Real-Time Quantitative Assessment of Accuracy and Precision of Blood Volume Derived from DCE-MRI in Individual Patients During a Clinical Trial

Madhava P. Aryal[1], Choonik Lee[1], Peter G. Hawkins[1], Christina Chapman[1], Avraham Eisbruch[1], Michelle Mierzwa[1], and Yue Cao[1,2,3]

Departments of [1]Radiation Oncology; [2]Radiology; and [3]Biomedical Engineering, University of Michigan, Ann Arbor, MI

Corresponding Author:
Madhava P. Aryal, PhD
Department of Radiation Oncology, University of Michigan,
519 W William St, Ann Arbor, MI 48103;
E-mail: mparyal@med.umich.edu

Key Words: quantitative imaging, repeatability, real-time assessment, dynamic contrast enhanced MRI, blood volume, head and neck cancer

Abbreviations: Quantitative imaging (QI), blood volume (BV), dynamic contrast-enhanced (DCE), magnetic resonance imaging (MRI), magnetic resonance (MR), arterial input function (AIF), quality assurance (QA), head and neck (HN), radiation therapy (RT), echo time (TE), repetition time (TR), 3-dimensional (3D), field of view (FOV), volumes of interest (VOIs), repeatability coefficient (RC), within-subject mean squares (WMS), confidence interval (CI), sternocleidomastoid muscle (SCM)

ABSTRACT

Accuracy and precision of quantitative imaging (QI) metrics should be assessed in real time in each patient during a clinical trial to support QI-based decision-making. We developed a framework for real-time quantitative assessment of QI metrics and evaluated accuracy and precision of dynamic contrast-enhanced (DCE)-magnetic resonance imaging (MRI)–derived blood volume (BV) in a clinical trial for head and neck cancers. Patients underwent DCE-MRI before and after 2 weeks of radiation therapy (2wkRT). A mean as a reference value and a repeatability coefficient (RC) of BV values established from *n* patients in cerebellum volumes of interest (VOIs), which were normal and affected little by therapy, served as accuracy and precision measurements. The BV maps of a new patient were called accurate and precise if the values in cerebellum VOIs and the difference between the 2 scans agreed with the respective mean and RC with 95% confidence. The new data could be used to update reference values. Otherwise, the data were flagged for further evaluation before use in the trial. BV maps from 62 patients enrolled on the trial were evaluated. Mean BV values were 2.21 (±0.14) mL/100 g pre-RT and 2.22 (±0.17) mL/100 g at 2wkRT; relative RC was 15.9%. The BV maps from 3 patients were identified to be inaccurate and imprecise before use in the clinical trial. Our framework of real-time quantitative assessment of QI metrics during a clinical trial can be translated to different QI metrics and organ-sites for supporting QI-based decision-making that warrants success of a clinical trial.

INTRODUCTION

Quantitative imaging (QI) metrics are emerging as a tool for therapeutic response assessment in cancer treatment (1). As QI tools have been technically validated, clinical trials start to make decisions based upon these imaging metrics, for example, quantitative parameters derived from dynamic contrast-enhanced (DCE)-magnetic resonance imaging (MRI) (1, 2).

The DCE-MRI-derived QI metrics can be affected by differences in MRI platforms, pulse sequences, acquisition parameters, image reconstruction schemes, pharmacokinetic models, and quantification software packages (3-8), which limits deployment of DCE-MRI in clinical trials and practice. MRI scanners from each vendor have unique hardware configuration, vendor-specific pulse sequences, and reconstruction schemes, which can cause a systematic bias in estimated QI metrics (4). In addition, selection of magnetic resonance (MR) acquisition parameters can influence quantification of these metrics (5, 9). Furthermore, QI metrics derived from different image-processing software packages can lead to substantial variations in the metrics, even when using the same pharmacokinetic model, T1 map, arterial input function (AIF), and region of interest (6, 7). To address these challenges, collaborative efforts under the initiatives of professional societies and government agencies have been made for development of DCE-MRI profiles, T1 phantoms, digital reference object, and statistical methods to harmonize imaging acquisition across different platforms, to validate imaging hardware and software, to test computer algorithms, and to assess technical performance (4-6, 10-16). All these efforts are absolutely necessary but not sufficient to warrant the accuracy and precision of QI metrics obtained in each individual

ISSN 2379-1381 http://dx.doi.org/10.18383/j.tom.2018.00029

patient during a clinical trial, which could affect decision-making and even clinical outcomes (1). Therefore, it is necessary to develop and implement a quantitative quality assurance (QA) procedure to measure QI metrics acquired in the patients who are on the trial (17).

