Translational Oncology
www.transonc.com

Volume 2 Number 4 December 2009 pp. 211–215 211

Magnetic Resonance Assessment of Response to Therapy: Tumor Change Measurement, Truth Data and Error Sources[1]

Edward F. Jackson*, Daniel P. Barboriak†, Luc M. Bidaut* and Charles R. Meyer‡

*Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; †Department of Radiology, Duke University Medical Center, Durham, NC, USA; ‡Department of Radiology, University of Michigan, Ann Arbor, MI, USA

Abstract

This article describes methods and issues that are specific to the assessment of change in tumor characteristics as measured using quantitative magnetic resonance (MR) techniques and how this relates to the establishment of quantitative MR imaging (MRI) biomarkers of patient response to therapy. The initial focus is on the various sources of bias and variance in the measurement of microvascular parameters and diffusion parameters as such parameters are being used relatively commonly as secondary or exploratory end points in current phase 1/2 clinical trails of conventional and targeted therapies. Several ongoing initiatives that seek to identify the magnitude of some of the sources of measurement variations are then discussed. Finally, resources being made available through the National Cancer Institute *Reference Image Database to Evaluate Response* (RIDER) project that might be of use in investigations of quantitative MRI biomarker change analysis are described. These resources include 1) data from phantom-based assessment of system response, including short-term (1 hour) and moderate-term (1 week) contrast response and relaxation time measurement, 2) data obtained from repeated dynamic contrast agent–enhanced MRI studies in intracranial tumors, and 3) data obtained from repeated diffusion MRI studies in both breast and brain. A concluding section briefly discusses issues that must be addressed to allow the transition of MR-based imaging biomarker measures from their current role as secondary/exploratory end points in clinical trials to primary/surrogate markers of response and, ultimately, in clinical application.

Translational Oncology (2009) 2, 211–215

Introduction

The wide range of image contrast mechanisms provided by magnetic resonance imaging (MRI) techniques and the continuous improvements in imaging acquisition rates have led to substantial interest in the application of ever-expanding quantitative MRI biomarkers' applications, particularly in phase 1/2 clinical trials of novel therapeutics and combination therapies. Whereas MR-based measures of lesion size, for example, Response Evaluation Criteria in Solid Tumors (RECIST) measures (http://imaging.cancer.gov/clinicaltrials/imaging/), are perhaps not as commonly used compared with computed tomography (CT), MR techniques that assess functional characteristics of lesions and normal tissue are available. Such functional MR techniques include, but are not limited to, those that provide an assessment of microvascular parameters, such as volume, flow, and permeability, and those that provide an assessment of cellular volume/density parameters as reflected in changes in water diffusion rates. Whereas such measures are currently being used in phase 1/2 clinical trials, very few investigations into the sources of measurement variance and bias, and the mitigation of such effects, have been reported. Such investigations, however, are critical to the transition from successful applications of such techniques in a limited number of isolated facilities to the routine use of such applications in multicenter clinical trials and, ultimately, to the routine use of such techniques in patient care.

Address all correspondence to: Edward F. Jackson, PhD, Professor and Deputy Chair, Department of Imaging Physics, Chief, Section of MR and Ultrasound Physics, University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Unit 056, Houston, TX 77030-4009. E-mail: ejackson@mdanderson.org

[1]This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contracts N01-CO-12400 and 27XS112. The content of this publication does not necessarily reflect the views of policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.
Received 21 August 2009; Revised 21 August 2009; Accepted 21 August 2009

DOI 10.1593/tlo.09241

It should not be surprising that the wide range of image contrast mechanisms and functional imaging applications available using MR techniques gives rise to an equally wide range of sources of variance and bias, with several such sources discussed by Tofts [1]. Accordingly, both the assessment of such sources and the means for mitigating their effects must be matched to the specific application. In this report, the main sources of variance and bias have been broken into three major groups, with the understanding that there may be significant interactions between sources of variance (e.g., covariances) between effects. These include 1) patient-related factors, 2) acquisition-dependent factors, and 3) image analysis factors.

