

Review

Quantitative MR in Multi-center Clinical Trials

Edward Ashton, PhD*

MRI has a wide variety of applications in the clinical trials process. MR has shown particular utility in the early phases of clinical development, when trial sponsors are interested in demonstrating proof of concept and must make decisions about allocation of resources to a particular compound based on the results from a small number of experimental subjects. This utility is largely due to the many different imaging endpoints that can be measured using MR, ranging from structural (tumor burden, hippocampal volume) to functional (blood flow, vascular permeability) to molecular (hepatic fat fraction, glycosaminoglycan content). The unique flexibility of these systems has proven to be both a blessing and a curse to those attempting to deploy MR in multi-center clinical trials, however, as differences among scanner manufacturers and models in pulse sequence implementation, hardware capabilities, and even terminology make it increasingly difficult to ensure that results obtained at one center are comparable to those at another. These problems are compounded by the differences between the procedures used in clinical trials and those used in routine clinical practice, which make trial-specific training for site technologists and radiologists a necessity in many cases. This article will briefly review the benefits of including quantitative MR imaging in clinical trials, then explore in detail the challenges presented by the need to develop and deploy a detailed MR protocol that is both effective and implementable across many different MR systems and software versions.

Key Words: clinical trials; protocol design; protocol harmonization; MRI

J. Magn. Reson. Imaging 2010;31:279–288.

© 2010 Wiley-Liss, Inc.

IMAGING IN GENERAL, and MRI in particular, has long had a role in the drug development process (1,2). MRI has much to recommend it as an assay in a clinical trial. It is a versatile technique which is able to assess both structure and function, and aside from some concerns around contrast toxicity (3), it is almost entirely noninvasive. MRI has played an important part recently in clinical trials in a wide variety

of indications, including acute stroke (4,5), advanced solid tumors (6), Alzheimer's disease (7), and heart failure (8). However, a close reading of the literature surrounding the use of MRI in the clinical trial process shows that, in most cases where MRI has played a prominent role in a particular trial, one (or in some cases both) of two conditions has prevailed: either the trial is very small, carried out at a few prominent research centers, or the endpoints derived from MRI are partially or mostly subjective, and heavily dependent on radiologist interpretation.

The large number of small trials involving MRI is partially explained by the role that MRI (and imaging in general) fills in the clinical development process. The number of image derived endpoints that have been used as primary justification for preliminary or final drug approval is small, and of that list (tumor burden assessed by means of RECIST criteria [9], changes in rheumatoid arthritis patients assessed by means of Sharp score [10], changes in numbers of enhancing multiple sclerosis (MS) lesions [11]) only MS lesions changes are assessed primarily through MRI. In contrast, a wide variety of MR imaging techniques has been used in early phase clinical trials in many different indications. These have included measurement of microvascular parameters (blood flow, vascular permeability) in solid tumors using dynamic contrast-enhanced MRI (DCE-MRI) (6,12), assessment of erosive damage and changes in the synovium in rheumatoid arthritis patients (13), and assessment of coronary wall thickness in patients with ischemic heart disease (14) among many others.

The scarcity of quantitative imaging endpoints in general, and MRI in particular, in larger late phase clinical trials is largely a result of the fact that drugs are approved based on clinical benefit, and it can be very difficult to definitively relate changes in any particular biomarker or finding to hard endpoints like survival or changes in disability scores. Image derived endpoints can often be much more easily related to acute drug effects. For example, tumors in a cancer patient who has received an effective dose of an anti-angiogenic or antivascular agent (15,16) should experience an acute reduction in blood flow and/or vascular permeability. These parameters can be measured directly in vivo using DCE-MRI (6,12,17). This assay may be very useful to a drug developer who is trying to determine whether his experimental compound is hitting its intended targets, but even a large reduction in tumor blood flow may or may not be directly related

VirtualScopies, Inc., Rochester, NY, USA.

*Address reprint requests to: E.A., VirtualScopies, Inc., 500 Linden Oaks, Second Floor, Rochester, NY 14625.

E-mail: ashton@virtualscopies.com

Received July 31, 2009; Accepted October 14, 2009.

DOI 10.1002/jmri.22022

Published online in Wiley InterScience (www.interscience.wiley.com).

to improved patient survival or quality of life. Techniques of this sort are, therefore, very useful in the early phases of drug development, when the emphasis is on learning about the mechanism of action and acute effects of a drug, but less so in the later stages, when the focus turns to achieving regulatory approval.