Accuracy, in general, refers to closeness of a measured QI metrics to a true or known value, while precision is an agreement between repeated measurements of a metrics (17). For any QI metrics that does not have its true value available, its deviation from a reference value, obtained as a group mean from a large sample study in any standard reference region, can serve as its measurement accuracy (17). Precision, more commonly known as repeatability, can be easily evaluated from repeated measurements, often called as test–retest studies, in a normal reference region that is not expected to have any changes during a time interval of test–retest studies (17). Under these principles, a reference value and repeatability coefficient (RC) of a QI metrics in a reference region under certain conditions or constraints of image acquisition and process can be determined from a sample of population with 95% confidence and used to assess accuracy and precision of the metrics measured from an individual patient.

DCE-MRI-derived blood volume (BV) is emerging as a promising QI metrics in assessing therapeutic response in head and neck (HN) cancers (18, 19). Tumor subvolumes characterized by low BV have been reported to be high-risk imaging biomarkers for tumor progression (19-22). Boosting those poorly perfused subvolumes with high radiation doses could improve local and regional control (23, 24). To test this clinical hypothesis, a randomized phase-II adaptive radiation therapy (RT) trial that targets persisting poorly perfused subvolumes of the tumor with high radiation doses in patients with poor prognosis HN cancers has been initiated (21, 22, 25). The persisting poorly perfused tumor subvolumes are defined on the basis of BV measurements pre-RT and 2 weeks after starting RT. Inaccurate and unrepeatable estimates of BV maps could generate false, poorly perfused subvolumes. Subsequently, intensifying radiation doses to these falsely classified subvolumes can lead to either tumor overdose or underdose, which could increase radiation toxicity or cause failure of disease control, respectively. To achieve the goal of the clinical trial, it is critical to ensure accuracy and precision of BV maps in each individual patient and thereby warrant proper segmentation of low BV tumor subvolumes.

The present study developed and evaluated a framework for real-time quantitative assessment of accuracy and precision of a QI metrics in individual patients during a clinical trial. The method was applied to DCE-MRI-derived BV maps acquired during an ongoing clinical trial for poor prognosis HN cancers. As the repeatability analysis cannot be done in treated tumor volume owing to expected therapy-caused changes, a normal tissue region in the cerebellum that has little therapy-induced change was used as a reference region for BV measurements and hence to assess the accuracy and precision of BV maps. Our study showed that inaccurate and imprecise BV maps could be detected in real time before clinical decision was made. This method can be extended to other QI metrics and body sites. This process should be a part of the workflow of a clinical trial.

MATERIALS AND METHODS

Human Subjects

Patients with advanced HN cancers were enrolled in an IRB-approved randomized phase-II clinical trial. The patients who have advanced human papillomavirus (HPV)-HN cancers (stage IV) or HPV+ T4/N3 HN cancers (stage III) were eligible for the trial. All patients gave their study-specific informed consent to participate in the trial. Patients underwent MRI scans before RT and after receiving 10 fractions (Fx) of 2 Gy per fraction of radiation.

MR Acquisition

All MRI scans were acquired on a 3 T MR scanner (Magnetom Skyra, Siemens Healthineers, Erlangen, Germany). Each patient underwent scanning in the radiation treatment position on a flat table top using the patient-specific immobilization face mask, head support, and bite bar. MRI series included 2-dimensional multislice pre- and postcontrast T1-weighted images with fat saturation (voxel size: 0.88 × 0.88 × 3.3 mm^3; echo time [TE]/repetition time [TR] = 8.4/1040 milliseconds), 2-dimensional T2-weighted images (voxel size: 0.78 × 0.78 × 3.3 mm^3; TE/TR = 89/11000 milliseconds), and 3-dimensional (3D) volumetric T1-weighted DCE images. The DCE image volumes were acquired using a 3D gradient-echo sequence in the sagittal orientation with a large field of view (FOV) in the superior and inferior directions to cover primary and nodal cancers, carotid artery, and cerebellum. The sagittal orientation allows us to achieve higher temporal resolution and avoid time-of-flight effects of blood-flow spins (Figure 1). Other acquisition parameters included flip angle/TE/TR = 10°/0.97/2.73 milliseconds, FOV = 300 × 300 × 150 mm^3, and voxel size ≈ 1.6 × 1.6 × 2.5 mm^3. Sixty dynamic scans were collected at 3 minute, with a temporal resolution of 3 second.

Extended Tofts Model for BV Quantification

Plasma volume maps (v_p) were generated from the T1-weighted DCE image series using the extended Tofts model (26);

$$C_{tiss}(t) = K^{trans}\int_0^t e^{-k_{ep}(t-\tau)}C_p(\tau)d\tau + V_pC_p(t), \qquad (1)$$

where $C_{tiss}(t)$ and $C_p(t)$ were respective tissue and plasma concentrations of the contrast agent, and K^{trans} and k_{ep} were respective transfer constant and rate. An assumption of $\Delta S/S_0 \propto C$ was used to fit equation (1). In-house software package of functional image analysis tool (FIAT) was used for image analysis and processing to generate parametric maps (20, 21), in which the implemented extended Tofts model has been validated using digital reference object (DRO) (5). To convert the plasma volume maps to the BV maps, a Hematocrit value of 0.45 was applied (27). A protocol-specific procedure of DCE analysis was established before initiation of the clinical trial, particularly regarding how to create an AIF. To obtain the AIF, a dynamic phase in which contrast just entered the carotid artery was chosen by visually inspecting the temporal profile of the dynamic image volumes. Then, an AIF was generated by thresholding 20 voxels with the largest intensity changes on the selected phase com-

