxel into the VOI). As a consequence, a nonlinear effect has been observed: Decrease of the Relative error and the mean ADC value. This point is further developed in Table 5.

© Sciences, MEDIAN Technologies - Valbonne/FR
Conclusion

Our QC results showed good compliance with QIBA metrics, except for ADC bias estimate slightly above the limit, and with a variability of about 9%. Results were independent of the value of b.

We questioned if the main part of error was due to our phantom design featuring a large off-axis volume of water and thermally suboptimal materials. Using repeated imaging of the phantom, we found however a good repeatability, suggesting acceptable thermal equilibrium.

SNR was also lower than QIBA’s recommendation but Malyarenko et al. [16] reported that low SNR has no impact on ADC assessments. Very low SNR without adequate post-processing would probably alter measurements as most software (including the one we used) compute ADC images in thresholding/removing low intensity voxels.

We highlighted a correlation between ADC measurement error and the distance of ROIs from the magnetic center. The error increased with bottom-shift (up to 24% when located 8cm out of isocenter). Conversely, with regard to lateral-shift we found no correlation with the magnitude of errors. This result can be explained by non-uniformity of gradient-fields.

As we found a correlation between variability of ADC assessments and contour segmentations, we concluded that partial volume effect was a major contributor to the variability.

A visual review of outliers in clinical data confirmed high variation of signal intensity in the tissue surrounding these lymph nodes: Consequently, even a small variation in segmentations led to a significant modification of ADC assessments (Figure 6)

We recommend to measure ADC by drawing ROIs smaller than the anatomical limits of the area of interest. How to optimize segmentation margins must be further investigated.

We found excellent repeatability and good reproducibility. This suggest that, if ADC intended to evaluate response to treatment, changes inferior to [-13%; +15%] may not be clinically relevant. However, longitudinal reproducibility would require further clinical studies to take into account all variability factors.

According to our dataset, the averaged ADC value for healthy subject's cervical lymph nodes was $0.87 \times 10^{-3} \pm 0.12 \times 10^{-3}$ mm²/s

Our results are well supported by the literature.
Regarding the correlation between variability of ADC assessments and contour segmentation, heterogeneous segmentation methods are available but several studies documented the reproducibility issues [17][18][19] affecting these methods. These different approaches are also reported as cumbersome and time-consuming [20].

Specific phantom studies have shown that gradient-field error would be scanner-dependent [21] and not significant within 4 cm from isocenter, explaining the good reproducibility of our ADC measurements and in other multi-centric studies. On multiple scanners, measurements at 12 cm from the isocenter showed an average error of -20% according to vertical shift and +7% horizontally.

Unlike our observation of an ADC value of $0.87 \pm 0.12 \times 10^{-3}$ mm²/s in healthy subject's cervical lymph nodes, Kwee et al. [10] reported a range of $[1.15 \times 10^{-3}$ mm²/s; $1.18 \times 10^{-3}$ mm²/s], with similar intra- and inter-observers' variabilities. The review of 12 studies including more than 1200 benign lymph nodes reports ADC values of $0.302 \pm 0.062 \times 10^{-3}$ mm²/s in inflammatory cervical nodes [22] and $2.38 \pm 0.29 \times 10^{-3}$ mm²/s for abdominal nodes [23]. Kwee et al. concluded that disparity of results could be due to the various segmentation methods used.

Our results for non-diseased ADC values overlap with metastatic or lymphomatous lymph nodes ADC measures $[0.410 \pm 0.105 \times 10^{-3}$ mm²/s; $1.84 \pm 0.37 \times 10^{-3}$ mm²/s] as reported by other groups [22][23], however the ADC values we found match with other non-diseased ADC studies. A radiological-pathological correlation study by Vandecaveye et al. [24] on 331 cervical lymph nodes proposed an ADC threshold of $0.94 \times 10^{-3}$ mm²/s for detecting nodes malignancy featuring a specificity of 94%. According to this threshold, 72% of our data would have been misclassified. The use of ADC values to assess cervical lymph nodes malignancy does not reach a consensus.

At this time, variabilities from different sources preclude a larger adoption of the ADC biomarker even though it is an important advance in cancer imaging. Generalizing quality controls and standardization of measurements is crucial to overcome these ADC variability issues.
Fig. 6: Example of inter-observer discordance in terms of volume and ADC on a level II lymph node. Top row: First reader's measurements, Bottom row: Second reader's measurements. Left: b1000 diffusion maps where VOIs are drawn. Right: corresponding ADC maps. To be noted: The heterogeneity of the node's environment featuring areas of high ADC values (green and red in the right images) without clear correspondence on the diffusion image.
Personal information

Contact: hubert.beaumont@mediantechnologies.com
References


