Inter-observer reproducibility of quantitative dynamic susceptibility contrast and diffusion MRI parameters in histogram analysis of gliomas

Hildebrand Dijkstra¹, Paul E Sijens¹, Anouk van der Hoorn¹ and Peter Jan van Laar¹,²

Abstract
Background: Dynamic-susceptibility contrast and diffusion-weighted imaging are promising techniques in diagnosing glioma grade.

Purpose: To compare the inter-observer reproducibility of multiple dynamic-susceptibility contrast and diffusion-weighted imaging parameters and to assess their potential in differentiating low- and high-grade gliomas.

Material and Methods: Thirty patients (16 men; mean age = 40.6 years) with low-grade (n = 13) and high-grade (n = 17) gliomas and known pathology, scanned with dynamic-susceptibility contrast and diffusion-weighted imaging were included retrospectively between March 2006 and March 2014. Three observers used three different methods to define the regions of interest: (i) circles at maximum perfusion and minimum apparent diffusion coefficient; (ii) freeform 2D encompassing the tumor at largest cross-section only; (iii) freeform 3D on all cross-sections. The dynamic-susceptibility contrast curve was analyzed voxelwise: maximum contrast enhancement; time-to-peak; wash-in rate; wash-out rate; and relative cerebral blood volume. The mean was calculated for all regions of interest. For 2D and 3D methods, histogram analysis yielded additional statistics: the minimum and maximum 5% and 10% pixel values of the tumor (min5%, min10%, max5%, max10%). Intraclass correlations coefficients (ICC) were calculated between observers. Low- and high-grade tumors were compared with independent t-tests or Mann–Whitney tests.

Results: ICCs were highest for 3D freeform (ICC = 0.836–0.986) followed by 2D freeform (ICC = 0.854–0.974) and circular regions of interest (0.141–0.641). High ICC and significant discrimination between low- and high-grade gliomas was found for the following optimized parameters: apparent diffusion coefficient (P < 0.001; ICC = 0.641; mean; circle); time-to-peak (P = 0.015; ICC = 0.986; mean; 3D); wash-in rate (P = 0.004; ICC = 0.826; min10%; 3D); wash-out rate (P < 0.001; ICC = 0.860; min10%; 2D); and relative cerebral blood volume (P < 0.001; ICC = 0.961; mean; 3D).

Conclusion: Dynamic-susceptibility contrast perfusion parameters relative cerebral blood volume and time-to-peak yielded high inter-observer reproducibility and significant glioma grade differentiation for the means of 2D and 3D freeform regions of interest. Choosing a freeform 2D method optimizes observer agreement and differentiation in clinical practice, while a freeform 3D method provides no additional benefit.

Keywords
Diffusion magnetic resonance imaging, glioma, reproducibility of results, perfusion

Date received: 26 September 2018; accepted: 28 April 2019

Introduction
Grading of gliomas, the most common primary brain tumor (1,2), is important since the prognosis and adjuvant therapy after surgery differ according to tumor grade (3,4). Conventional imaging parameters, such as contrast enhancement on T1-weighted magnetic resonance imaging (MRI), do not provide a reliable indicator of tumor grade (5,6). Diffusion-weighted imaging (DWI)

¹Department of Radiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
²Department of Radiology, Ziekenhuis Groep Twente, Almelo-Hengelo, The Netherlands

Corresponding author:
Hildebrand Dijkstra, University Medical Center Groningen, Department of Radiology, EB44, PO Box 30001, 9700 RB Groningen, The Netherlands.
Email: h.dijkstra01@umcg.nl
and perfusion-weighted imaging (PWI) provide a plethora of additional parameters to overcome the shortcomings of conventional MRI. Differentiation between low and high glioma tumor grades has been demonstrated with dynamic susceptibility-weighed (DSC)-MRI, yielding the relative cerebral blood volume (rCBV) (7–11). Additionally, the additional value has been documented of dynamic contrast-enhanced (DCE)-MRI derived parameters such as blood plasma volume (Vp) (7,8), volume transfer constant (Ktrans) (7,8,10,12), rate constant (Kep) (9,10), and normalized tumor cerebral blood flow (nCBF) (10,12,13). In addition, quantitative DWI expressed in the apparent diffusion coefficient (ADC) shows significant differences between low- and high-grade gliomas (8,13,14).