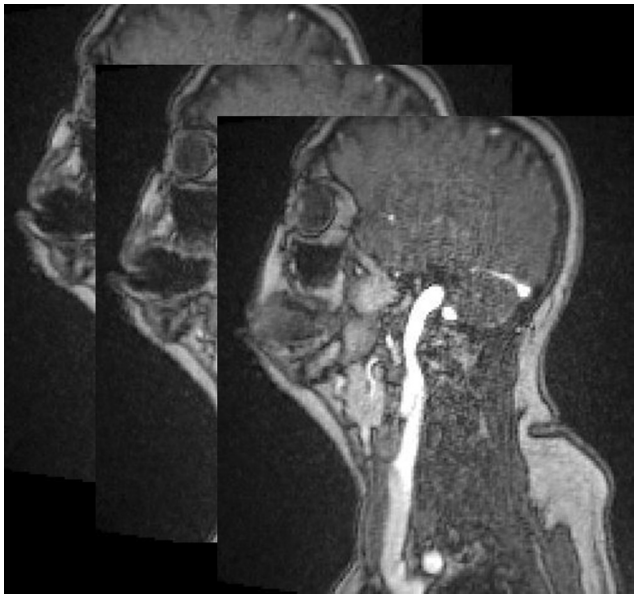

Figure 1. T1-weighted dynamic contrast-enhanced (DCE) images acquired using a 3-dimensional gradient-echo sequence in the sagittal orientation. As shown in the figure, these images were collected with a large field-of-view (FOV) in the superior and inferior directions to cover the primary and nodal tumors, carotid artery, and normal tissue region in cerebellum. The latter region, that is, the normal cerebellum region, was used as a reference region for quality assessment of blood volume (BV) measurement in each individual examination.

pared with the average baseline image intensities. Finally, the AIF was visually inspected to make sure that its voxels were located within the carotid artery and had the expected dynamic profile. BV maps were derived from the extended Tofts model using the patient-specific AIF, and then coregistered to the postcontrast T1-weighted images at pre-RT using rigid-body transformation (20).

System-Level QA

To ensure quality of quantitative parametric maps, QA of hardware and software at system-level was performed routinely. System-level QA of the MRI scanner was performed daily, weekly, and yearly using an ACR water phantom following the ACR protocol. Daily signal-to-noise ratio variations were recorded and were stable. Also, in an NCI Quantitative Imaging Network (QIN) multicenter collaborative project, we evaluated accuracy, repeatability, and interplatform reproducibility of T1 quantification from variable flip angles using an NIST T1 water phantom on our scanner, compared to others (4). For software QA, performance of our implementation of the extended Tofts model was evaluated using a digital reference object, that is, synthesized DCE phantoms with and without noise, which was fully reported previously (5). Also, we participated in an NCI QIN multicenter AIF challenge to validate and compare our AIF delineation procedure with others' (15). Based upon these evaluation and validation, *imFIAT* has been granted a level-2 benchmark by NCI QIN (28).

Individual-Level Assessment of Accuracy and Precision of BV Maps

Our pilot study indicates that repeatability of BV values in the cerebellum is stable and ~18% (unpublished data). Also, cerebellums in our patients received a mean radiation dose <3Gy after 10 Fx of 2 Gy treatment. Therefore, we chose cerebellum as a reference region and manually drew bilateral volumes of interest (VOIs) across 2–3 slices having a volume of ~4 cc

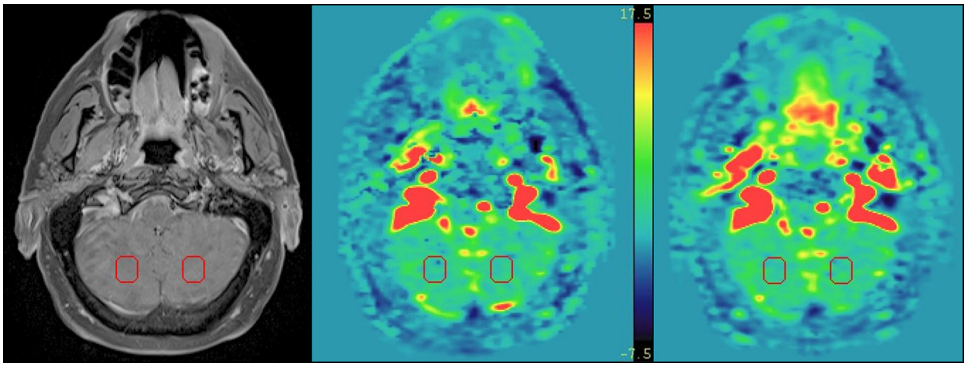

Figure 2. The coregistered postcontrast T1-weighted image (left), and the BV maps pre-radiation therapy (RT) (middle) and after 10 fractions of radiation therapy (right) from a sample study. The postcontrast T1-weighted images was used to delineate the tumor volumes and to locate the normal cerebellum region as a reference region. Red contours (~4cc in volume) represent the volumes of interest (VOIs) in the normal cerebellum, which was used as the reference region for the accuracy and precision analysis.