In considering those late phase trials that do use MRI in a significant way, the prominent use of subjective or categorical rather than quantitative endpoints is at first glance somewhat puzzling. Continuous, quantitative endpoints (lesion volume, blood flow) are more easily dealt with from a statistical standpoint than categorical or subjective endpoints, and should be more easily correlated to pharmacokinetic variables. An example of such a correlation is given in Figure 1, which shows the relationship established in a phase I clinical trial between acute change in the vascular parameter K^{Trans} (18) measured using DCE-MRI and exposure to the intravenously administered vascular disruptive agent MN-029 (19). The demonstration of the existence of a strong PK-PD relationship in this relatively small trial provides a good indication that this compound is behaving as expected in the human population despite the fact that this trial was not powered sufficiently to demonstrate efficacy by means of RECIST (9) or overall survival.

Depending on how they are measured, quantitative imaging endpoints may also avoid the problem of extreme dependence on reader interpretation. Quantitative measures can be conveniently characterized in terms of intra- and interobserver variability (20–22), making it relatively simple to model the effects of using two or more readers. Variability between readers tends to be both greater and more difficult to model for categorical or subjective endpoints (23,24). This difficulty is well understood in the drug development community. Inter-reader variability in studies using heavily reader-dependent endpoints is sometimes mitigated by requiring that a single reader analyze all cases for a given cohort or entire study. While this can be an effective approach for relatively small studies, it rapidly becomes impractical for studies that are very large or that stretch over several years.

The great advantage of subjective or categorical endpoints in large clinical trials is that the mediation of the reader partially obviates the need to standardize and verify the acquisition protocol across sites. A radiologist's interpretation of radiological progression in an osteoarthritis patient may be relatively insensitive to the differences between one imaging system and another, while the precise measurement of cartilage thickness, for example, may be much less so (25).

As recently as 2005, an editorial in the *Journal of Clinical Oncology* expressed some surprise that quantitative blood flow measurements using MRI could be carried out successfully and consistently in even a three-site study (26). It is clear from the study referenced in this editorial (6) as well as from later, larger efforts at generating quantitative MR data across multiple sites and multiple imaging platforms, including the Alzheimer's Disease Neuroimaging Initiative (ADNI) (27) and the Osteoarthritis Initiative (OAI) (28),

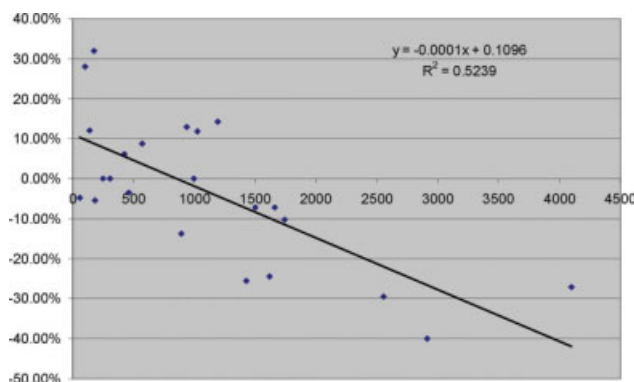


Figure 1. Change in K^{Trans} versus AUC(0–24 h) for MN-029. The correlation seen here is statistically significant ($P < 0.001$). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

that careful attention to cross-platform protocol design, site training, equipment assessment including periodic phantom scanning, and ongoing site monitoring are vital to achieving reliable results in a multi-center trial involving quantitative MRI.

CHALLENGES

Protocol Development

One of the most important challenges in implementing a complex quantitative MRI protocol in a multi-site clinical trial is developing a protocol that is sufficiently specific to guarantee the ability of each site to meet the informational requirements of the study while also allowing enough flexibility to accommodate the different scanner makes and models, coils, and software versions available at the various sites. This frequently entails the development of what are sometimes derisively referred to as “least common denominator” protocols—the implication being that these protocols are inadequate because they fail to take advantage of the full capabilities of the sites with the most advanced equipment. However, it should be recognized that the importance of obtaining comparable results across sites in a study generally outweighs the small improvements in precision that might be obtained by, for instance, using a 3 Tesla (T) system at one site when another site is limited to 1.5T.

As an example, consider the development of a protocol for a multi-site study whose intent is to measure changes in abdominal fat and hepatic fat fraction—a parameter set which is of interest for programs addressing obesity, type 2 diabetes mellitus, and as a safety marker in a variety of drug development programs. Options for a protocol of this sort include in-phase and out-of-phase (IP/OP) GRE imaging (29), two-point or three-point Dixon techniques (30), or variations such as General Electric's IDEAL, which makes use of asymmetric echo times to resolve dominant fraction ambiguities (31,32).

