Pre-trial quality assurance of diffusion-weighted MRI for radiomic analysis and the role of harmonisation

Zelda Paquier a,*, Shih-Li Chao a, Giacomo Bregni b, Ana Veron Sanchez c, Thomas Guiot a, Jennifer Dhont a, Akos Gulyban a, Hugo Levillain a, Francesco Sclafani b, Nick Reynaert a, Maria Antonietta Bali c

a Medical Physics Department, Institut Jules Bordet – Université Libre de Bruxelles, Brussels, Belgium
b Medical Oncology Department, Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium
c Radiology Department, Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium

ARTICLE INFO

Keywords:
MRI
ADC
Quality assurance
Radiomics

ABSTRACT

Purpose: The aim of this study was to perform a quantitative quality assurance of diffusion-weighted MRI to assess the variability of the mean apparent diffusion coefficient (ADC) and other radiomic features across the scanners involved in the REGINA trial.

Materials and methods: The NIST/QIBA diffusion phantom was acquired on six 3 T scanners from five centres with a rectum-specific diffusion protocol. All sequences were repeated in each scan session without moving the phantom from the table. Linear interpolation to two isotropic voxel spacing (0.9 and 4 mm) was performed as well as the ComBat feature harmonisation method between scanners. The absolute accuracy error was evaluated for the mean ADC. Repeatability and reproducibility within-subject coefficients of variation (wCV) were computed for 142 radiomic features.

Results: For the mean ADC, accuracy error ranged between 0.1% and 8.5%, repeatability was <1% and reproducibility was <3% for diffusivity range between 0.4 and 1.1x10^{-3} mm^2/s. For the other radiomic features, wCV was below 10% for 24% and 15% features for repeatability with resampling 0.9 mm and 4 mm, respectively, and 13% and 11% feature for reproducibility. ComBat method could improve significantly the wCV compared to reproducibility without ComBat (p-value < 0.001) but variation was still high for most of the features.

Conclusion: Our study provided the first investigation of feature selection for development of robust predictive models in the REGINA trial, demonstrating the added value of such a quality assurance process to select conventional and radiomic features in prospective multicentre trials.

Introduction

Diffusion weighted imaging (DWI) in magnetic resonance imaging (MRI) assesses the random Brownian motion of water molecules non-invasively. It is widely used for diagnostics due to its unique soft tissue contrast [1,2] that is reflecting the tissue microstructure [3]. The apparent diffusion coefficient (ADC) provides a quantitative measure of DWI that is derived from monoexponential modelling of the signal intensity as a function of b-value [4]. ADC is considered a promising biomarker of therapy response [5], notably for systemic therapy, by offering indicators with the potential to predict treatment response earlier than conventional size parameters [6,7]. Combining conventional DWI metrics with radiomics – the extraction of supplementary phenotypic information from medical images [8] – is offering further opportunities [9,10].

The results of DWI and radiomics are however variable across scanner platforms, sequences and data processing methods [11,12]. This lack of reproducibility limits the application of potential quantitative biomarkers from scientific research to decision-making in clinical practice [13]. Standardisation of DWI acquisition techniques across multiple MRI platforms is thus warranted, especially for multicentre clinical trials to reduce bias. This is one of the goals of the Quantitative Imaging Biomarkers Alliance (QIBA), which provides technical performance standards for various imaging modalities, including DWI [14].

* Corresponding author at: Institut Jules Bordet, Rue Meylaeenss 90, 1070 Brussels, Belgium.
E-mail address: zelda.paquier@bordet.be (Z. Paquier).

https://doi.org/10.1016/j.ejmp.2022.10.009
Received 16 May 2022; Received in revised form 30 September 2022; Accepted 8 October 2022
Available online 27 October 2022
1120-1797/© 2022 Associazione Italiana di Fisica Medica e Sanitaria. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
The National Institute of Standards and Technology (NIST), alongside QIBA, has developed a phantom which allows precise assessment of the ADC range that might be encountered in human tissue. The phantom can be used to standardize DWI acquisition schemes across multiple vendors.