In our opinion, a systematic review of the robustness of the various quantitative measures of perfusion and diffusion in brain tumors, and of the impact of the methodology of region of interest (ROI) selection, is lacking. The purpose of this study was therefore to compare inter-observer reproducibility of DSC-MRI and DWI-derived tumor characteristics in 30 gliomas. Three observers analyzed the MRIs according to the common practice of manual delineation of the tumor as visualized on the respective PWI and DWI series, and of the equally widespread use of small circular ROIs to simulate fast evaluation in clinical practice. Interclass correlation coefficients (ICC) were determined and differentiation between low- and high-grade glioma was assessed for each parameter.

**Material and Methods**

**Study population**

The protocol of this retrospective study was approved by the hospital’s institutional review board and patients’ written informed consent was waived. Between March 2006 and March 2014, patients with suspected brain lesions and scheduled for MRI were recruited. Inclusion criteria were: MRI scan protocol including both DSC and DWI series; and lesion diameter ≥2 cm on MRI. This resulted in the inclusion of 30 consecutive patients (16 boys/men; mean age = 40.6 years; age range = 5–78 years) with 30 lesions: 13 were low-grade and 17 were high-grade. Tissue samples were obtained in all patients using neuro-navigational guided core biopsy (n = 13) or surgical resection (n = 17).

**MR protocols**

All participants were examined on a 1.5-T MRI system (Magnetom Sonata, Siemens Medical Solutions, Erlangen, Germany) using the eight-channel head coil. Diffusion tensor imaging (b = 0 and 1000 s/mm²) was scanned with the following: single shot echo-planar imaging (EPI); TR = 5800 ms; TE = 92 ms; flip angle (FA) = 90°; slice thickness = 3 mm; slice gap = 3 mm; field of view (FOV) = 230 × 230 mm; matrix = 256 × 256; bandwidth = 1565 Hz/pixel; no averages. Diffusion gradients (40 mT/m) in 12 directions; 37 transverse slices; acquisition time = 2.2 min. ADC maps were calculated on the scanner. Finally, DSC perfusion: 70 transversal time series (EPI, epfid2d1); temporal resolution = 1.0 s; gadoterate meglumine (Gd-DOTA, Dotarem, Guerbet, 0.1 mmol/kg): TR = 1000 ms; TE = 40 ms; FA = 60°; slice thickness = 6 mm; slice gap = 7.8 mm; FOV = 230 × 230 mm; matrix = 256 × 256; and bandwidth = 1500 Hz/pixel.

**Defining the lesions**

All patients’ images were clinically interpreted by neuroradiologists with >10 years of experience. Three independent observers then analyzed the lesions: an expert neuro-radiologist with >10 years of experience (Observer 1); a radiology resident (Observer 2), and a medical physicist without any experience in brain imaging (Observer 3). All lesions were analyzed off-line (Matlab 2014a, The Mathworks, Natick, MA, USA). Three different methods were used to define the ROI. In method 1, a circle was drawn independently and with varying diameters to include the lowest pixel values on the ADC map and highest pixel values on the DSC time series at the time-point with the highest contrast enhancement (Fig. 1). These small ROIs had an average diameter of 9.1 mm (range = 3.7–17.8 mm). The lowest or highest region in the tumor was located in the tumor by visual evaluation and quantitative measurements of the pixel values. The diameter of the circle was adjusted according to the estimation of the observer. Additionally, freeform ROIs were drawn to encompass the whole tumor at the largest MRI cross-section only (method 2: 2D; Fig. 2; area = 184–5140 mm²) or summed MRI cross-sections (method 3: 3D; Fig. 3; area = 577–46,121 mm²).