Each of these options carries advantages and disadvantages in terms of both effectiveness and ease of deployment. IP/OP imaging is probably the simplest

Table 1
Examples of Differences in Basic Terminology Among MRI System Vendors

Sequence or term	Siemens	GE	Philips
Spoiled gradient echo	FLASH	SPGR	T1FFE
Steady state free procession	TrueFISP	FIESTA	Balanced-FFE
Parallel imaging	IPAT	ASSET	SENSE
Repeated measurements	Acquisitions, no. of averages	NEX	NSA
Oversampling in phase	Phase oversampling	No Phase Wrap	Fold over suppression
Half Fourier imaging	Half Fourier	1/2 NEX, Fractional NEX	Half scan, HS
Partial echo	Asymmetric echo	Fractional echo	Partial echo

of these techniques to implement, and it is a relatively straightforward matter to develop comparable IP/OP protocols across the major system vendors. However, this method fails to account for the effects of iron in liver tissue—an important consideration in some indications. It is also unable in its most basic form to distinguish between water dominant and fat dominant pixels and requires additional estimation of T2* variations (33). Dixon techniques also fail to compensate for iron effects, require significantly more complex postprocessing, and are sensitive to both phase and motion errors. IDEAL may provide the most accurate and robust analytical results (34), but its availability is severely limited. This method can currently be implemented only on a subset of GE systems, and T2*-IDEAL (35), which incorporates estimation of T2* variations, is currently available at only six sites worldwide.

Given these constraints, the most appropriate method will depend on the nature of the study being planned. For a small study across few sites, particularly if it is possible to limit the study to a single scanner manufacturer, three-point Dixon or IDEAL may provide the best results. For a larger study across many sites, some of which may be community imaging centers rather than major research institutions, IP/OP imaging may provide the best chance at obtaining reliable and consistent results.

Once a basic approach has been selected, the next step is to design a separate specific protocol for each system vendor that will be represented in the study. This is necessary both because of differences in the availability of specific sequences as well as differences between vendors in basic terminology, some examples of which are given in Table 1. It is not generally necessary to design distinct protocols for different models produced by the same vendor. However, in some cases (1.5T versus 3T systems, for example) this may also be required.

It should be understood that the goal of this process is harmonization rather than standardization, which is generally not possible. As an example, consider the design of a DCE-MRI protocol to be implemented on both Siemens and GE 1.5T systems. The design goal is to produce an oblique coronal three-dimensional SPGR sequence using the inherent body coil for both transmit and receive, with a slab thickness of 8–10 cm, 5- to 8-mm slice thickness, and temporal resolution of 6–10 s per slab. It is possible to meet these design goals using either system, but there will be

specific differences between the two prescriptions, some of which are outlined in Table 2.

Once these protocols have been designed, it is critical to test them to ensure that they will produce comparable results. In the case of DCE-MRI, the most important performance characteristic is the interaction between signal changes observed in the images, changes in baseline T1, and changes in gadolinium concentration in tissue. This relationship can be assessed by means of phantom studies (36). Typical results of this assessment for GE and Siemens systems are given in Figure 2. Note that, in addition to trivial scaling differences, there are some differences between these two protocols in linearity—in particular, the Siemens system begins to show roll-off in response at longer T1 values than the GE system. This difference is not critical, but will need to be accounted for in the modeling to prevent it from translating into a significant difference in measured parameters.

Site Evaluation and Training

Clinical trial sites, even in studies in which imaging plays an important role, are generally selected with little or no consideration of their ability to implement the required imaging protocol. The driving force in the selection process is most often the site's ability to recruit patients, for the simple reason that patient recruitment is the factor with the single largest influence on the duration of the trial. Because of this reality, it is often the case that the selected sites do not have the optimal equipment or expertise to implement a complex imaging protocol. It is the responsibility of the organization managing the imaging component of the study to evaluate the imaging capabilities of the selected sites and determine whether they will be able to meet the requirements of the study. This process generally begins with a survey in which the site is

Table 2
Differences Between GE and Siemens Implementations of a Basic DCE-MRI Protocol

Parameter	GE	Siemens
Sequence type	SPGR	SPGR
Sequence variants	SS\SP\SK	SP\OSP
Sequence options	FAST\WB\EDR\MP\FFF	FPF
TR/TE/FA	4.62/1.05/30	5.00/1.57/30