Acquiring a phantom, however, comes with an extra cost and is futile in the context of retrospective studies. To overcome this, the ComBat harmonisation method was introduced in radiomic analysis. This data-driven method that does not require the acquisition of phantom data is a post-processing step to remove the variability between feature distributions due to scanner and protocol effect, without altering the biological information [15]. The approach has already been validated on different imaging modalities with phantoms and patient data [15–19].

The REGINA trial (NCT04503694) is a phase II trial of neoadjuvant regorafenib, in combination with nivolumab and short-course radiotherapy for intermediate-risk stage II-III rectal cancer [20]. One of the exploratory objectives of this trial was to perform a quality assurance of DWI acquisition using a physical imaging phantom in all centres participating in the REGINA trial, as is recommended by QIBA [14]. The repeatability and reproducibility of the mean ADC and other radiomic features were also assessed, with and without ComBat harmonisation between scanners.

Materials and methods

Study design

This phantom study is composed by multiple steps. At first, mean ADC values of the NIST/QIBA diffusion phantom were compared with the ground truth (accuracy). Secondly, repeatability and reproducibility were assessed, not only for mean ADC but also for other radiomic features. Finally, the ComBat method for the harmonisation of these feature values was evaluated. The workflow of our phantom study is shown in Fig. 1.

Phantom

The phantom used in this study was the NIST/QIBA diffusion phantom (Model 128, Qalibre MD, Boulder, Colorado USA). It consists of 13 vials filled with different polymer concentrations (from 0 % to 50 %) in an aqueous solution, giving different ADC values \([0.128–1.127 \times 10^{-3} \text{mm}^2/\text{sec}]\). One vial with 0 % polymer is in the central position and the others are positioned on two circles around it (inner and outer circle), as shown on Fig. 2. The vials are surrounded by ice water to control the phantom’s temperature at 0 °C, as the diffusion process is thermally-driven.

A single phantom was used across the different institutions. It was positioned in a systematic way between scanners always by the same person, with uniform cylindrical phantoms on both sides to improve the filling factor and with foam pads for stabilisation, as shown in Fig. 3. Laser positioning was used to align the centre of the phantom to the isocentre of the magnet, with its central axis parallel to the z-axis of the magnet. Finally, the body-matrix coil was carefully fixed above the phantom.

MRI

Phantom measurements were performed on six MR systems across five medical centres. All MR systems had a 3 T field strength, but were from different manufacturers (Siemens, GE, Philips). Prior to the centre visit, a DWI data acquisition protocol (Supplementary Data 1, Table 1) designed for the REGINA trial was sent to each centre with the request to comply with the parameters within the limits of their system. Centres didn’t need a specific research environment to implement it, it remained within the condition of the clinical one. This protocol was introduced by each institution’s physicist or manufacturer’s technician.

After the acquisition of a localiser to center the phantom on the images, diffusion data were obtained consisting of an echo-planar imaging (EPI) sequence with nine b-values. Details of each scanner acquisition protocol are given in Table 1. DWI was performed three times within each scanning session without moving the phantom.
between each acquisition to assess the short-term repeatability. On scanner B, measurements at three time points set (one week apart from each other) were acquired, to assess long-term repeatability.

**Image processing and segmentation**

Each dataset was analysed by the coordinating centre, first qualitatively with a visual check for artefacts and then quantitatively. The ADC maps were calculated with an in-house developed software, with the nine b-values and assuming a mono-exponential signal decay. The noise was estimated using the image with the 0 sec/mm² b-value (b₀). The mean and standard deviation (SD) of the signal were calculated in the b₀ image, excluding a four pixel margin at the periphery. The ADC calculation was rejected if the signal was less than (mean + 3SD) × NoiseFactor, with the NoiseFactor set to 1.2. We also considered the signal drop between b₀ and b1500 images by excluding the calculation when the value was less than SD × 5. Non-computed pixels were set to 0.