Primary Sources of Bias and Variance in MR Imaging Biomarker Applications

Patient-related factors. These factors are addressed primarily in the companion manuscript focused on non–modality-dependent factors, so only those issues specific to MR studies are discussed. The primary MR application factor, of course, relates to voluntary and involuntary patient motion and/or lesion motion. In MR, such motions can lead to a variety of artifacts and will often strongly affect measurement variance. The untoward effects of motion can, in many cases, be minimized by appropriate choice of the frequency- and phase-encoding directions, by using respiratory compensation and/or respiratory gating techniques (and/or cardiac gating techniques), or by using navigator echo techniques. Lesion location can strongly affect measurement variance due to lesion motion and/or motion of nearby normal tissues, for example, dynamic contrast agent-enhanced (DCE) studies of lesions near the heart and/or great vessels. In addition, effects due to therapies other than the target therapy in a given clinical trial may affect measurement variance, for example, steroids, particularly if the dose is varied during treatment. Finally, some MRI biomarkers of interest, including blood flow and blood volume, are affected by injection technique (site and rate of injection, use of saline flush), prandial status, caffeine consumption, and so on. Minimization of such patient-related factor variance is a very important, albeit often underappreciated, component of good study design.

Scanner hardware and acquisition-dependent factors. Fundamentally, the lower limits on measurement variance and bias are established by MR scanner calibration and stability and other hardware-related factors. Important factors in this regard include the magnetic field homogeneity, the radiofrequency subsystem calibration and stability, the gradient subsystem calibration and stability, gradient nonlinearity corrections (in-plane and, for some scanners and pulse sequences, through-plane), and radiofrequency (B_1) homogeneity and receiver coil sensitivity characteristics. Variance in imaging biomarker measures due to imaging hardware design and performance characteristics can be particularly problematic in MR applications where there are substantial differences in, for example, gradient subsystem performance and associated image acquisition rates. Equally important, if not more so, are numerous sources of variance across systems due to specific imaging protocol implementations. The contrast response for a particular pulse sequence class, for example, fast spoiled gradient echo, varies across vendors and field strengths and can also vary across software releases for a given vendor's platform. It is widely recognized that for studies for which repeat measurements will be obtained, for example, measurement of MRI biomarkers before and during or after therapy, requiring that the imaging be performed on the same MR system can be useful for decreasing measurement variance. This requirement, however, can make scheduling the studies more challenging and can also be subverted in practice by software and/or hardware upgrades, which can have a substantial effect on the measured parameter(s). This "upgrade dilemma" is a significant challenge in MR studies because software upgrades are frequent and are desirable to maintain "state-of-the-art" imaging capabilities. More advanced acquisition strategies, such as parallel imaging options, and image intensity correction techniques can have a substantial impact on MRI biomarker measurement variance, particularly in a multicenter study. Undoubtedly, a rigorous quality assurance program is critical to the minimization of measurement variance and bias. Unfortunately, such programs are relatively uncommon in MR facilities. With some notable exceptions, MR quality assurance programs are either nonexistent (beyond what is required by the vendor's preventive maintenance procedures) or are based on quality assurance programs, such as the American College of Radiology MR Accreditation Program, which were established to achieve goals that are not specifically related to *quantitative* imaging biomarker studies.

Image analysis factors. Most of the general image analysis sources of bias and variance are addressed in the companion manuscript on non–modality-dependent factors. However, there are some analysis factors that are unique to MR applications, including image intensity corrections to account for variations in B_1 radiofrequency coil sensitivity characteristics. More advanced MRI biomarker applications, such as DCE-MRI, are particularly dependent on image analysis factors, including vascular input function characterization, native tissue T_1 relaxation time measurements, imaging biomarker parameter normalization factors, and pharmacokinetic modeling assumptions, and algorithm implementations.