**DSC-MRI quantification**

Voxelwise analysis of the DSC curve was performed to obtain the following parameters: maximum contrast enhancement (MCE); time-to-peak (TTP); wash-in rate (W_in); wash-out rate (W_out); and the relative cerebral blood volume (rCBV).

Three time-points were defined: t onset = 10 s after start of the sequence; t max = time at the peak of the signal; and t end = last time point of DSC curve. W_in was calculated by the slope of a linear fit of the signal between t onset and t max. W_out was calculated by the slope of a linear fit of the signal between t max and
tend. MCE was defined as \((S_{\text{max}} - S_{\text{onset}})/S_{\text{onset}}\), where \(S_{\text{max}}\) is the signal intensity at maximum contrast enhancement and \(S_{\text{onset}}\) is the signal without contrast enhancement at \(t_{\text{onset}}\). TTP is the time between \(t_{\text{onset}}\) and \(t_{\text{max}}\). The rCBV was calculated by numerical integration of \(\Delta R^2 = -\ln(S/S_{\text{onset}})/TE\) (15), where S is the signal intensity. \(\Delta R^2\) was not divided by the result for contralateral or centrum ROIs in order to obtain ICC values not complicated by additional ROI settings.

**Comparisons and statistics**

Statistical analyses were performed using SPSS (SPSS 23, Chicago, IL, USA). All data were tested for normality using Shapiro–Wilk tests. The ICCs between observers was calculated for all three methods using a two-way mixed model, single measures after applying a log transformation to normalize the PWI data. Low- and high-grade lesions were compared using independent t-tests for ADC data and Mann–Whitney tests for PWI data. The mean of all voxels in the circular ROI was taken to define the ADC, rCBV, MCE, TTP, \(W_{\text{in}}\), and \(W_{\text{out}}\) for each lesion for all methods. For the freehand 2D and 3D methods, next to the mean, four additional statistics were calculated by histogram analysis for ADC, rCBV, MCE, TTP, \(W_{\text{in}}\), and \(W_{\text{out}}\): the 5% and 10% lowest values in the lesion (min5% and min10%) and the 5% and 10% highest values in the

---

**Fig. 1.** Method 1, conventional small circle ROI. A 54-year-old man with a glioblastoma (high-grade). Circles of free diameter were drawn independently to include the lowest pixel values on the ADC map (left) and the highest pixel value on the PWI map (right). ROI: region of interest; ADC: apparent diffusion coefficient; PWI: perfusion-weighted imaging.

**Fig. 2.** Method 2, 2D freeform ROI method. A 54-year-old man with a glioblastoma (high-grade). A freeform ROI encompassing the whole tumor was drawn at the largest tumor cross-section, independently on the ADC map (left) and on the PWI map (right). ROI: region of interest; ADC: apparent diffusion coefficient; PWI: perfusion-weighted imaging.
lesion (max5% and max10%) for each parameter. For both freehand methods, the best statistic was determined based on the highest ICC and best differentiation (highest significance) between low- and high-grade lesions. For the conventional small circle ROI, only the mean of the circle was used. A $P_{\text{two-sided}} < 0.05$ was considered a statistically significant difference for all tests.

Results
Histopathology of 30 lesions resulted in 13 low-grade and 17 high-grade gliomas: 12 astrocytoma (Grade II); one oligoastrocytoma (Grade II); 13 glioblastoma (Grade IV); three anaplastic astrocytoma (Grade III); and one anaplastic oligodendroglioma (Grade III).

Inter-observer reproducibility
The Shapiro–Wilk tests showed that within the chosen ROIs the pixel value distributions for ADC and five PWI parameters (rCBV, MCE, TTP, $W_{\text{in}}$, $W_{\text{out}}$) were generally normal after the log transformation, so ICC were assessed for the mean values (Table 1). Low-grade gliomas presented with sufficient signal change on both ADC and PWI maps to accurately draw the contours, as demonstrated in the example in
Fig. 4. For all six parameters, inter-observer agreement of the mean values was better for 2D and 3D freeform ROIs (ICC = 0.836–0.986) than for circular ROIs (ICC = 0.141–0.641).