The signal-to-noise ratio (SNR) was computed as described by the QIBA protocol [21] in MICEd tool v2021.1.0 (NOMPI Medical AB, Umeå, Sweden) [22]. The maps were then spatially resampled to an isotropic voxel spacing to be rotationally invariant in the same software [23]. Two schemes were analysed in this work: upsampling (0.9x0.9x0.9 mm³) and downsampling (4x4x4 mm³) using linear interpolation. This step was performed to ensure an identical volume between scanners during the segmentation. These new images were then sent to MIM v7.1.5 (MIM Software Inc., Cleveland, OH) where 13 spherical volumes of interest (VOI) with a diameter of 15 mm were drawn in the centre of each vial at the central plane of the phantom. Finally, delineations were exported into a single DICOM RT Structure Set file per patient.

**Feature extraction**

RadiomiX Research Toolbox (Radiomics, Liège, Belgium) was used for radiomic processing. To extract the radiomic features, the DICOM data and the corresponding segmentation file (.RTStruct file) were imported into the software tool. Except for image discretisation, no image processing was performed inside the application (neither spatial resampling nor normalisation). Intensity discretisation was applied with a fixed bin size (FBS), set to the average standard deviation of the signal intensity in all VOI [17,18], resulting in an FBS of 10. Calculation settings were left at their default values (GLCM distance of 1).

Investigated features were categorised by intensity (n = 46) and texture (n = 96). The shape category was not analysed because identical VOI were used for each image. The intensity category combines 2 local intensity features (LocInt), 19 intensity-based statistical features (Stats), and 25 intensity histogram features (IH). The texture category includes 26 features from grey level co-occurrence matrix (GLCM), 16 features from grey level run length matrix (GLRLM), 16 features from grey level size zone matrix (GLSZM), 16 features from grey level distance zone matrix (GLDZM), 5 features from neighbourhood grey tone difference matrix (NGTDM), and 17 features from neighbouring grey level dependence matrix (NGLDM).

**ComBat harmonisation**

ComBat is a realignment method that can be directly applied to radiomic features to correct for the scanner effect [15]. It matches the statistical distributions of the feature values measured in VOI j for each scanner i by the equation [18]:

\[ y_{ij} = \alpha + \gamma_i + \delta_i \epsilon_{ij} \]

where \( \alpha \) is the average value for feature \( y_{ij} \); \( \gamma_i \) is the additive scanner effect on features; and \( \delta_i \) is the multiplicative scanner effect affected by...
Diffusion-weighted imaging (DWI) magnetic resonance imaging protocol. Abbreviations: TR = Repetition Time; TE = Echo Time.

<table>
<thead>
<tr>
<th>Scanner A</th>
<th>Scanner B</th>
<th>Scanner C</th>
<th>Scanner D</th>
<th>Scanner E</th>
<th>Scanner F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer</td>
<td>Siemens</td>
<td>Siemens</td>
<td>Siemens</td>
<td>Siemens</td>
<td>Philips</td>
</tr>
<tr>
<td>Model</td>
<td>Magnetom Vida</td>
<td>Skyra</td>
<td>Skyra</td>
<td>Magnetom Vida</td>
<td>Achieva dStream</td>
</tr>
<tr>
<td>Field strength (T)</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Number of coil channels</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>30</td>
<td>32</td>
</tr>
<tr>
<td>Parallel imaging factor</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Percent sampling</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>98.8</td>
</tr>
<tr>
<td>Pixel spacing (mm)</td>
<td>1.488/1.488</td>
<td>1.488/1.488</td>
<td>1.488/1.488</td>
<td>1.488/1.488</td>
<td>1.477/1.477</td>
</tr>
<tr>
<td>Slice thickness (mm)</td>
<td>4.5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Spacing between slices (mm)</td>
<td>4.95</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
<td>4.3</td>
</tr>
<tr>
<td>Number of slices</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>31</td>
</tr>
<tr>
<td>Acquisition time (min)</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>TR (msec)</td>
<td>7500</td>
<td>6700</td>
<td>6700</td>
<td>7300</td>
<td>4680</td>
</tr>
<tr>
<td>TE (msec)</td>
<td>6700</td>
<td>7300</td>
<td>4680</td>
<td>4500</td>
<td>102.1</td>
</tr>
<tr>
<td>b-values (sec/mm²)</td>
<td>average(s):</td>
<td>average(s):</td>
<td>average(s):</td>
<td>average(s):</td>
<td>average(s):</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>50</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>150</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>250</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>400</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>600</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>900</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1200</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1500</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Bandwidth (Hz)</td>
<td>1653</td>
<td>1655</td>
<td>1655</td>
<td>1653</td>
<td>3591</td>
</tr>
</tbody>
</table>

$\epsilon_i$, an error term.