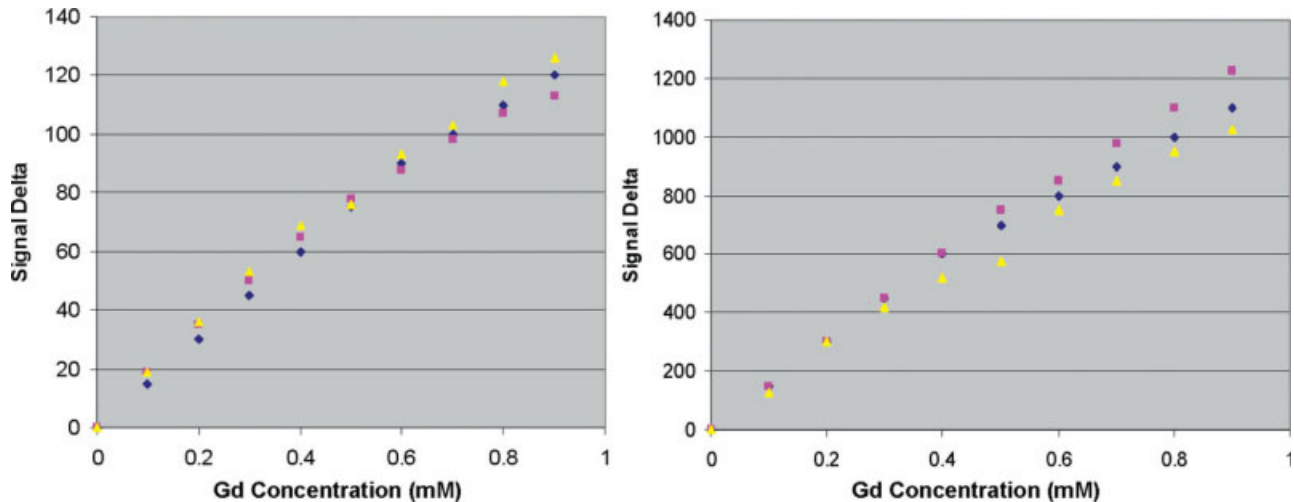


Figure 2. Plot of signal change versus gadolinium concentration for GE (l) and Siemens (r) systems. Baseline T1 values were 90 ms (yellow), 300 ms (pink), and 700 ms (blue). Note that the Siemens system shows roll-off in signal response at significantly longer T1 than the GE system.

asked to list pertinent information including available equipment, software versions, experience, and personnel qualifications. This may be followed by more detailed discussions by means of telephone or email in cases where the site's capabilities are unclear. Finally, a determination is made as to the suitability of the site for the proposed protocol.

If the answer is no, it may be necessary to locate another facility near enough to the clinical site that patients can be conveniently sent there for imaging. Otherwise, the next step is to develop an imaging manual—a document describing the imaging protocol in detail—and to train the site personnel in the implementation of the imaging protocol and in procedures for proper handling of clinical trial data (37,38). In general it is not advisable to deliver the imaging manual to the sites without further contact or follow-up and assume that the protocol will be properly implemented. The extent of training that is necessary will depend on the complexity of the protocol and its similarity to protocols used in clinical practice.

As an example, the typical protocol for a multiple sclerosis clinical trial (39) is very similar to that which is used in the clinical evaluation of MS patients. Consequently, in an MS trial with reasonably basic imaging requirements the technologists and radiologists at any site that handles a substantial number of MS patients will be generally familiar with this protocol and should be able to implement it with a minimum of guidance. Training in this case should touch lightly on the imaging protocol, but should focus on necessary clinical trial and data handling procedures (de-identification of patient data, image archival and transfer, etc.), which do differ substantially from day-to-day clinical practice. Training of this sort can usually be conducted remotely as a webinar or even as a teleconference.

In contrast, the protocols typically used in a trial involving DCE-MRI (40) are not commonly used in a clinical setting, and are likely to be very unfamiliar to personnel at sites other than certain major research

institutions. Training in this case will be by necessity more extensive and will in most cases require the training personnel to be present at the imaging site. In addition to the issues mentioned above, training in a case such as this should include actual implementation of the imaging protocol on the site magnet and a complete test run using an appropriate phantom and/or a human volunteer (use of a volunteer during training activities is permitted at some sites and prohibited at others). The phantom should be complex enough to demonstrate that the system is functioning properly, but does not necessarily need to mimic relevant human anatomy. Some examples of phantom designs are given here for assessment of uniformity and linearity (41), relationship between signal and T1 (36), and signal-to-noise ratio (SNR) and contrast assessment (27).

After the completion of training, data acquired during the training session should be carefully evaluated in terms of both conformance to the imaging protocol and general image quality. The imaging site should not be approved to begin scanning clinical trial patients until all training data have been evaluated and judged to be acceptable.