Interactions between sources of variance. As also noted in the companion articles [2–5], there are known and, undoubtedly, as yet unknown interactions between some of the aforementioned sources of variance, and these must be addressed in an application specific manner.

Methods

RIDER Investigations into Sources of Bias and Variance for MR Measures

In preliminary investigations performed by the authors under the auspices of the RIDER initiative, the techniques investigated were those most commonly used currently in phase 1/2 clinical trials. Companion articles discuss generic challenges related to quantitative imaging [2] and the importance of addressing these challenges [3]. Additional companion articles discuss modality-specific quantitative imaging issues as related to volume computed tomography and positron emission tomography/computed tomography measures [4,5]. This report discusses issues related to quantitative imaging applications using DCE-MRI and diffusion-weighted (DW) and diffusion tensor imaging (DTI) MRI techniques. Deidentified source image data and data analyses from each of the three projects were uploaded to the National Cancer Institute's National Biomedical Imaging Archive (NBIA, http://ncia.nci.nih.gov/collections). Three specific subprojects were developed:

Project 1: Phantom data obtained on four scanners from two vendors and at two field strengths to address 1) repeatability

(short-term and long-term) of T_1 measurements, 2) contrast-to-noise ratio and signal intensity stability, 3) limits of agreement of T_1 measurements using two acquisition techniques, and 4) between-vendor contrast differences and effect on computed contrast agent concentration results. (Institution A)

Project 2: Repeat, that is, assumed zero pathologic change, DCE-MRI and DT-MRI of brain tumors in patients with recurrent glioblastoma multiforme. The data obtained using DCE-MRI techniques, when appropriately analyzed using pharmacokinetic models, provide a means of assessing the endothelial transfer constant (K^{trans}), which describes the rate of contrast agent transfer from the vascular space to the extravascular, extracellular space (EES) and, particularly with currently approved low–molecular weight contrast agents, represents the extraction-flow product, k_{ep}, the reflux rate of the contrast agent from the EES back to the vascular space, v_p, the vascular volume fraction, and v_e, the EES volume fraction. The data obtained using DT-MRI techniques can be used to assess the apparent diffusion coefficient (ADC) and fractional anisotropy (FA) within enhancing lesions in these patients. (Institution B)

Project 3: Repeat, that is, assumed zero pathologic change, DW and ADC MRI of breast lesions. Changes in ADC have been previously demonstrated to compare favorably with changes in cellular volume fraction or density. (Institution C)

The three projects were chosen to satisfy the general goals of the RIDER group at large and were designed to address the following specific issues:

Accuracy, bias, and reproducibility using a test object:

Accuracy: If the same phantom is rescanned on the same device, how much variance is observed?

Reproducibility: If the phantom is scanned on a different device (same vendor as well as different vendor), how much variance is observed?

Bias: How closely, on average, can acquisitions on one vendor's scanner be matched on a scanner, at the same field strength, from a different vendor?

Patient accuracy:

If repeated scans are obtained on a given patient under a zero pathologic change assumption, how much variation is observed when the same acquisition, image analysis, and image-derived measurement calculation are used?

A third primary goal, focused on outcome measures, could not be addressed in the short period during which these studies were completed. This third important question addresses the critically important question of MRI biomarker change in patients, that is, when a patient is imaged over time, how much change can reliably be observed, and how well does such a change predict subsequent clinical outcome?

Each project provided to RIDER deidentified source data as well as results of test-retest reproducibility. All data and analyses have been made available through the NBIA so that they might be used by other researchers interested in change analysis in DCE-MRI, DW/ADC-MRI, or DT-MRI.

