Multisite Concordance of DSC-MRI Analysis for Brain Tumors: Results of a National Cancer Institute Quantitative Imaging Network Collaborative Project


AJNR Am J Neuroradiol 2018, 39 (6) 1008-1016
doi: https://doi.org/10.3174/ajnr.A5675
http://www.ajnr.org/content/39/6/1008
Multisite Concordance of DSC-MRI Analysis for Brain Tumors: Results of a National Cancer Institute Quantitative Imaging Network Collaborative Project


ABSTRACT

BACKGROUND AND PURPOSE: Standard assessment criteria for brain tumors that only include anatomic imaging continue to be insufficient. While numerous studies have demonstrated the value of DSC-MR imaging perfusion metrics for this purpose, they have not been incorporated due to a lack of confidence in the consistency of DSC-MR imaging metrics across sites and platforms. This study addresses this limitation with a comparison of multisite/multiplatform analyses of shared DSC-MR imaging datasets of patients with brain tumors.

MATERIALS AND METHODS: DSC-MR imaging data were collected after a preload and during a bolus injection of gadolinium contrast agent using a gradient recalled-echo–EPI sequence (TE/TR = 30/1200 ms; flip angle = 72°). Forty-nine low-grade (n = 13) and high-grade (n = 36) glioma datasets were uploaded to The Cancer Imaging Archive. Datasets included a predetermined arterial input function, enhancing tumor ROIs, and ROIs necessary to create normalized relative CBV and CBF maps. Seven sites computed 20 different perfusion metrics. Pair-wise agreement among sites was assessed with the Lin concordance correlation coefficient. Distinction of low- from high-grade tumors was evaluated with the Wilcoxon rank sum test followed by receiver operating characteristic analysis to identify the optimal thresholds based on sensitivity and specificity.

RESULTS: For normalized relative CBV and normalized CBF, 93% and 94% of entries showed good or excellent cross-site agreement (0.8 ≤ Lin concordance correlation coefficient ≤ 1.0). All metrics could distinguish low- from high-grade tumors. Optimum thresholds were determined for pooled data (normalized relative CBV = 1.4, sensitivity/specificity = 90%:77%; normalized CBF = 1.58, sensitivity/specificity = 86%:77%).

CONCLUSIONS: By means of DSC-MR imaging data obtained after a preload of contrast agent, substantial consistency resulted across sites for brain tumor perfusion metrics with a common threshold discoverable for distinguishing low- from high-grade tumors.

ABBREVIATIONS: AIF = arterial input function; HGG = high-grade glioma; LCCC = Lin concordance correlation coefficient; LGG = low-grade glioma; NAWM = normal-appearing white matter; nCBF = normalized cerebral blood flow; nRCBV = normalized relative cerebral blood volume; SN = sensitivity (in the context of receiver operating characteristic analysis); SP = specificity (in the context of receiver operating characteristic analysis); TI+ C = postcontrast TIWI

Current Response Assessment in Neuro-Oncology criteria,1 which include anatomic imaging only, are often insufficient for determining tumor response and progression after therapy. Frequently, imaging changes on postcontrast MR imaging or FLAIR are not the result of increased tumor activity but rather a consequence of the treatment itself. These changes, which can include edema, inflammation, or increased vascular or blood-brain barrier permeability, make it difficult to distinguish true

This work was supported by the National Institutes of Health/National Cancer Institute: U01 CAI66040 (T.L.C., B.H.), U01 CA154601 (X.D., Y.-F.Y., J.K.-C.), U01 CA160045 (B.E., P.K.), U01 CA 183848 (Y.C., M.P.A.), R01 CA158079 (L.B., C.C.Q.), R01 CA082500 (K.M.S., M.A.P., S.D.R.), R50 CA211270 (M.M.), and P01 CA085878, U24CA180927, U24CA180918 (J.K.-C.).


Please address correspondence to Kathleen M. Schmainda, PhD, Medical College of Wisconsin, Department of Radiology, 8701 W Watertown Plank Rd, Milwaukee, WI 53226; e-mail: kathleen@mcw.edu

* Indicates open access to non-subscribers at www.ajnr.org

http://dx.doi.org/10.3174/ajnr.A5675
tumor progression from treatment response.2-6 Because of these difficulties, patients are often precluded from switching to potentially more effective therapies within treatment windows of 3–5 months.1 Clearly, better indications of tumor response that are not confounded by these treatment adverse effects are needed.