Monitoring Site Performance

The importance of quality control does not diminish once a site has been approved to participate in a clinical study and has begun recruiting and imaging patients. Many factors can cause a site that has been screened and approved to fall out of compliance with protocol and quality requirements. Sites should be asked to report any events that might cause a disruption in image quality, including loss of key personnel such as the lead technologist or radiologist for the study, significant imaging system maintenance, or upgrades to the system hardware or software. Any of these events should prompt at a minimum a phantom re-scan to verify that the site is still able to effectively implement the imaging protocol, and that data

acquired postevent will be comparable to that acquired earlier. This is particularly critical if the event occurs while patients who have already received baseline scans are still on study.

Even in the absence of an event such as those described above, it is important to continuously monitor both image quality and adherence to protocol for all sites throughout the course of a clinical trial, and particularly so for sites that recruit relatively few patients, because weeks or even months may pass between such sites seeing study patients, and critical points of the protocol or data handling procedures may be forgotten. This is generally accomplished through two channels: periodic phantom re-scans, and continuous monitoring of acquired patient data.

Scans of all qualification phantoms should be repeated on a quarterly basis at a minimum, and more frequently in cases such as cartilage thickness measurement (42), where minor system drift might cause significant problems with data interpretation. Many imaging sites are reluctant to perform frequent phantom scans, generally because they absorb valuable magnet time. It is important, therefore, to make certain that sites are appropriately compensated for phantom scans. Otherwise, compliance with the phantom regimen is likely to be low, and the risk of introducing flawed patient data into the study is likely to be commensurately high.

It should be noted that there are two possible ways to use the phantom data once they have been acquired. The most conservative (and common) approach is to use the phantoms strictly as a quality check on the site's equipment and procedures. Interpretation in this case is fairly straightforward. The data are compared against a reference standard and graded as either pass or fail, with failure resulting in an escalation process generally beginning with a phantom re-scan, progressing through system trouble shooting and potential maintenance, and leading eventually to site disqualification if no better resolution is possible.

Another less commonly used approach is to use phantom data to make adjustments or corrections to the acquired data before analysis. An example is the use of data from a uniformity and linearity phantom (see Fig. 3) on magnet drift to make adjustments to measured thicknesses and volumes of cartilage. While it may be necessary in some limited cases, this approach is problematic from a statistical standpoint. This is because there is an error associated with any correction factor measured from the acquired phantom data. When the correction factor is applied to the patient data, this error will propagate through to the measured parameters. Therefore, in the case (which one would hope is most common) where there is no true drift, application of the correction factor will necessarily degrade the quality of the patient data. Application of phantom derived correction factors is only advisable in cases where it can be demonstrated that the improvement in absolute accuracy introduced through this process outweighs the necessary degradation in precision.

In addition to phantom evaluations, it is vital to provide continuous, real-time evaluation of all

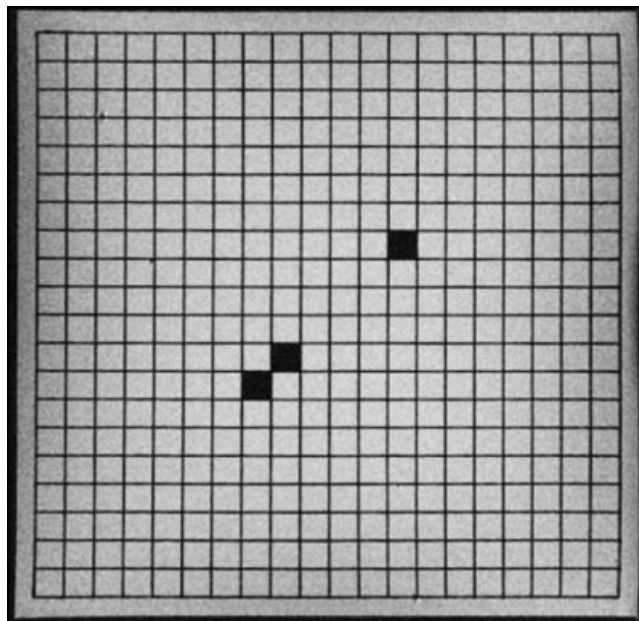


Figure 3. A coronal plane scan of a uniformity and linearity phantom. Knowledge of the true phantom grid size allows correction of minor errors in pixel resolution in the reconstructed image.

acquired patient data. Data need not be analyzed as they are acquired (although this is desirable) but they should be quality controlled by the analyzing organization in as close to real time as is practically possible. At a minimum, the data headers should be checked against the imaging protocol to ensure continued compliance, and all images should be visually inspected for artifacts and any signs of imaging system degradation. Maintaining a rigorous inspection regimen will not absolutely prevent loss of patient data, because in many protocols re-scan of incorrectly acquired data is impractical or impossible. It will, however, help to limit such losses to a small number of isolated events.