(number of voxels, ~1600) to extract mean BV values (Figure 2).

For each patient, MRI scanning was performed pre-RT and repeated after 10 Fx of radiation (2wkRT), which were considered as test and retest studies. An RC of BV values in the cerebellum VOIs was estimated using 1-way analysis of variance (ANOVA) model (29). First, within-subject mean squares (WMS) was estimated from n patients. Then RC and relative RC were estimated by $RC = 2.77 \times \sqrt{WMS}$ and $rRC = 100 \times RC/\hat{X}$, respectively, where $\hat{X}$ was the grand mean of overall observations from n patients. Because the WMS for 2 repeated measurements was distributed as $\chi^2_n wSD^2/n$, the 95% confidence interval (CI) of the estimated RC was given by $RC_L = RC \times \sqrt{n/\chi^2_n(0.975)}$ and $RC_U = RC \times \sqrt{n/\chi^2_n(0.025)}$, where $\chi^2_n(a)$ was the a^{th} percentile of the χ^2 distribution with n degrees of freedom.

To assess accuracy and precision of BV values in each individual patient, a group mean (M_n) of BV in cerebellum VOIs as a reference value with a 95% CI defined by standard deviation (SD_n), and an RC_n with a 95% CI defined by RC_L and RC_U were computed from n patients. For the next new scan, it was determined whether the mean BV value in the cerebellum VOI was between $M_n - 2SD_n$ and $M_n + 2SD_n$. If yes, the BV map was deemed accurate with 95% confidence. For each new patient, a difference of BV between the 2 scans (test and retest) was determined whether it was within $-RC_n$ and RC_n. If yes, the BV maps of this new patient were considered repeatable with 95% confidence. When the new patient's data passed both tests, the BV maps could be used to update the reference value and RC. Otherwise, the BV maps from this individual patient were flagged for further evaluation or correction before used in the clinical trial.

Other Statistical Analysis

A paired t test was performed to examine whether there was any difference between mean BVs measured at test and a retest with P-value <0.05 as statistically significant. The distribution of differences in mean BV values between the 2 scans was tested for normality using the Shapiro–Wilk test. Similarly, to detect a potential relationship between the measurement error and the magnitude of the combined mean BV values between 2 scans, a rank correlation coefficient (Kendall's tau) test between absolute differences against their combined means was performed.

Association Between Repeatability of AIF Peak and BV

As noted, we used a fixed imaging protocol to minimize variations in acquisition. However, it was unknown how repeatability of AIF was associated with repeatability of BV values. To examine this association, we measured the AIF peak value for each scan and calculated the RC from the 2 scans. We compared percentage differences of AIF peaks between the 2 scans with those of BV values measured in the cerebellum VOIs.

RESULTS

At the time of this report, 62 consecutive patients (median age, 62 years; male, 52; female, 10) were enrolled in the clinical trial. For the first 10 patients, the mean (±SD) BV values from test and retest were 2.22 (±0.13) mL/100 g and 2.21 (±0.19) mL/100 g, respectively, and not significantly different (P-value = 0.79: paired t test), yielding the overall group mean (±SD) of 2.21 (±0.16) mL/100 g (see Table 1). The difference in the BV values between test–retest studies was independent to the combined mean (P-value = 0.21: Kendall tau test), indicating that the measurement error was independent to the magnitude of measured BV values. Also, the Shapiro–Wilk test showed that the differences in BV values between the 2 examinations were normally distributed. An RC of BV values between the 2 tests was estimated to be 0.37, yielding a relative RC (rRC) of 16.7% with a 95% CI of (11.7%, 29.4%). Using the leave-1-out cross-validation, we did not find any outlier from the first 10 patients. Therefore, we used M_{10} and RC_{10} as starting reference values to evaluate the next patient (Table 1).