Perfusion MR imaging methods, which have repeatedly demonstrated the ability to provide biologically relevant information for treatment management, have the potential to overcome these limitations. For brain perfusion, the DSC-MR imaging methods have been most commonly used. With DSC-MR imaging, T2- or T2*-weighted images are acquired with high temporal resolution during a bolus administration of a gadolinium contrast agent.7 The derived relative CBV (rCBV) maps have demonstrated the ability to predict tumor grade6,8 and survival,10 distinguish treatment effects from recurrent tumor,11,12 and predict response to antiangiogenic therapy more reliably than standard MR imaging.13-15

Despite this promise, the translation of DSC-MR imaging for routine clinical use has been hindered by a lack of consistency in the methods used and the rCBV values reported to make the noted distinctions. However, often a threshold determined for one purpose, such as distinguishing low- from high-grade tumor,16 is used for another purpose such as predicting outcomes.17 Consequently, the present confusion may be due to the lack of well-defined studies performed under carefully controlled conditions that test a specific outcome. This study addresses these limitations by providing carefully curated DSC-MR imaging datasets of low-grade glioma (LGG) and high-grade glioma (HGG) to multiple sites that participate in the National Cancer Institute Quantitative Imaging Network. With this approach, variations in image acquisition and preprocessing are eliminated and postprocessing methods can be directly compared in their ability to distinguish LGGs from HGGs. In addition, the threshold for this distinction can be identified both for each individual site and as a consensus recommendation.

MATERIALS AND METHODS

Patients

All subjects recruited from a single site provided informed written consent according to institutional review board policy. Subjects with histologically confirmed, newly diagnosed, and treatment-naive glial tumors who had preoperative DSC-MR imaging were included in this study. Subjects with purely oligodendroglial lesions were not included due to demonstrated differences in rCBV compared with astrocytic tumors.18 Due to the disparity in the number of subjects histologically diagnosed with low- and high-grade tumors, consecutive subjects with low-grade tumors between 2008 and 2014 and high-grade tumors from 2010 to 2014 were identified. Subjects were excluded if anatomic images were not available for lesion delineation or when DSC-MR imaging data were of poor quality.

Imaging

Studies were performed on 1.5T or 3T MR imaging systems. Standard precontrast MRIs were acquired including FLAIR (TE/TR/NEX/matrix = 126/9000 ms/2200–2500 ms/2/252 × 215) and T1WI spin-echo (TE/TR/NEX/matrix = 11 ms/650 ms/2/256), after which gadolinium contrast agent was administered. Either 0.05 mmol/kg of gadobenate dimeglumine (MultiHance; Bracco Diagnostics, Princeton, New Jersey) (n = 48) or 0.1 mmol/kg of gadodiamide (Omniscan; GE Healthcare, Piscataway, New Jersey) (n = 1) was administered before obtaining the postcontrast T1-weighted images. This initial gadolinium administration serves as a preload for the subsequent DSC-MR imaging acquisition. A preload results in more accurate rCBV because it reduces the T1 leakage effects that can confound the DSC-MR imaging data analysis.9,19-21 After collecting the postcontrast T1WI (T1+C), using parameters equivalent to the precontrast settings, we acquired DSC-MR imaging data (gradient recalled-echo–EPI: TR = 1100–1250 ms, TE = 30 ms, flip angle = 70°–80°, 12–17 slices, slice thickness = 4–5 mm, interslice gap = 0–1.5 mm, matrix = 96 × 96/128 × 128, FOV = 165–240 × 165–240 mm²). The gradient recalled-echo–EPI data were collected for 120 seconds with 0.1 mmol/kg of gadolinium injected at 60 seconds. For ease of coregistration of the DSC-MR images to anatomic images, a spin-echo “reference” scan (TE/TR/NEX/matrix = 14 ms/300 ms/1/256 × 192) was obtained using a slice prescription matching the more limited slice prescription of the DSC-MR imaging examination.

Central Preprocessing

The preprocessing workflow is schematized in Fig 1. All preprocessing was performed in Osirix Imaging Software (http://www.osirix-viewer.com) using the IB Delta Suite (Imaging Biometrics, Elm Grove, Wisconsin). Six datasets provided for each case included the following: 1) T1+C images; 2) DSC-MR imaging time-series co-registered to the T1+C images; 3) an arterial input function (AIF), which included 3 AIF locations in each file; 4) a whole-brain mask and ROIs of 5) normal-appearing white matter (NAWM) and 6) tumor. The NAWM was used to compute normalized rCBV (nRCBV) and normalized CBF (nCBF) maps.