CASE STUDIES

Imaging Cartilage

It is well accepted that one of the markers of advancing osteoarthritis is a breakdown and eventual loss of cartilage in the joints (43). The ability to effectively image cartilage is critical for the development of both disease modifying osteoarthritis drugs (DMOADs) (44) and cartilage repair devices (45). Structural changes in cartilage, including loss of cartilage volume and reduction of joint space, are typically measured using either x-ray or MRI. However, these parameters change slowly, and are not sensitive to the changes in cartilage composition that are early hallmarks of osteoarthritis. Moreover, cartilage volume has not been shown to be a good correlate to advancing disease. Consequently, a clinical study designed to assess the effects of a DMOAD or cartilage repair device will typically include both structural endpoints (cartilage volume or thickness) and functional or

molecular endpoints such as assessment of cartilage water and collagen content by means of T2 measurement (46,47) or assessment of cartilage glycosaminoglycan content by means of delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) (48,49), which requires direct measurement of T1 values.

The first step in preparing for a clinical trial of this sort is the development of a general protocol. This will most likely include a T1-weighted SPGR with fat saturation and a T2* weighted GRE sequence to provide structural measurements. These sequences are reasonably straightforward, and can be implemented readily on most platforms. T1 and T2 measurement, in contrast, present several options, each with advantages and disadvantages. The basic process for T2 measurement is relatively simple: images are acquired at multiple echo times, and T2 is estimated from the negative slope of the natural log of the resulting signal intensities. There are three commonly used techniques for acquiring the necessary images: (i) multiple single-echo SE, (ii) single series multiple echo SE, and (iii) dual echo FSE.

The first option will theoretically provide the most accurate results. However, to provide significant advantage over the simpler dual echo option, at least four acquisitions are required, with each acquisition taking roughly 6 min of table time. While this may not seem excessive, 24 min without movement is a difficult requirement for an osteoarthritis patient, and the morphology of articular cartilage virtually guarantees significant errors in pixels affected by motion.

Acquiring four or more echoes in a single series is substantially faster and, therefore, less prone to motion artifacts than the first option. However, this approach will result in biased measurements due to the presence of stimulated echoes, which will affect the magnitude of each echo after the second (50).

A single dual echo acquisition, while not as robust to widely varying T2 values as multiple single echo acquisitions, is fast enough to be clinically practical and has been shown to be accurate within a reasonably broad range of T2 values which is defined by the specific echo times selected (51). In many cases, this technique will provide the best option for T2 measurement in a large multi-site clinical trial.

Similarly, several options are available for making the T1 measurements required by dGEMRIC: (i) acquisition using multiple inversion recovery times (48), (ii) acquisition using multiple flip angles (52), and (iii) acquisition using the Look-Locker method (53).

Multiple IR acquisition is generally considered the gold standard for accuracy in this case. However, because this method requires a nonlinear fit involving three parameters, 5 or more acquisitions totaling 30–40 min of table time are required to achieve adequate results.

Both the multiple flip angle and Look-Locker methods are substantially faster, generally requiring less than 10 min in either case, but either may be somewhat less robust, particularly to T1 values outside the expected normal range in cartilage. Additionally, Look-Locker requires a somewhat more complex acquisition, which may necessitate more rigorous site

training. The optimal technique for a given study will depend on the size of the study, the number of sites that will be involved, and the number of other sequences that are being acquired. Multiple IR acquisition, for instance would probably not be acceptable in a protocol requiring 40 min or more in addition to the dGEMRIC acquisition. If only dGEMRIC were being acquired, however, multiple IR acquisition might be a more attractive option.

Once a protocol has been finalized, it will be necessary to design vendor-specific implementations and develop appropriate site documents as outlined previously, and to train the individual imaging sites in the conduct of the protocol. The critical question at this point is whether remote training by means of Web conference will be adequate, or if instead an on-site visit will be required for each imaging site. This is an important question in terms of both budget and timelines. On-site training is typically more expensive than remote training by a factor of 5 or more, and in a large study this can have a substantial impact on the overall imaging budget. Perhaps more significantly, on-site training requires a much greater commitment of time on the part of both the training organization and the sites being trained. This may make it more difficult to schedule training sessions, and, therefore, more difficult to bring sites on-line quickly enough to satisfy overall study timelines.

In a study of this sort, however, on-site training will probably be necessary unless the study is limited to sites with substantial experience in T2 and dGEMRIC measurements. The T1 and T2 mapping techniques used in this protocol are not ones that are commonly used in clinical practice, and the patient preparation for dGEMRIC in particular is fairly complex. Conducting remote training sessions in this case will run a high risk of faulty data acquisition which will quickly swamp any cost savings associated with avoiding site visits.