ComBat uses a maximum likelihood approach to estimate $\alpha_i$, $\gamma_i$, and $\delta_i$, denoted as $\hat{\alpha}_i$, $\hat{\gamma}_i$, and $\hat{\delta}_i$. The normalised value of feature $y_{ij}$ is then obtained by:

$$y_{ij}^{\text{ComBat}} = \frac{y_{ij} - \hat{\alpha}_i - \hat{\gamma}_i}{\hat{\delta}_i} + \hat{\alpha}$$

It was computed in Python v3.7.7 using the library neurocombat v0.2.12, by Fortin et al [24,25], using the non-parametric form of the model and without empirical Bayes assumption.

Statistical analysis

Statistical analysis was performed in Python v3.7.7. Accuracy error was calculated for each vial as:

$$\text{Accuracy Error} = \left| \frac{\text{Measured} - \text{Reference}}{\text{Reference}} \right| \times 100\%$$

with the reference values specified by the phantom manufacturer and the measured values corresponding to the mean ADC taken after resampling on an isotropic voxel of 0.9x0.9x0.9 mm$^3$.

For short/long-term repeatability and reproducibility, the within-subject coefficient of variation (wCV) in percentage was calculated as specified in QIBA recommendations [14]. It quantifies the intra-subject variability between measurements in order to detect the minimum detectable difference when two longitudinal measurements are used (repeatability) or different imaging methods (reproducibility). When the wCV is small, minimal changes in the feature can be detected, whereas when the wCV is large, considerable variations in the feature are required before one can be confident that a real alteration has occurred. For the repeatability/reproducibility analysis using wCV, a threshold of 10 % was used to consider the features as good and suitable for patient analysis, the strictest in the literature [26-29]. Vials using 40 % and 50 % were excluded from these analyses.

Additional statistical tests were calculated with the pingouin library v0.3.11 [30]. Differences in mean ADC between vial position in the phantom were tested with a Mann-Whitney U test. Differences in wCV between resampling and before/after ComBat were tested with a paired Wilcoxon’s signed-rank test. Friedman test was used to assess the difference between centres for each feature, before and after ComBat.

realignment. P-values $<0.05$ were interpreted as statistically significant.

Results

Accuracy

Two issues were revealed during the initial visit to two centres, one related to the sequence of scanner A, the other to the software of scanner E, which were resolved promptly. Scanner F images contained N/2 ghosting artefacts but it was deemed acceptable. The SNR results, as well as other QIBA DWI profile requirement measures, can be found in Supplementary Data 1, Table 2.

After resolving these issues, the quality control was successful for all scanners. 87 % vials with a polymer concentration of 0–30 % had $<5$ % difference from the ground truth, and all of them were under 10 % difference (Fig. 4). The median accuracy error for all vials was 1.3 % for scanner A, 2.7 % for scanner B, 0.9 % for scanner C, 2.6 % for scanner D, 4 % for scanner E, and 1.8 % for scanner F. The difference of the mean ADC between the inner and outer circle of vials was not statistically significant.

Repeatability

For the short-term repeatability, the mean ADC had a wCV between 0.1 % and 1 % depending on the scanner for both resampling schemes. The percentage of radiomic features that had good short-term repeatability (wCV $\leq 10$ %) over all the six scanners was 24 % (15/64 intensity features and 19/96 texture features) for a 0.9 mm resampled voxel spacing and 15 % (10/64 intensity features and 12/96 texture features) for a 4 mm resampled voxel spacing, with a significant difference in wCV between resampling schemes (p-value $<0.001$). Results for each scanner are shown in Fig. 5.