The DSC-MR imaging volume was co-registered to the T1+C images via the reference scan obtained with the same slice prescription as the DSC-MR imaging. The AIF locations were determined semiautomatically using IB Neuro (Imaging Biometrics) with manual adjustments when necessary. An average signal generated from 3 pixels constituted the AIF. The whole-brain mask was made available to prevent disparity in values that could result from threshold variations that each software platform might use. Using the IB Delta Suite, we determined tumor ROIs from deltaT1 maps, which are standardized difference maps22 that enable clear visualization of enhancing lesions free of bright signal from blood products or proteinaceous material. Nonenhancing lesions, apparent as areas of dark signal on T1WI, were delineated by a neuroradiologist with >20 years of experience. Each ROI was created as a 16-bit binary DICOM file that included only whole voxels rather than contoured points. This approach prevents differences in the applied ROIs because each software platform applies different rules regarding whether a voxel is considered inside or outside an ROI. Anonymized datasets were uploaded to The Cancer Imaging Archive,23,24 where they were further vetted to ensure the compatibility of complete datasets for the analysis platform of each site. All sites were blinded to tumor grade.
Image and Statistical Analyses

Seven sites (1–7) using 7 different software platforms provided 20 different rCBV measurements (Table 1) and 12 different CBF measurements (Table 2). Details for each software platform are listed in the tables. Several sites used >1 platform or analysis method. When applicable, the analysis for measurements of standardized rCBV was grouped separately from nRCBV due to image-scale differences. Agreement between each pair of values was assessed by computing the Lin concordance correlation coefficient (LCCC). An LCCC $>0.8$ indicates good agreement, and LCCC $>0.9$ indicates excellent agreement.

The ability of each metric to distinguish LGG from HGG was determined using the Wilcoxon rank sum test, with $P<.05$ indicating significance. A receiver operating characteristic analysis was performed to identify the threshold that provides the optimal sensitivity (SN) and specificity (SP) to distinguish LGG from HGG. The DeLong test for comparing $\approx2$ receiver operating characteristic curves was used to determine whether there were differences in the ability of each measurement to classify tumors.

To determine “consensus” cutoff points, we created boxplots of the sum of SN and SP. Optimal thresholds were identified as those with maximum SN + SP mean, according to the Youden selection criteria with small variance. The random effects model was used to assess the reliability of measurements across sites and platforms. The reliability is quantified by the following:

Table 1: Teams and rCBV analysis methods

<table>
<thead>
<tr>
<th>Team/Entry No.</th>
<th>Software</th>
<th>Scaling</th>
<th>Leakage Correction</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>IB Neuro</td>
<td>NAWM</td>
<td>Yes</td>
<td>Modified BSW$^{21}$</td>
</tr>
<tr>
<td>B1</td>
<td>IB Neuro</td>
<td>NAWM</td>
<td>Yes</td>
<td>Modified BSW$^{21}$</td>
</tr>
<tr>
<td>C1</td>
<td>IB Neuro</td>
<td>NAWM</td>
<td>Yes</td>
<td>Modified BSW$^{21}$</td>
</tr>
<tr>
<td>B2</td>
<td>Matlab$^a$</td>
<td>NAWM</td>
<td>Yes</td>
<td>120-point TI (BSW$^{21}$)</td>
</tr>
<tr>
<td>B3</td>
<td>Matlab$^a$</td>
<td>AIF</td>
<td>Yes</td>
<td>120-point TI (BSW$^{21}$)</td>
</tr>
<tr>
<td>D1</td>
<td>nordicICE</td>
<td>NAWM</td>
<td>Yes</td>
<td>120-point TI (BSW$^{21}$)</td>
</tr>
<tr>
<td>D2</td>
<td>nordicICE</td>
<td>Population AIF</td>
<td>Yes</td>
<td>120-point TI (BSW$^{21}$)</td>
</tr>
<tr>
<td>E1</td>
<td>PGUI (LC)$^b$</td>
<td>NAWM</td>
<td>Yes</td>
<td>AUC of C(t) ratio for ROI/AIF</td>
</tr>
<tr>
<td>E2</td>
<td>PGUI (C)</td>
<td>NAWM</td>
<td>Yes</td>
<td>Bayesian parametric FT$^c$</td>
</tr>
<tr>
<td>E3</td>
<td>PGUI (PM LC)</td>
<td>NAWM</td>
<td>Yes</td>
<td>120-point TI (BSW$^{21}$)</td>
</tr>
<tr>
<td>D3</td>
<td>Slicer$^d$</td>
<td>NAWM</td>
<td>Yes</td>
<td>Modified Weisskoff$^{45,46}$</td>
</tr>
<tr>
<td>F1</td>
<td>FIAT$^e$</td>
<td>NAWM</td>
<td>Yes</td>
<td>Modified Weisskoff$^{45,46}$</td>
</tr>
<tr>
<td>G1</td>
<td>MCcbv$^f$</td>
<td>NA</td>
<td>Yes</td>
<td>Modified BSW$^{21}$.2006</td>
</tr>
<tr>
<td>A2</td>
<td>IB Neuro</td>
<td>NA (standardized)$^{47}$</td>
<td>No Leakage correction not used</td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>IB Neuro</td>
<td>NAWM</td>
<td>No Leakage correction not used</td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td>IB Neuro</td>
<td>NAWM</td>
<td>No Leakage correction not used</td>
<td></td>
</tr>
<tr>
<td>E4</td>
<td>PGUI</td>
<td>NAWM</td>
<td>No Leakage correction not used</td>
<td></td>
</tr>
<tr>
<td>E5</td>
<td>PGUI</td>
<td>NAWM</td>
<td>No Deconvolution of residue function (sSVD)</td>
<td></td>
</tr>
<tr>
<td>E6</td>
<td>PGUI</td>
<td>NAWM</td>
<td>No Deconvolution of residue function (oSVD)</td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>IB Neuro</td>
<td>N/A (standardized)$^{47}$</td>
<td>No Leakage correction not used</td>
<td></td>
</tr>
</tbody>
</table>

Note:—TI indicates trapezoidal integration; BSW, Boxerman-Schmainda-Weisskoff; NA, not applicable; FT, Fourier Transform; sSVD, standard singular value decomposition; oSVD, block circulant singular value decomposition; C(t) ratio, concentration time course.

$^a$MathWorks, Natick, Massachusetts.

$^b$Perfusion Graphical User Interface; MINDlab, Center of Functionally Integrative Neuroscience, Aarhus University Hospital Norrebrogade, Denmark.$^{44}$


$^d$Functional Image Analysis Tool (University of Michigan).

$^e$Mayo Clinic cerebral blood volume [Mayo Clinic, Rochester, Minnesota].$^{37}$

$^f$Schmainda Jun 2018 www.ajnr.org
lant singular value decomposition; FT, Fourier Transform.

Note:—SVD indicates singular value decomposition; sSVD, standard singular value decomposition; oSVD, block circulant singular value decomposition; FT, Fourier Transform.

Sixty-three subjects met inclusion criteria for this study, with 14 excluded for the following reasons: The contrast agent bolus was delayed during DSC acquisition, preventing capture of the postbolus steady-state signal (n = 4); the contrast agent bolus was injected too slowly and in an irregular pattern (n = 4); the contrast agent bolus was not present during acquisition of images (n = 2); there were severe ghosting and motion artifacts of DSC images (n = 2); and anatomic images were not available for lesion delineation (n = 2). Forty-nine coregistered LGG (n = 13) and HGG (n = 36) DSC-MR imaging datasets were preprocessed, anonymized, and uploaded to The Cancer Imaging Archive (Fig 1).

Tumor grade was confirmed with histopathology a median of 3 days (range, 0–4 days) following MR imaging. Examples of postprocessed datasets are shown in Fig 2. LCCC results are displayed in a matrix listing each nRCBV (Fig 3) or nCBF (Fig 4) entry on both the x- and y-axes. For tumor nRCBV, 75% of the entries showed excellent agreement with LCCC ≥ 0.9 and 19% with good agreement (0.80 ≤ LCCC < 0.90), leaving only 6% with poor concordance (LCCC < 0.80). The concordance was best for nRCBV values determined with leakage correction. For nCBF, only 59% had 0.90 ≤ LCCC ≤ 1.0 and 34% had 0.80 ≤ LCCC < 0.89.