**Differentiation of low- and high-grade gliomas**

For PWI, both freeform 2D and 3D methods yielded significant differences between low- and high-grade gliomas when comparing the mean, max5%, or max10%
of either TTP (\(P \leq 0.015\)) or rCBV (\(P \leq 0.001\)), as documented in Table 2. For \(W_{in}\) (\(P \leq 0.004\)) or \(W_{out}\) (\(P \leq 0.007\)) the min5\% and min10\% resulted in significant differences for both methods. MCE as assessed by any ROI method yielded no significant differences between low and high grade (\(P \geq 0.127\)).

For DWI, the ADC differentiated significantly (\(P \leq 0.048\)) between low- and high-grade gliomas when choosing the minimum 5\% or 10\% of pixel values of the 2D and 3D freeform ROIs; however, reproducibility was low (ICC = 0.532–0.622). In addition, for the circular ROI method, the ADC was the only parameter yielding significant differences with reasonable inter-observer agreement (\(P < 0.001\), ICC = 0.641).

Overall, the best ICC with significant grade differentiation could be obtained for the TTP parameter measured on a freeform 3D (ICC = 0.986, \(P = 0.015\)) and using the mean as measurement statistic (Table 3), closely followed by the result from a freeform 2D ROI (ICC = 0.974, \(P = 0.004\)).

**Discussion**

This study showed that inter-observer agreement increased when gliomas were encompassed with a 2D

---

**Table 1. Inter-observer reproducibility.**

<table>
<thead>
<tr>
<th>Method</th>
<th>Statistic</th>
<th>ADC</th>
<th>MCE</th>
<th>TTP</th>
<th>(W_{in})</th>
<th>(W_{out})</th>
<th>rCBV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circle</td>
<td>Mean</td>
<td>0.641</td>
<td>0.204</td>
<td>0.314</td>
<td>0.325</td>
<td>0.141</td>
<td>0.405</td>
</tr>
<tr>
<td>2D</td>
<td>Mean</td>
<td>0.854</td>
<td>0.890</td>
<td>0.974</td>
<td>0.876</td>
<td>0.892</td>
<td>0.923</td>
</tr>
<tr>
<td></td>
<td>Min5%</td>
<td>0.532</td>
<td>0.839</td>
<td>0.923</td>
<td>0.552</td>
<td>0.620</td>
<td>0.861</td>
</tr>
<tr>
<td></td>
<td>Min10%</td>
<td>0.554</td>
<td>0.844</td>
<td>0.926</td>
<td>0.677</td>
<td>0.860</td>
<td>0.705</td>
</tr>
<tr>
<td></td>
<td>Max5%</td>
<td>0.639</td>
<td>0.905</td>
<td>0.860</td>
<td>0.888</td>
<td>0.922</td>
<td>0.814</td>
</tr>
<tr>
<td></td>
<td>Max10%</td>
<td>0.749</td>
<td>0.903</td>
<td>0.874</td>
<td>0.890</td>
<td>0.926</td>
<td>0.842</td>
</tr>
<tr>
<td>3D</td>
<td>Mean</td>
<td>0.836</td>
<td>0.897</td>
<td>0.986</td>
<td>0.911</td>
<td>0.935</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td>Min5%</td>
<td>0.622</td>
<td>0.904</td>
<td>0.878</td>
<td>0.743</td>
<td>0.745</td>
<td>0.708</td>
</tr>
<tr>
<td></td>
<td>Min10%</td>
<td>0.588</td>
<td>0.905</td>
<td>0.873</td>
<td>0.826</td>
<td>0.769</td>
<td>0.507</td>
</tr>
<tr>
<td></td>
<td>Max5%</td>
<td>0.755</td>
<td>0.912</td>
<td>0.927</td>
<td>0.923</td>
<td>0.973</td>
<td>0.908</td>
</tr>
<tr>
<td></td>
<td>Max10%</td>
<td>0.821</td>
<td>0.907</td>
<td>0.935</td>
<td>0.929</td>
<td>0.966</td>
<td>0.923</td>
</tr>
</tbody>
</table>