Table 1. Summary Statistics for BV Measurement at Normal Cerebellum Region

Statistical Parameters	Preliminary Statistics (n = 10)	Updated Statistics (n = 58)
Mean BV (±SD) (mL/100 g)		
Test study	2.22 (±0.13)	2.21 (±0.14)
Retest study	2.21 (±0.19)	2.22 (±0.17)
Overall	2.21 (±0.16)	2.22 (±0.15)
Paired t test (P-value)	0.79	0.73
Kendall's tau test (P-value)	0.21	0.67
WMS	0.02	0.02
RC (rRC%)	0.37 (16.7)	0.35 (15.9)
95%CI on rRC (%): rRC_L, rRC_U	11.7, 29.4	13.5, 19.5

BV measurements from 62 patients were evaluated in real time, and 3 patients were identified to have inaccurate BV values in 1 of the 2 scans (Figure 3). Mean BVs measured from these 3 patients were in the range of 3.05–3.95 mL/100 g, which were much higher than those measured from the group mean + 2 × SD value (2.52 mL/100 g). The repeatability tests found that the percentage differences of BV values between the 2 scans of the 3 patients were much greater than the uncertainty range defined by −RC and RC. Note that our procedure detected large variations of BV values in 3 scans in real time, but not in retrospective analysis. The consequences of the BV maps for decision-making with and without correction were evaluated and discussed with the physicians during the clinical trial.

As the patients were enrolled into the clinical trial, the data from the 3 patients were excluded from the updated reference values for accuracy and precision measurements. One additional patient who had BV values within the normal range for both test and retest was excluded owing to partial coverage of cerebellum in 1 scan and mismatched slices in cerebellum between the 2 scans. As a result, the data from 58 patients were included to update the reference values. A group mean (±SD) of BV values was of 2.21 (±0.14) mL/100 g at test, and 2.22 (±0.17) mL/100 g at retest, which were not significantly different (P-value = 0.

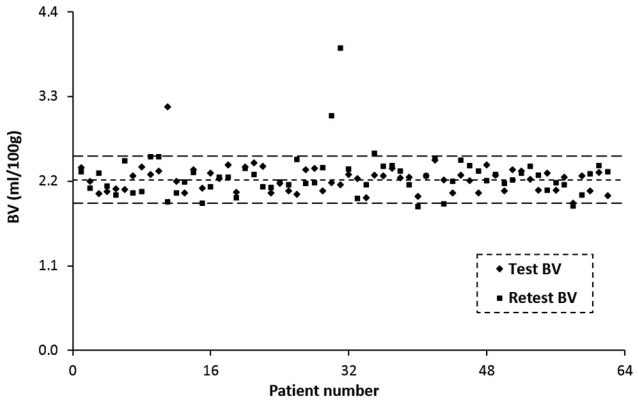

Figure 3. Mean BV values obtained in the cerebellum VOIs in each study plotted against the patient number. A center dotted line represents the overall group mean of BV, while 2 dashed lines depict the 95% confident interval (±1.96 × SD from the group mean). Note that 3 BV values are far away from the confident range, and are identified as inaccurate BV measurements.

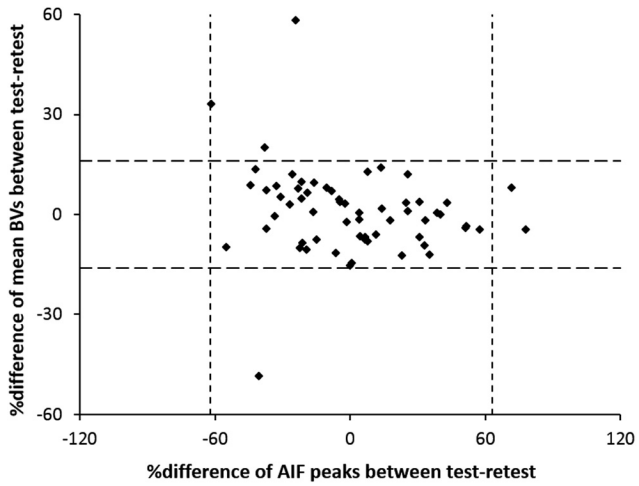

Figure 5. Scatter plot of percentage differences of mean BVs versus percentage differences of arterial input function (AIF) peaks between the 2 scans, with their corresponding RC ranges (horizontal dashes lines for BV and vertical ones for AIF peak).

73: paired *t* test; see Table 1), suggesting stability of the quantified BV maps. Also, the absolute difference was independent of their combined means (*P*-value = 0.67: Kendall tau test). ANOVA led to an RC of 0.35, and an rRC of 15.9% with a 95% CI of (13.5%, 19.5%). Note that the 95% CI (uncertainty) of estimated RC decreased with an increase in the number of patients. Figure 4 shows a plot of percentage differences of BV values between test–retest studies versus their combined means. As shown in the plot, percentage differences of mean BVs from the 3 patients, who had inaccurate mean BVs, were much large than the RC interval (% difference > 33% at the lowest), indicating the imprecision in the repeated measures.