MR Data Made Publicly Available through RIDER on the NCI NBIA

Project 1: Phantom data to assess system bias and variance (Institution A). Data were obtained using an 18-compartment EuroSpin TO5 contrast response phantom (Diagnostic SONAR, Ltd, Livingston, West Lothian, Scotland). Data were obtained in "coffee break" fashion on the same day as well as 1 week later. In the coffee break measurement setting, the radiofrequency coil and phantom were positioned and data obtained and the phantom was then removed and positioned and scanned on a second nearby scanner. The phantom was subsequently returned to the first scanner, the radiofrequency coil and phantom were repositioned, and the data were obtained using a new examination ID but using the same stored data acquisition protocol. Data were obtained at all three time points from three 1.5-T scanners (two with differing gradient subsystems from a single vendor and one from a second vendor) and from one 3.0-T scanner. Data acquired included T_1 measurements using a two-dimensional inversion recovery (IR) sequence (once at 1.5 T and twice at 3.0 T) and a three-dimensional multiple flip angle fast-spoiled gradient-recalled echo (MF-FSPGR) sequence (each time on each scanner). A DCE-MRI data set was also acquired each time on each scanner using a three-dimensional FSPGR sequences with sections matching those acquired for the MF-FSPGR T_1 measurements.

For the fast spin-echo IR T_1 measurements, 10 inversion times were used (50, 100, 250, 500, 750, 1000, 1500, 2000, 2500, and 3000 milliseconds). For the MF-FSPGR T_1 measurements, seven flip angles were used (2°, 5°, 10°, 15°, 20°, 25°, and 30°). For the DCE-MRI acquisitions, the total scan time was approximately 7 minutes with a temporal resolution of approximately 9 seconds (dependent on specific scanner). Echo (TE) and repetition times (TR) were matched as closely as possible across scanners, with TE ranging from 0.90 to 1.35 milliseconds and TR ranging from 4.09 to 5.10 milliseconds. The flip angle was 30° for all DCE-MRI scans. For all acquisitions, the field of view was 24 × 19 cm, the acquisition matrix was 256 × 192, and the acquisition bandwidth was ±31.25 kHz.

Project 2: Repeated DCE-MRI and diffusion tensor imaging data in brain tumor patients (Institution B). In this study, repeat DCE-MRI data sets in 19 patients with recurrent glioblastoma multiforme and repeat diffusion tensor imaging (DTI) data sets in 17 of the 19 patients were performed. The interval between scans was 2 days or less. All images were obtained on the same 1.5 T scanner (Siemens Avanto, Siemens Medical Solutions, USA, Inc, Malvern, PA). In brief, for DCE-MRI, T_1 mapping using a multiflip angle approach with six flip angles was performed. Dynamic imaging was performed every 4.8 seconds during the intravenous infusion of 0.1 mmol/kg of gadolinium–diethylenetriamine penta-acetic acid at 3 ml/sec. DTI was performed using a 12 diffusion-sensitizing gradient direction sequence, TR of 6000 milliseconds, TE of 100 milliseconds, flip angle of 90°, signal averages of four, acquisition matrix of 128 × 128, voxel size of 1.7 × 1.7 × 5.0 mm, and a b value of 1000 sec/mm^2.

In addition to the source data, parameter maps for fractional plasma volume, volume of the extracellular extravascular space, and K^{trans} (from the DCE-MRI data), and ADC and FA (from the DTI data) were transferred to the NBIA. Segmentations of tumor-related enhancement and fluid-attenuated IR (FLAIR) signal abnormality volumes were also transferred.

Project 3: Repeated ADC diffusion of breast cancer (Institution C). In this study, during the first cycle of neoadjuvant chemotherapy, the following imaging protocol was implemented:

1. Two preinitiation (pretherapy) baseline MRI scans were typically obtained within 15 minutes, where initiation of neoadjuvant chemotherapy was no later than 1 week from the last baseline MRI; and
2. One postinitiation MRI was obtained within 8 to 11 days after initiation of neoadjuvant chemotherapy

The second cycle of neoadjuvant chemotherapy was implemented with no further imaging studies. In brief, the data were acquired in the axial plane using a Philips 3 T scanner (Philips Medical System, Eindhoven, the Netherlands) and the following acquisition parameters: field of view = 350 mm; acquisition matrix = 196 × 86; sensitivity encoding factor 2; 30 4-mm-thick sections; two-dimensional spin-echo–echo-planar imaging three-axis DWI at b = 0 and 800 sec/mm^2; short tau IR (fat suppression), TR/TE/TI = 12,000/62/150 milliseconds; signal average = 2; acquisition time = 4 hours 41 minutes.