Three different methods were used to define the ROI (circle, 2D, and 3D). For each method, the six MRI parameters were calculated by using the mean or 5\% or 10\% lowest or respectively highest values in the ROI. The intraclass correlation coefficient (ICC) was determined for all combinations.

ROI: region of interest; ADC: apparent diffusion coefficient; MCE: maximum contrast enhancement; TTP: time-to-peak; \(W_{in}\): wash-in rate; \(W_{out}\): wash-out rate; rCBV: relative cerebral blood volume.
freeform ROI compared to the conventional small circle ROI, while 3D freeform ROIs yielded no additional benefits.

Our finding of best reproducibility of whole tumor freeform ROI for the five DSC parameters, as well as ADC, compared with small circles of variable size positioned at highest PWI map and lowest ADC, is hard to compare with previous studies. Some only used small circular or elliptical ROIs of 20–40 mm² (9–11,16) compared to others that only used freeform ROIs to delineate tumors (7,8,12–14), sometimes supplemented with “minimum ADC” (8) or lower and pixel value means for ADC (12). In just one study, both small presumably circular “hot spot” ROIs of 25–30 mm² were set for each DSC parameter and compared with freeform ROI (7). Their conclusion that “hot spot ROI” analysis in DSC showed the best correlation with grading of gliomas appears to be at odds with our result. No assessment of reproducibility by ICC comparison was provided in that comparatively small study of nine low-grade gliomas and 17 high-grade ones (7).

In our case, reproducibility and differentiation was best for rCBV and TTP which apparently reflected the comparatively high ICC values obtained for these two DSC parameters in 2D and 3D freeform ROI. That the freehand mean and max5–10% offered the best differentiation for these parameters leads to the interpretation that the highest blood volume and enhancement delay encountered in glioma are most characteristic for tumor grade. In contrast, the rates of W in and W out offered best differentiation at freeform ROI min5–10%. This fits with the knowledge that increased transfer rates are typical for malignancy (7–10,12).

<table>
<thead>
<tr>
<th>Table 2. Differentiation between low- and high-grade gliomas.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
</tr>
<tr>
<td>Circle</td>
</tr>
<tr>
<td>2D</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>3D</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Three different methods were used to define the ROI (circle, 2D and 3D). For each method the six MRI parameters were calculated by using the mean or 5% or 10% lowest or respectively highest values in the ROI. The significance of the differences between low- and high-grade was determined for all combinations.

*P < 0.05.

ROI: region of interest; ADC: apparent diffusion coefficient; MCE: maximum contrast enhancement; TTP: time-to-peak; W in: wash-in rate; W out: wash-out rate; rCBV: relative cerebral blood volume.

<table>
<thead>
<tr>
<th>Table 3. Comparison of optimal ROI settings and statistic.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
<tr>
<td>Conventional circle</td>
</tr>
<tr>
<td>2D freeform</td>
</tr>
<tr>
<td>3D freeform</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

For the 2D and 3D freeform method the best statistic is indicated based on the highest ICC and best differentiation between low- and high-grade lesions. For the conventional circle ROI only the mean of the circle was used.

*P < 0.05.

ROI: region of interest; ADC: apparent diffusion coefficient; MCE: maximum contrast enhancement; PWI: perfusion-weighted imaging; ROI: region of interest; TTP: time-to-peak; W in: wash-in rate; W out: wash-out rate; rCBV: relative cerebral blood volume.
Note that the most basic PWI parameter, MCE, though highly reproducible at especially 3D-freeform ROI max5% and max10% (ICC = 0.907–0.912), did not differentiate between low- and high-grade gliomas ($P > 0.7$). This is in line with a previous study showing that rCBV is a stronger predictor of glial tumor grade than the degree of enhancement (5).