For all software platforms, both nRCBV and nCBF showed statistically significant differences between LGG and HGG (Tables 3 and 4), with a mean nRCBV = 1.4 ± 0.13 and mean nCBF = 1.57 ± 0.24. The SN/SP for nRCBV ranged from 81%–97%/77%–85% and was slightly worse for nCBF with SN/SP = 64%–97%/69%–85%. By means of the DeLong test, no significant differences were found among the 18 nRCBV (P = .72) metrics to distinguish LGG from HGG. While differences among the nCBF metrics were borderline significant (P = .05), if the entry with the lowest area under the curve (0.658) was excluded, there was no significant difference between the remaining measures (P = .49). The DeLong test for the standardized rCBV showed no significant distinction between the 2 submissions for this measure (P = .23).

Alternatively, the data can be pooled, as shown by the boxplots of SN + SP (Fig 5), for which median and quartile values are indicated. The maximum sums were the following: nRCBV = 1.4

Finally, to assess the clinical relevance of the study observations, we determined the false-positive rate from T1 images, the false-positive rate was defined as the proportion of low-grade tumors thought to be aggressive and/or high-grade, as indicated by the decision for tumor resection, relative to all tumors resected. The false-negative rate was determined because data from all patients, including those who did not undergo an operation, were not available. By means of the T1+C images, the false-negative rate was defined as the proportion of tumors that are low-grade and had contrast agent–enhancing lesions. For the perfusion parameters, a false-positive rate is defined as a low-grade tumor with a value above the threshold determined for distinguishing low- from high-grade tumors.

RESULTS
For these consensus thresholds, the minimum individual SN/SP was 83%/77% for nRCBV and 80%/70% for nCBF.

For the 18 nRCBV measurements, the reliability was determined to be 0.93, indicating that 93% of the variation can be attributed to differences in tumors, and 7%, to differences in analysis methods. The reliability was 95% for nRCBV determined with leakage correction and 93% for the group without leakage correction. For the nRCBV computed with one of the most common leakage-correction algorithms (Boxerman-Schmainda-Weisskoff\textsuperscript{21}), the reliability improved to 98%. The reliability of standardized rCBV was 96%. For the 12 nCBF measurements, the reliability was 61%.

Finally, the false-positive rate using TI+ C was found to be 31%. In comparison, the false-positive rate across all software platforms was 15%–23% for nRCBV, 15%–31% for nCBF, and 8%–15% for standardized rCBV.

**DISCUSSION**

By means of carefully curated DSC-MR imaging datasets, obtained with a single acquisition approach, all nRCBV and nCBF metrics, processed by 7 different sites, could distinguish LGG from HGG. The optimal nRCBV and nCBF thresholds varied by only 9% and 15%, respectively. Unique to this study, consensus thresholds of nRCBV \(= 1.4\) and nCBF \(= 1.58\) were determined, indicating good accuracy overall and for each individual site. These results should bolster confidence in the ability of DSC-MR imaging to provide reliable and consistent cross-platform perfusion metrics for the evaluation of brain tumors and, specifically, for distinguishing low- from high-grade gliomas.

The range of nRCBV threshold values determined in this study is much tighter than the 0.7–3.0 range previously reported for distinguishing tumor grade,\textsuperscript{25,26} predicting differences in survival,\textsuperscript{10,17,27-29} and distinguishing true progression from pseudoprogression\textsuperscript{30} and tumor from treatment effect.\textsuperscript{11,12,21}

While this large range of threshold values has been attributed to different acquisition and postprocessing schemes,\textsuperscript{20} differences in patient populations and the clinical questions addressed also contribute to the variabilities. While it is unlikely that a single threshold can be universally applied for all clinical questions, these studies suggest that with well-defined studies to address a specific outcome under carefully controlled conditions, it is possible to reach consensus.