Analyses of Bias and Variance

For each project, analyses of bias and variance were performed as follows:

Project 1. Limits of agreement and coefficients of repeatability, as proposed by Bland and Altman [6], were computed for the T_1 measurements obtained from each scanner using the IR-based and multiple flip angle–based data (limits of agreement) or multiple flip angle measurements at each time point (coefficients of repeatability) [7]. T_1 measurement correlation analysis was also performed using the data obtained at each time point on each scanner. Contrast response was assessed at each time point on each scanner using the multiple compartment phantom data, and stability of the CNR and signal intensity was assessed from each of the DCE scans obtained at each time point on each scanner. Simulated DCE uptake curves were also generated for each scanner using measured data from the multiple compartment phantom and commonly assumed signal intensity response for an ideal fast spoiled gradient echo sequence.

Project 2. A similar analysis of repeatability was performed on both the DCE-MRI and DTI data sets from brain tumor patients. For DCE-MRI parameters, mean values were obtained at both time points in areas of tumor-related enhanced segmented from three-dimensional isotropic T_1-weighted contrast-enhanced FLASH images. The coefficient of repeatability (Bland and Altman [6]) and the 95% confidence interval for percent change in parameter were computed [8,9]. For DTI parameters, mean values were obtained at both time points in both the areas of tumor-related enhancement and the areas of FLAIR signal abnormality segmented from the three-dimensional isotropic contrast-enhanced FLAIR images [10].

Project 3. Registration of the tumor in the interval examinations was implemented using MIAMI Fuse. Tumor volumes of interest were drawn on the anatomic image volume and were warped from the anatomic volume onto one of the pretherapy diffusion volumes denoted as the reference; warping is necessary owing to the susceptibility artifacts in the diffusion acquisitions that are not present in the anatomic volumes. Subsequent registrations, either between the two pretherapy examinations or the two pretherapy and posttherapy scans, were also warped to account for repositioning deformations to the breast as well as any small compartmental changes to the tumor. Warping was accomplished using thin-plate splines where the degrees of freedom (DOF) of the warp are related to the volume of the tumor. The user only needed to pick the location of three control points in the homologous tumor volume that approximate their loci in the reference tumor volume. The multiscale registration first implemented rigid-body registration, then low DOF warping, and finally full DOF warping.

ADC volumes were computed from the interleaved b = 0 sec/mm^2 and b = 800 sec/mm^2 acquisitions. For each pair of registered ADC images, a 128 × 128 joint density histogram (JDH) was constructed by incrementing the count of the two-dimensional histogram defined by the two ADC values of the registered tumors. For the JDH of the two pretherapy examinations, bias was removed from this realization and variance was generalized, that is, increased, by adding the transpose of its JDH to itself. Linear regression was then performed after rotating the JDH onto its principal component axes. The resulting linear estimate, the estimates of its 95% confidence limits, and the 95% confidence limits of the JDH were computed and plotted on the modified JDH. The means of both JDHs were computed and plotted. Note that the pretherapy/pretherapy JDH represents estimates of sources of noise associated with the null hypothesis, that is, the presence of all sources of noise encountered in estimating change between short-interval examinations, but no change in the tumor [11,12].

The detailed results from the analyses of bias and variance for each of the three projects will be provided on the NBIA Web site (http://ncia.nci.nih.gov/collections), along with all source data, and are not replicated in this article in the interest of brevity.

All human subject studies were approved by the respective institutional review board and written informed consent was provided by each subject.