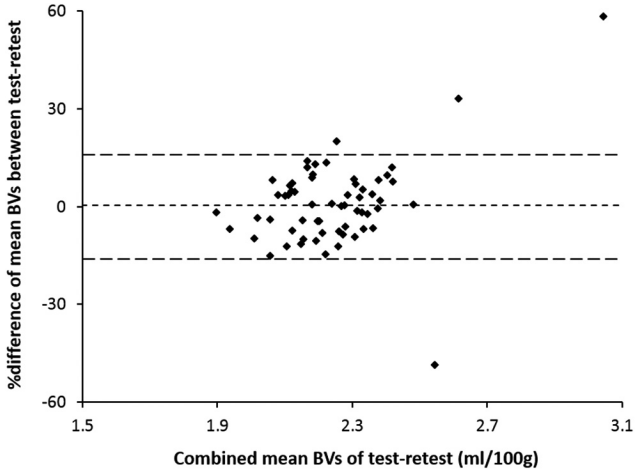

Figure 4. Bland–Altman plots of the percentage difference in mean BV between 2 studies plotted against their combined mean. A center dotted line shows the mean percentage difference of BV between the 2 scans, while 2 dashed lines represent the estimated repeatability coefficient (RC) interval (−RC, RC).

Finally, the relative RC of the AIF peak values was of 61.8%. Figure 5 shows a scatter plot of percentage differences of BV values in the cerebellum VOIs versus those of AIF peak values between the 2 scans. Note that there was no association or even a trend between the 2 differences, suggesting the variation of AIF peaks could not explain the variation in BV measurements.

DISCUSSION

In this study, we developed and evaluated a methodology and metrics for real-time quantitative assessment of accuracy and precision on DCE-MRI derived metrics using reference values in a normal reference tissue region. It is critical to establish such a real-time QA test in the workflow of a clinical trial to identify unreliable estimates of QI metrics before used in a trial. A subsequent action should be planned in the design of a clinical trial. A real-time QA procedure of QI metrics in individual patients would enhance the ability of the trial to achieve its objectives and increase reliability of scientific findings. Our method can be extended to other QI metrics and body-sites to support individualized therapy and improve therapeutic outcomes.

It would be worth noting that accuracy and precision of BV values investigated in this study do not represent how accurate the QI metrics measure a true physiological BV. As discussed in the Introduction, they are measures of bias and variation of BV values as a QI metrics quantified from HN DCE-MRI using the extended Tofts model to reference values. Our data show that the

group mean and RC of BV values in the cerebellum are stable, suggesting that it is a great candidate used as a reference region. As anticipated, the 95% CI of estimated RC decreases with an increase in the sample size. Using these reference values, we are able to detect unreliable QI measures of individual patients in real time during the clinical trial. Our test is different from test and retest analysis performed before therapy. The latter helps us understand the general technical behavior of a QI metrics in a sample of population, but it does not tell us whether the metrics acquired in each patient in a clinical trial is reliable or not. Finally, the impact of uncertainty of a QI metrics in a decision-making process needs to be investigated in future.

As shown in this study, reference values have to be established in a reference tissue region to perform the proposed QA test. The reference tissue region chosen may depend upon the image type and body site of interest. However, the QI metrics in a reference region has to be stable, less affected by therapy, and within the FOV of the scan. In our preliminary investigation, we tested sternocleidomastoid muscle (SCM) contralateral to tumor as a possible tissue reference region. We found that the BV values in SCM were not as stable as those in the cerebellum, possibly owing to low BV in SCM. Also, in some cases, tumors are distributed bilaterally, in which there is no noninvolved SCM that can be used as a reference region. On the other hand, the cerebellum tissue receives few Gy radiation doses (<3 Gy) for HN cancer treatment, and BV changes in cerebellum VOIs after 10 Fx of RT do not show any positive or negative trend (Figure 4), suggesting that the treatment effect within the cerebellum is minimum and can be ignored. Reference values of BV in the cerebellum VOIs are adequate for evaluation of the overall quality of BV maps, as MRI data are acquired in the k-space and BV maps are determined by a single AIF. However, local motion, e.g., swallowing, can cause local degradation in DCE-MRI, which cannot be captured by the analysis performed in the normal reference region. However, it still needs to be cautious to use QI metrics during a therapeutic trial.

In our study, patient positioning, scanner, image protocol, acquisition procedure, and analysis software and process are controlled carefully to maintain consistency of QI metrics delineation during the clinical trial. The factors that can influence repeatability of DCE-MRI-derived QI metrics include patient positioning, image registration, AIF delineation, image noise, image process, treatment effect, and unknown physiological fluctuation. We further investigated repeatability of AIF peaks, as well as its influence on repeatability of BV maps, but found no relationship among differences in the BV values and the AIF peaks between the 2 scans (Figure 5). These findings indicate that the AIF peak variation cannot solely explain one in the BV measures.