For the long-term repeatability assessed on scanner B, the mean ADC had a wCV of 1.3 % for both resampling schemes. 22 % (10/46) intensity features were repeatable for both resampling schemes, and 19 % (18/96) and 13 % (12/96) for texture features with a resampled voxel spacing of 0.9 mm and 4 mm, respectively. Those results are slightly worse than the short-term repeatability results for scanner B. The difference between resampling on wCV was significant (p-value $<0.05$).
Reproducibility

Mean ADC had a wCV of 2.7 % for the two resampling approaches. The percentage of radiomic features presenting good reproducibility (wCV ≤ 10 %) was 13 % (10/46 intensity features and 8/96 texture features) for a resampling 0.9 mm and 11 % (10/46 intensity features and 6/96 texture features) for a resampling 4 mm, with a significant difference of wCV between resampling (p-value < 0.01). Only five categories (LocInt, Stats, GLCM, GLRLM and GLDZM) had reproducible features. All features for which good reproducibility was observed also showed good long-term repeatability.

ComBat

Without harmonization, 18 out of 142p-values calculated when testing the difference between centres for each feature were greater than 0.05. After ComBat, this number increased to 142p-values greater than 0.05, indicating a scanner effect was present but could be resolved. The wCV improved significantly (p-value < 0.001) compared to the reproducibility without ComBat. However, the percentage of radiomic features presenting good reproducibility (wCV ≤ 10 %) was still low, with 15 % (10/46 intensity features and 12/96 texture features) for a 0.9 mm isotropic voxel spacing and 14 % (10 intensity and texture features) for a 4 mm isotropic voxel spacing, with no significant difference of wCV between resampling schemes (p-value = 0.23). A good realignment
between centres could be observed qualitatively (Fig. 6). Variation greater than 10%, however, still existed for most features, with only 4 texture features (for both resampling schemes) passing the 10% threshold in addition to the features already reproducible without ComBat.

All the wCV results for repeatability and reproducibility before and after ComBat can be found in the Supplementary Material 2.

Discussion

The first objective of this work was to perform a quality assurance using the NIST/QIBA diffusion phantom of DWI as part of the REGINA

---

**Fig. 6.** Realignment of (a) NGLDM_LDE and (b) GLCM_diffVar features using ComBat to remove the scanner effect. Abbreviation: wCV = within-subject Coefficient of Variation.
The quality assurance highlighted different issues in two scanners. The mean ADC computed for scanner E had an accuracy error of 25% for the central vial (0% polymer) when it was computed with the accompanying manufacturer’s software. Nevertheless, when calculating the ADC map with our in-house software, a mean ADC in the correct range was observed. This encourages computing the ADC maps on a single software platform, especially in the context of a clinical trial. The second issue was due to numeric overflow that appeared in low b-value images (from 0 to 400 s/mm$^2$) in scanner A for the phantom acquisition. It disappeared after changing the receiver image scaling from three to one, and the mean ADC in each vial was more consistent with the manufacturer’s values. This problem does not occur in patients, likely due to the higher concentration of protons, and thus signal, in the phantom. For this reason, the sequences were only adjusted for phantom acquisition.

The accuracy and the short- and long-term repeatability and reproducibility of the mean ADC were good for the vials with 0% to 30% polymer concentration, comparable with results in literature [32,35–42]. The accuracy of the vials with higher polymer concentration was poorer, as was already observed in other studies using the NIST/QIBA diffusion phantom as well [35,37,42]. This behaviour could be due to the lower diffusivity present in those vials. It was also worse for scanners E and F. The b0 images from these two scanners had the lowest SNR, possibly due to a higher bandwidth and a lower number of signal averages. Insufficient SNR may have influenced the assessment of the ADC in the highly concentrated vials. Images of scanner F also presented a N/2 ghosting artefact which could be caused by eddy currents. The presence of this artefact explains the substantial outlier for the outer circle 50% vials. The ADC ranges of these concentrations are however below what is observed in human tissue (≤0.5 x10$^{-3}$ mm$^2$/s) [11]. It is also mentioned in literature that ADC values can be heavily influenced by the off-set position of the ROI in the magnet because of the gradient nonlinearity present in each system [43–45]. However, this influence was not applicable for the ROI position in the phantom in this study.