Discussion and Conclusions

The focus of this article, like its companion articles addressing other modalities [4,5], has been the identification of sources of bias and variance and the measurement of bias and variance at the system level (phantom measures) and patient level ("coffee break" studies) in studies focused on applications of DCE-MRI, DW-MRI, and DT-MRI. The data obtained and analyzed by the authors, and made available through RIDER on the NCI NBIA, were described along with the rationale for the selection of these particular projects.

The characterization and investigations of the sources of bias and variance introduced in this article are essential, but are clearly not sufficient, for the establishment of the MRI biomarkers investigated in this study as surrogate markers of response. This essential remaining step requires carefully designed, multicenter trials with standardized acquisition protocols, quality assurance programs, and outcome measures highly correlated with disease status.

References

[1] Tofts P (2003). *Quantitative MRI of the Brain: Measuring Changes Caused By Disease.* Chichester, UK: John Wiley & Sons, Ltd.

[2] Meyer CR, Armato SG III, Fenimore CP, McLennan G, Bidaut LM, Barboriak DP, Gavrielides MA, Jackson EF, McNitt-Gray MF, Kinahan PE, et al. (2009). Quantitative imaging to assess tumor response to therapy: common themes of measurement, truth data, and error sources. *Transl Oncol* **2** (4), 198–210.

[3] Clarke LP, Croft BS, Nordstrom R, Zhang H, Kelloff G, and Tatum J (2009). Quantitative imaging for evaluation of response to cancer therapy. *Transl Oncol* **2** (4), 195–197.

[4] McNitt-Gray MF, Bidaut LM, Armato SG III, Meyer CR, Gavrielides MA, McLennan G, Petrick N, Zhao B, Reeves AP, Beichel R, et al. (2009). Computed tomography assessment of response to therapy: tumor volume change measurement, truth data, and error. *Transl Oncol* **2** (4), 216–222.

[5] Kinahan PE, Doot RK, Wanner-Roybal M, Bidaut LM, Armato SG III, Meyer CR, and McLennan G (2009). PET/CT assessment of response to therapy: tumor change measurement, truth data, and error. *Transl Oncol* **2** (4), 223–230.

[6] Bland JM and Altman DG (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**, 307–310.

[7] Bosca R and Jackson E (2009). Stability of T_1 relaxation time and DCE-MRI measures. Annual Meeting of the American Association of Physicists in Medicine, Anaheim, CA. *Med Phys* **36**, 2705, 2009.

[8] Roberts C, Issa B, Stone A, Jackson A, Waterton JC, and Parker GJ (2006). Comparative study into the robustness of compartmental modeling and model-free analysis in DCE-MRI studies. *J Magn Reson Imaging* **23**, 554–563.

[9] Paldino MJ, Desjardins A, Friedman HS, Vredenburgh JJ, and Barboriak DP (2009). Repeatability of parameters derived from dynamic contrast-enhanced MR imaging. *Proceedings of the American Society of Neuroradiology (ASNR),* Vancouver, BC, Canada.

[10] Paldino MJ, Barboriak D, Desjardins A, Friedman HS, and Vredenburgh JJ (2009). Repeatability of quantitative parameters derived from diffusion tensor imaging in patients with glioblastoma multiforme. *J Magn Reson Imaging* **29**, 1199–1205.

[11] Meyer CR, Schott AF, Chenevert TL, Galban C, Johnson TD, Rehemtulla A, Hamstra DA, and Ross BD (2008). Parametric response mapping (PRM): voxel-based analysis of quantitative diffusion MRI changes for individualized assessment of primary breast cancer response to therapy. *Proceedings of World Molecular Imaging Congress (WMIC'08),* Nice, France. Poster 500.

[12] Meyer C, Chenevert T, Galban C, Johnson T, Hamstra D, Rehemtulla A, and Ross B (2009). Parametric response mapping: a voxel-based analysis of quantitative diffusion MRI changes for individualized assessment of primary breast cancer response to therapy. *Proceedings 17th Scientific Meeting, International Society for Magnetic Resonance in Medicine (ISMRM),* Hawaii.