In conclusion, the present study developed and evaluated a methodology for quantitative assessment of accuracy and precision of DCE-MRI derived BV maps in a phase-II randomized clinical trial for poor prognosis HN cancers. The outlined framework was able to detect outliers, that is, identify the individual patients who had unreliable BV values in real time during the clinical trial. Because accuracy and precision of QI metrics influence decision-making in the individualized and adaptive cancer therapy, individual QA testing of such QI metrics needs to be integrated into a clinical trial workflow to warrant success of the trial.

ACKNOWLEDGMENTS

This work is supported in part by NIH/NCI grants of 1U01CA183848 and RO1CA184153.

Disclosures: No disclosures to report.

Conflict of Interest: The authors have no conflict of interest to declare.

REFERENCES

1. Abramson RG, Arlinghaus LR, Dula AN, Quarles CC, Stokes AM, Weis JA, Whisenant JG, Chekmenev EY, Zhukov I, Williams JM, Yankeelov TE. MR imaging biomarkers in oncology clinical trials. Magn Reson Imaging Clin N Am. 2016;24:11–29.
2. NCT02031250, NCT00581906, NCT02070705, and NCT02878109. Available on National Institute Health US. National Library of Medicine Clinical Trials. 2018.
3. Kim H. Variability in quantitative DCE-MRI: sources and solutions. J Nat Sci. 2018;4. pii: e484.
4. Bane O, Hectors SJ, Wagner M, Arlinghaus LL, Aryal MP, Cao Y, Chenevert TL, Fennessy F, Huang W, Hylton NM, Kalpathy-Cramer J, Keenan KE, Malyarenko D, Mulkern RV, Newitt DC, Russek SE, Stupic KF, Tudorica A, Wilmes LJ, Yankeelov TE, Yen YF, Boss MA, Taouli B. Accuracy, repeatability, and interplatform reproducibility of T1 quantification methods used for DCE-MRI: Results from a multicenter phantom study. Magn Reson Med. 2018;79:2564–2575.
5. Cao Y, Li D, Shen Z, Normolle D. Sensitivity of quantitative metrics derived from DCE MRI and a pharmacokinetic model to image quality and acquisition parameters. Acad Radiol. 2010;17:468–478.
6. Huang W, Li X, Chen Y, Li X, Chang MC, Oborski MJ, Malyarenko DI, Muzi M, Jajamovich GH, Fedorov A, Tudorica A, Gupta SN, Laymon CM, Marro KI, Dyvorne HA, Miller JV, Barbodiak DP, Chenevert TL, Yankeelov TE, Mountz JM, Kinahan PE, Kikinis R, Taouli B, Fennessy F, Kalpathy-Cramer J. Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: a multicenter data analysis challenge. Transl Oncol. 2014;7:153–166.
7. Heye T, Davenport MS, Horvath JJ, Feuerlein S, Breault SR, Bashir MR, Merkle EM, Boll DT. Reproducibility of dynamic contrast-enhanced MR imaging. Part I. Perfusion characteristics in the female pelvis by using multiple computer-aided diagnosis perfusion analysis solutions. Radiology. 2013;266:801–811.
8. Sourbron SP, Buckley DL. Tracer kinetic modelling in MRI: estimating perfusion and capillary permeability. Phys Med Biol. 2012;57:R1–R33.
9. Heisen M, Fan X, Buurman J, van Riel NA, Karczmar GS, ter Haar Romeny BM. The influence of temporal resolution in determining pharmacokinetic parameters from DCE-MRI data. Magn Reson Med. 2010;63:811–816.
10. Kurland BF, Gerstner ER, Mountz JM, Schwartz LH, Ryan CW, Graham MM, Buatti JM, Fennessy FM, Eikman EA, Kumar V, Forster KM, Wahl RL, Lieberman FS. Promise and pitfalls of quantitative imaging in oncology clinical trials. Magn Reson Imaging. 2012;30:1301–1312.
11. Obuchowski NA, Reeves AP, Huang EP, Wang XF, Buckler AJ, Kim HJ, Barnhart HX, Jackson EF, Giger ML, Pennello G, Toledano AY, Kalpathy-Cramer J, Apanasovich TV, Kinahan PE, Myers KJ, Goldgof DB, Barboriak DP, Gillies RJ, Schwartz LH, Sullivan DC, Algorithm Comparison Working Group. Quantitative imaging biomarkers: a review of statistical methods for computer algorithm comparisons. Stat Methods Med Res. 2015;24:68–106.
12. Raunig DL, McShane LM, Pennello G, Gatsonis C, Carson PL, Voyvodic JT, Wahl RL, Kurland BF, Schwarz AJ, Gönen M Zahlmann G, Kondratovich MV, O'Donnell K, Petrick N, Cole PE, Garra B, Sullivan DC; QIBA Technical Performance Working Group. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res. 2015;24:27–67.