Radiomic features from the ADC maps were less robust than the mean ADC. Despite not moving the phantom from the table between scans, very few features were repeatable. By doing a long-term repeatability analysis with one week interval, the results were not getting a lot worse compared to the short-repeatability analysis for scanner B. Intensity and GLCM categories gave the best results, comparable to previous studies [28,46]. More repeatable features were achieved when resampling to a smaller voxel size, as was also reported in a PET/CT phantom study [47]. To our knowledge, Dreher et al [48] is the only group who also assessed radiomic feature repeatability in an ADC map on a phantom. They showed high repeatability in the features, but any comparison with our results would be inappropriate due to large differences in terms of study design (i.e., different phantom and repeatability indicator).

By acquiring diffusion phantom data with six different scanners, we also showed that the ComBat procedure was able to realign radiomic feature distributions for ADC maps. The high variation in radiomic features between scanners was improved, but this was not sufficient for most of the features to pass below the pre-defined threshold (<10%). These results indicate that ComBat cannot overcome variability in all features, more so if there is a big discrepancy between scanners as was shown in Fig. 6b. Assessing the success of the harmonisation should not be only based on statistical tests, but also on some variation indicator as wCV, and also by looking at the data as closely as possible. Even if ComBat does not need phantom experiments, a phantom study permits to control which features can be harmonised with ComBat, as was also performed in a previous study using phantom CT images [19], and to select only the “ComBatable” ones for future clinical studies using that method. Other harmonisation solutions could be further explored, e.g. normalisation through deep learning, which is increasingly being studied [49,50].

A limitation of this study is the use of the NIST/QIBA diffusion phantom for radiomic analysis, as it is not a heterogeneous phantom. As shown by Mackin et al. [51] for CT, different textural patterns are not affected in the same way by differences in scanners, and the design of the phantom has an impact on the reproducibility that is found between different scanners [52]. Our results for texture features on a homogeneous phantom are therefore not directly transferable to patients, but are nevertheless a preliminary opportunity for investigating radiomic features in DWI. Efforts to create heterogeneous phantoms for the standardisation of radiomic features are ongoing [53]. It should also be noted that the within-subject coefficient of variation, which was used in this study to evaluate repeatability and reproducibility, lacks the assessment of variation between different ROIs. This means that features with low variation between scanners may also show low variation between patients, and would then not be useful in a radiomic model. The use of wCV in combination with another indicator that would quantify inter-scanner variability relative to inter-tumor variability (e.g. the intraclass correlation coefficient) would be preferable if there is some variation between ROIs texture corresponding to what might be observed in patients. Finally, different field strength systems were not included in this study as all scanners participating in the REGINA trial were at 3 T. Previous studies [32,36,38,41] showed however invariance of the ADC map to field-strength, as long as b-values were not too high [36].

Conclusion

As part of the ongoing REGINA trial, accuracy, repeatability and reproducibility of mean ADC were assessed using the NIST/QIBA diffusion phantom across six MRI systems. This quality assurance was a success for all centres, and showed that mean ADC could be largely standardised to be used for prediction purposes in the trial. Radiomic features are however still fragile even in images with good standardisation and post-processing harmonisation.

Declaration of Competing Interest

The radiology and medical physics departments of Institut Jules Bordet has a research agreement with Siemens. The authors declare that they have no other known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We wish to thank the centres participating in this study and who welcomed us for the phantom measurements in their radiology department: AZ Groeninge (Kortrijk, Belgium), Centre Hospitalier Régional Sambre et Meuse (Namur, Belgium), Erasme (Bruxelles, Belgium), Grand Hôpital de Charleroi (Charleroi, Belgium).

Funding

Zelda Paquier has a research grant from “L’Association Jules Bordet “. The REGINA trial is funded by Bayer.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org.
References