Quality control will also be critical in a study of this sort, particularly in the early stages when new sites are being brought on-line. Continuous monitoring of incoming data, with particular attention to the first patient data acquired at each new site, will help to ensure that any errors that occur are limited to individual data sets and not propagated through all patients from a given site.

Imaging Blood Flow and Vascular Permeability

There are numerous cancer therapeutics either approved or currently in development that are either anti-angiogenic or vascular disruptive agents (6,15,16). The ability to accurately and precisely estimate parameters related to blood flow and vascular permeability is critical to the early evaluation of these compounds, as this gives the most direct window into their targeted biological effects. The most common technique for estimating these parameters is DCE-MRI (18).

DCE-MRI involves the periodic acquisition of T1-weighted images before, during and after injection of a gadolinium labeled tracer such as gadopentetate dimeglumine. The change over time in signal intensity

in a voxel or region of interest in this time series can then be related to contrast agent concentration in that tissue. By making use of a two-compartment model, with one compartment representing blood plasma and the other extravascular extracellular space (EES), the observed enhancement curves in tissue and plasma can be used to estimate various physiological parameters (54,55).

The design of a DCE-MRI protocol for a multi-site clinical trial involves several tradeoffs. The speed at which DCE-MRI data must be acquired—typically from 2 to 10 s per slab depending on the specific protocol—is stressing, and requires concessions in terms of both anatomical coverage and signal quality. The first question that must be answered, therefore, is this: what information does the protocol require, and what points can be compromised?

In particular, it is important to know whether the protocol requires the ability to measure flow and permeability independently. This is because the modeling used to estimate these parameters requires continuous time-concentration curves (see Fig. 4) which are reconstructed from discrete samples obtained at the per-slab imaging rate. To reconstruct these curves with reasonable accuracy, the imaging rate must be equal to or faster than the Nyquist rate (56) for the curves, which is determined by their rate of change. The time-concentration curve in arterial plasma, generally referred to as the arterial input function (AIF), has two distinct phases. During the first 15 s or so after injection it changes rapidly, with a Nyquist rate of ~ 2 s. Thereafter it changes much more slowly, with a Nyquist rate closer to 15 s. This is critical, because models that allow the separation of flow and permeability (57) require accurate reconstruction of the first bolus passage, and, therefore, require very rapid imaging. Models that produce composite parameters (18) generally do not require an accurate reconstruction of the first bolus passage, and so can make use of substantially slower protocols with correspondingly higher SNR and/or greater spatial coverage. Additionally, if a high speed protocol is selected, it should be understood that this will limit the sites that are able to participate in the study to those with newer and higher-end equipment.

Given these considerations, it becomes important to understand what compounds may require direct evaluation of blood flow and vascular permeability, and which can be accurately evaluated using composite parameters such as K^{Trans} (18). The general effect of most anti-angiogenic (6) and antivascular (16) compounds is to prune away, either temporarily or permanently, the immature or poorly formed microvasculature that is commonly found in malignancies.

This process, if successful, will have several measurable effects. Vascular permeability should decline as small vessels with poorly formed endothelia are destroyed. The effects on flow, however, may vary depending on the extent of the vascular disruption and the size of vessels that are damaged. If only the smallest and most immature vessels are removed, flow may actually increase as the vascular bed regularizes and interstitial pressure within the tumor

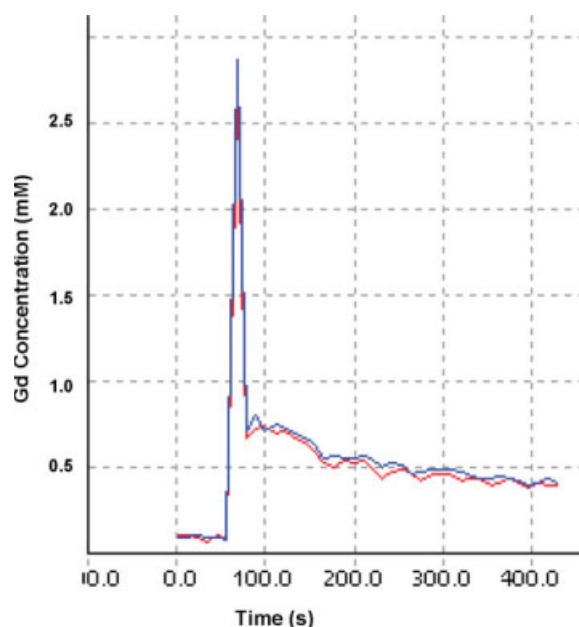


Figure 4. Typical time-concentration curve, with concentration of Gd (mM) plotted against time (s), for arterial plasma taken from a DCE-MRI protocol with temporal resolution of ~ 8 s per slab. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

decreases. In this case, a high speed protocol able to separate flow and permeability effects will be necessary to detect the treatment effects. If the damage progresses to the tumor's feeder vessels, however, flow will decrease—dramatically, in some cases (6). In this case, it is sufficient to measure a composite parameter, because flow and permeability will change in concert.