This range of thresholds should be well thought out. For example, in one study,\textsuperscript{32} 2 commercial software packages (nordicICE; NordicNeuroLab, Bergen, Norway and BrainSTAT; GE Healthcare) were compared. Like the present study, 1 dataset of 24 patients with de novo glioblastoma was used and ROIs of tumor and reference brain were predetermined. However, unlike the present study, 5 of 10 algorithms relied on the determination of the AIF.\textsuperscript{32} Using AIF to compute nRCBV resulted in coefficients of variation of 15%, but only 2% when AIF was not used. The challenges of reliably determining the AIF are well-known and may largely explain the poor repeatability.\textsuperscript{33,34} Most software platforms in this study did not incorporate AIF for nRCBV calculation and may therefore also explain the excellent concordance across sites. Yet, the computa-
tion for leakage correction. The largest differences between the in-house and commercial software occurred with the tool that required the most user interaction (nordicICE), further motivating the development of more automated workflows with less need for user interaction. Yet another study comparing these same 3 packages also found significant differences, with the outlying package depending heavily on the type of rCBV metric used.19 This finding again suggests that it is imperative that the same output metric be used when making such comparisons.

Of relevance to the current study, rCBV maps generated with IB Neuro showed superior leakage correction and stronger correlation with image-guided microvessel quantification as well as higher accuracy in distinguishing tumor recurrence from pseudoprogression/radiation necrosis compared with other software platforms.19 These results are relevant, given the number of sites in the present study that chose to use IB Neuro for their processing.

A limitation of the current study is the use of a DSC-MR imaging dataset that was obtained at a single center using a single approach. Use of a range of acquisition methods would likely result in greater variation in the DSC-MR imaging perfusion results. A previous study confirmed this by comparing a range of acquisition and analysis methods, which also influenced the ability to distinguish high-grade tumor from reference brain.20 However, a consensus regarding best practices for DSC-MR imaging data acquisition is being reached, as described in a recent review,40 and includes the approach used for this study. Specifically, use of a preload of contrast agent and a flip angle <90° is proving to be one of the most accurate approaches, further confirmed by 2 recent studies,19,41 both incorporating sophisticated simulations of DSC-MR imaging data representative of brain tumor. Use of a preload might also be an important reason for greater consistency across postprocessing methods in this study compared with previous studies (eg, Orsingher et al12). Collecting DSC-MR imaging data after preload was shown to decrease the dependence of tumor rCBV on the chosen method of analysis.20

An additional limitation of this study is the use of laboratory or proprietary commercial packages for which many of the details of the algorithmic implementation are not available and thus cannot be further evaluated as potential sources of differences. Also, the software platforms used for this study were dictated entirely by platforms being used at each participating site. Consequently, this is not a comprehensive comparison of

<table>
<thead>
<tr>
<th>Team/Entry No.</th>
<th>Software Platform</th>
<th>Leakage Correction Used</th>
<th>LG vs HG (P Value)</th>
<th>Threshold</th>
<th>SN</th>
<th>SP</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>IB Neuro</td>
<td>Yes</td>
<td>5.34E-06</td>
<td>1.605</td>
<td>0.861</td>
<td>0.846</td>
<td>0.889</td>
</tr>
<tr>
<td>B1</td>
<td>IB Neuro</td>
<td>Yes</td>
<td>5.36E-06</td>
<td>1.605</td>
<td>0.861</td>
<td>0.846</td>
<td>0.889</td>
</tr>
<tr>
<td>C1</td>
<td>IB Neuro</td>
<td>Yes</td>
<td>5.39E-06</td>
<td>1.605</td>
<td>0.861</td>
<td>0.846</td>
<td>0.889</td>
</tr>
<tr>
<td>D1</td>
<td>nordicICE</td>
<td>Yes</td>
<td>3.30E-06</td>
<td>1.305</td>
<td>0.972</td>
<td>0.769</td>
<td>0.892</td>
</tr>
<tr>
<td>E1</td>
<td>PGUI (LC)</td>
<td>Yes</td>
<td>7.32E-06</td>
<td>1.615</td>
<td>0.861</td>
<td>0.846</td>
<td>0.892</td>
</tr>
<tr>
<td>F1</td>
<td>FIAT</td>
<td>Yes</td>
<td>2.50E-05</td>
<td>1.235</td>
<td>0.944</td>
<td>0.769</td>
<td>0.893</td>
</tr>
<tr>
<td>G1</td>
<td>McCBV</td>
<td>Yes</td>
<td>1.26E-06</td>
<td>1.425</td>
<td>0.944</td>
<td>0.769</td>
<td>0.891</td>
</tr>
<tr>
<td>A2</td>
<td>IB Neuro No</td>
<td></td>
<td>1.55E-06</td>
<td>1.330</td>
<td>0.917</td>
<td>0.769</td>
<td>0.895</td>
</tr>
<tr>
<td>B2</td>
<td>IB Neuro No</td>
<td></td>
<td>1.56E-06</td>
<td>1.330</td>
<td>0.917</td>
<td>0.769</td>
<td>0.895</td>
</tr>
<tr>
<td>E4</td>
<td>PGUI</td>
<td>No</td>
<td>6.94E-06</td>
<td>1.400</td>
<td>0.861</td>
<td>0.846</td>
<td>0.887</td>
</tr>
<tr>
<td>E6</td>
<td>PGUI No</td>
<td></td>
<td>1.19E-05</td>
<td>1.395</td>
<td>0.861</td>
<td>0.769</td>
<td>0.873</td>
</tr>
</tbody>
</table>