The DWI parameter ADC was the only one providing significant grade differentiation along with reasonable reproducibility (ICC > 0.641) for the circular ROI, despite its lower ICC values compared with freeform 2D and 3D ROI. The worse result in terms of grade differentiation when just taking the mean of a freeform 2D or 3D ROI indicates that low ADC values are probably clustered in a limited part of the tumor and thus averaged out when taking the mean of the entire ROI. This is confirmed by the finding by Perez et al., that the ADC was worse in differentiating between glioma grades than a couple of DSC parameters when just taking the mean of the ROI (8). However, when taking the minimal ADC value of the ROI, they observed significant differentiation between glioma grades by the ADC, which is in accordance with our results. In addition, other compounds, such as fatty acids and calcifications, potentially present in oligodendrogloma, can affect the minimal ADC of a tumor which can be avoided by a manually selected area of low diffusion with a circle ROI (17,18).

Results obtained from multiple summed tumor slices (3D freeform ROI) were similar to those obtained with 2D freeform ROI, both in terms of reproducibility and in terms of tumor grade differentiation. It was expected that the circular ROIs would result in a lower reproducibility than the freeform methods because the circular diameters and locations were set differently by different observers. Another fair assumption is that the freeform 3D method, including the whole tumor, would result in better reproducibility than the freeform 2D method. Due to the heterogeneity of the tumor, especially for the high-grade glioma, one slice analysis, as applied in the freeform 2D method, might not reflect the features of the whole tumor. Our results of 3D analysis were, however, similar to the 2D analysis. The statistics such as min5% and max5% are already sensitive for the small number of pixels deviating from the mean in the 2D freeform slice. For tumor characterization, the 3D analysis therefore had no incremental value beyond the 2D freeform region. This might indicate that in the analysis of glioma, evaluation of a single 2D freeform region of the largest cross-section of the tumor might be sufficient, especially considering the time needed to draw the regions.

A limitation in our study was that only 30 patients were included. However, the results of this study documented that this was sufficient, considering that the reported observations generally meet high levels of significance ($P < 0.001$). Another issue was that ROIs were set independently for the DWI analysis and for the DSC analyses. This had the limitation that it made no sense to associate the ADC parameter values to those derived from DSC. Apart from that, it might even be best to set “hot spot” ROIs independently for each DSC parameter to account for tumor heterogeneity (7,9,10). Another limitation was that our analyses were performed off-line by Matlab. More consistent results could be obtained by more or less automated software packages such as Olea or NordicIce, although inter-software reproducibility has been found to be poor (16). The three observers in our study were of mixed levels of experience with neuroradiology, which turned out to be no real limitation as evidenced by the tabulated ICC values of up to 0.99. It has been shown that combining perfusion and DWI parameters can improve glioma grading accuracy (7,9,13,14), but exploration of that was beyond the scope of the present study aiming at the effects of ROI definition on the robustness of separate DSC parameters and ADC in relation to their potential to differentiate between low- and high-grade gliomas. Another topic, not explored in this study, with standard clinical DWI acquisitions of just 2 b values, is that with biexponential intravoxel incoherent motion (IVIM) analysis of DWI data differentiation between low- and high-grade glioma and other brain tissue types may improve (19).

In conclusion, DSC perfusion parameters rCBV and TTP yielded high inter-observer reproducibility and significant glioma grade differentiation for the means of 2D and 3D freeform ROIs. However, the more time-consuming alternative of 3D freeform ROI did not provide addition benefit. The small circular ROI yielded only significant differences with reasonable reproducibility for the DWI parameter ADC.

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

**ORCID iD**

Hildebrand Dijkstra https://orcid.org/0000-0001-6664-902X

Paul E Sijens https://orcid.org/0000-0003-4184-3736
References