13. Keenan KE, Ainslie M, Barker AJ, Boss MA, Cecil KM, Charles C, Chenevert TL, Clarke L, Evelhoch JL, Finn P, Gembris D, Gunter JL, Hill DLG, Jack CR, Jr., Jackson EF, Liu G, Russek SE, Sharma SD, Steckner M, Stupic KF, Trzasko JD, Yuan C, Zheng J. Quantitative magnetic resonance imaging phantoms: A review and the need for a system phantom. Magn Reson Med. 2018;79:48–61.
14. QIBA. Profile: DCE MRI quantification version 1.0; Accessed June 28, 2011. Available from: http://www.rsna.org/uploadedFiles/RSNA/Content/Science_and_Education/QIBA/DCE-MRI_Quantification_Profile_v1%200-Reviewed-Draft%208-8-12.pdf.
15. Huang W, Chen Y, Fedorov A, Li X, Jajamovich GH, Malyarenko DI, Aryal MP, LaViolette PS, Oborski MJ, O'Sullivan F, Abramson RG, Jafari-Khouzani K, Afzal A, Tudorica A, Moloney B, Gupta SN, Besa C, Kalpathy-Cramer J, Mountz JM, Laymon CM, Muzi M, Schmainda K, Cao Y, Chenevert TL, Taouli B, Yankeelov TE, Fennessy F, Li X. The impact of arterial input function determination variations on prostate dynamic contrast-enhanced magnetic resonance imaging pharmacokinetic modeling: a multicenter data analysis challenge. Tomography. 2016;2:56.
16. QIBA. Synthetic DCE-MRI Data. Available from: https://qibawikirsnaorg/index-php/Synthetic_DCE-MRI_Data. 2009.
17. Cercignani M, Dowell NG, Tofts PS. Quantitative MRI of the Brain; Quality Assurance: Accuracy, Precision, Controls and Phantoms 1. 2nd ed. CRC Press: Boca Raton, Florida; 2018.
18. Cao Y, Popovtzer A, Li D, Chepeha DB, Moyer JS, Prince ME, Worden F, Teknos T, Bradford C, Mukherji SK, Eisbruch A. Early prediction of outcome in advanced head-and-neck cancer based on tumor blood volume alterations during therapy: a prospective study. Int J Radiat Oncol Biol Phys. 2008;72:1287–1290.
19. Agrawal S, Awasthi R, Singh A, Haris M, Gupta R, Rathore R. An exploratory study into the role of dynamic contrast-enhanced (DCE) MRI metrics as predictors of response in head and neck cancers. Clin Radiol. 2012;67:e1–e5.
20. Cao Y. WE-D-T-6C-03: development of image software tools for radiation therapy assessment. Med Phys. 2005;32:2136–2136.
21. Wang P, Popovtzer A, Eisbruch A, Cao Y. An approach to identify, from DCE MRI, significant subvolumes of tumors related to outcomes in advanced head-and-neck cancer. Med Phys. 2012;39:5277–5285.
22. Hawkins PG, Lee JY, Lee C, Green M, Mierzwa ML, Aryal MP, Arnould GS, Worden F, Swiecicki PL, Spector ME, Schipper M, Cao Y, Eisbruch A. Adaptive chemoradiation therapy for head and neck cancer based on multiparametric MRI: interim results of a prospective randomized trial. Int J Radiat Oncol Biol Phys. 2017;99:E339.
23. Wang J, Zheng J, Tang T, Zhu F, Yao Y, Xu J, Wang AZ, Zhang L. A randomized pilot trial comparing position emission tomography (PET)-guided dose escalation radiotherapy to conventional radiotherapy in chemoradiotherapy treatment of locally advanced nasopharyngeal carcinoma. PloS One. 2015;10:e0124018.
24. Overgaard J. Hypoxic modification of radiotherapy in squamous cell carcinoma of the head and neck–a systematic review and meta-analysis. Radiother Oncol. 2011;100:22–32.
25. Teng F, Aryal M, Lee J, Lee C, Shen X, Hawkins P, Mierzwa M, Eisbruch A, Cao Y. Adaptive boost target definition in high-risk head and neck cancer based on multi-imaging risk biomarkers. Int J Radiat Oncol Biol Phys. 2018;102:969–977.
26. Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, Larsson HB, Lee TY, Mayr NA, Parker GJ, Port RE, Taylor J, Weisskoff RM. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. J Magn Reson Imaging. 1999;10:223–232.
27. Sourbron SP, Buckley DL. Classic models for dynamic contrast-enhanced MRI. NMR Biomed. 2013;26:1004–1027.
28. Farahani K, Kalpathy-Cramer J, Chenevert TL, Rubin DL, Sunderland JJ, Nordstrom RJ, Buatti J, Hylton N. Computational challenges and collaborative projects in the NCI Quantitative Imaging Network. Tomography. 2016;2:242–249.
29. Barnhart HX, Barboriak DP. Applications of the repeatability of quantitative imaging biomarkers: a review of statistical analysis of repeat data sets. Transl Oncol. 2009;2:231–235.