Note: —AUC indicates area under the curve; LG, low-grade; HG, high-grade.

<table>
<thead>
<tr>
<th>Team/Entry No.</th>
<th>Software</th>
<th>LG vs HG (P Value)</th>
<th>Threshold</th>
<th>SN</th>
<th>SP</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>A5</td>
<td>IB Neuro</td>
<td>2.90E-05</td>
<td>1.815</td>
<td>0.806</td>
<td>0.846</td>
<td>0.861</td>
</tr>
<tr>
<td>B5</td>
<td>IB Neuro</td>
<td>2.38E-05</td>
<td>1.555</td>
<td>0.917</td>
<td>0.769</td>
<td>0.864</td>
</tr>
<tr>
<td>C2</td>
<td>IB Neuro</td>
<td>3.82E-06</td>
<td>1.415</td>
<td>0.944</td>
<td>0.769</td>
<td>0.880</td>
</tr>
<tr>
<td>D6</td>
<td>MatLab</td>
<td>3.61E-04</td>
<td>1.580</td>
<td>0.861</td>
<td>0.846</td>
<td>0.866</td>
</tr>
<tr>
<td>E7</td>
<td>PGUI</td>
<td>1.30E-05</td>
<td>1.575</td>
<td>0.806</td>
<td>0.846</td>
<td>0.846</td>
</tr>
<tr>
<td>E8</td>
<td>PGUI</td>
<td>3.85E-05</td>
<td>1.375</td>
<td>0.917</td>
<td>0.692</td>
<td>0.845</td>
</tr>
<tr>
<td>E9</td>
<td>PGUI</td>
<td>1.67E-04</td>
<td>1.460</td>
<td>0.833</td>
<td>0.769</td>
<td>0.812</td>
</tr>
<tr>
<td>E10</td>
<td>PGUI</td>
<td>7.86E-03</td>
<td>1.920</td>
<td>0.639</td>
<td>0.692</td>
<td>0.658</td>
</tr>
<tr>
<td>D6</td>
<td>Slicer</td>
<td>1.04E-06</td>
<td>1.405</td>
<td>0.944</td>
<td>0.769</td>
<td>0.887</td>
</tr>
<tr>
<td>F2</td>
<td>FIAT</td>
<td>2.65E-04</td>
<td>1.460</td>
<td>0.944</td>
<td>0.769</td>
<td>0.816</td>
</tr>
</tbody>
</table>

Note: —AUC indicates area under the curve; LG, low-grade; HG, high-grade.
FIG 5. Boxplots showing the sum of sensitivity and specificity values for nRCBV (A) and nCBF (B) for all sites and metrics. From these data, values that gave a maximum sum of specificity plus sensitivity and minimal variance were chosen as the consensus threshold values.

CONCLUSIONS

This study demonstrates that nRCBV and nCBF can be used to distinguish LGG from HGG in a consistent fashion and using a single consensus threshold. This result should increase confidence in using nRCBV primarily, but also nCBF, on a routine basis, potentially motivating its incorporation into the updated Response Assessment in Neuro-Oncology criteria. Finally, these results provide strong motivation for the development of more automated preprocessing workflows that are less dependent on subjective user interaction.

REFERENCES

11. Hu LS, Eichbacher JM, Heiserman JE, et al. Reevaluating the imag-
ing definition of tumor progression: perfusion MRI quantifies recurrent glioblastoma tumor fraction, pseudoprogression, and radioablation to predict survival. *Neuro Oncol* 2012;14:919–30


44. Bjornerud A, Sorensen AG, Mouridsen K, et al. T1- and T2*-dominant