In general, vascular disruptive agents operate by inducing reversible ischemia leading to cell death, and so will need to produce large reductions in blood flow to provide clinical benefit. VEGF and multi-kinase inhibitors, particularly those intended for use as monotherapies, will likewise need to produce large flow reductions to be effective. These compounds may be effectively evaluated using slower acquisition protocols and composite parameters. Platelet-derived growth factor inhibitors (59) and other anti-angiogenic agents designed for use in combination therapy often operate by increasing flow to the malignancy while reducing vascular permeability. The effects of these compounds will not be detectable using composite parameters like K^{Trans} . For these compounds, the higher speed protocol will be necessary.

A second question is whether T1 mapping will be required (36), and if so, how it will be acquired—options here are similar to those in the previous section, with the added complication that image acquisition will often be in the chest or abdomen rather than the knee for this protocol.

Because DCE-MRI analysis is reliant both on the speed of acquisition and the ability to accurately infer gadolinium concentration in tissues from observed changes in MR signal, site evaluation by means of phantom studies (36) before the initiation of patient

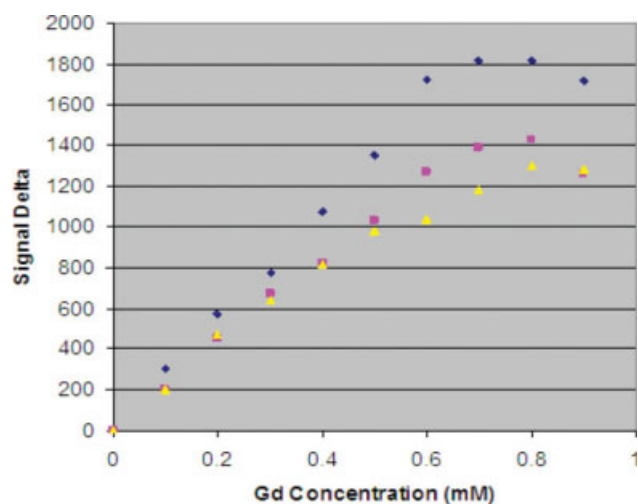


Figure 5. Plot of signal change versus millimolar gadolinium concentration using an 8 channel torso phased array coil. Note the non-monotonic relationship as well as severe dependence on initial T1 in comparison to the plots shown in Figure 2.

scanning is particularly critical in this sort of trial. Additionally, because DCE-MRI is very dissimilar to most routine clinical protocols, training of site personnel is also extremely important. As with the prior example, it is advisable in this case to conduct training in person rather than by means of Web conference. Phantom studies (and if possible a volunteer study without gadolinium injection) should be carried out during the course of the training, allowing the trainer to verify first that the site scanner is able to acquire data at the necessary rate, and second that the site personnel fully understand the implementation of the protocol. Phantom and volunteer studies should be closely evaluated by the training organization before the initiation of patient scanning, and phantom studies should be repeated on a quarterly basis throughout the trial to ensure continued compliance.

CONCLUSIONS

MR imaging has proven over the past 20 years to be a valuable tool in the drug development and clinical trials process, with established applications in a wide range of indications including joint disease, neurology, oncology, and cardiovascular disease. However, there are several challenges yet to be met as the field transitions from subjective interpretation to quantitative measurement. The extreme flexibility and adaptability of MR compared with other imaging platforms is both a blessing and a curse. It is this flexibility that permits MR to be useful in such a wide range of indications and applications—but at the same time, it is the very flexibility of the system that makes it so easy for things to go disastrously wrong.

A good example of this dilemma is seen in the choice of coils for acquisition of DCE-MRI data in the

chest and abdomen. Because DCE-MRI analysis is dependent on being able to accurately infer gadolinium concentration from changes in the MR signal, it is vital for the relationship between signal delta and gadolinium concentration to be both consistent and well characterized. However, because the imaging must be done very quickly and, therefore, at low SNR, it is also important to use the most sensitive receiver coil possible. This imperative often leads to the selection of a phased array coil for this sort of protocol rather than the inherent body coil.

While the use of a phased array coil does lead to images with a more pleasing visual appearance and better SNR, it may in some cases have severe negative effects on the consistency of the relationship between signal change and gadolinium concentration. A plot of this relationship from a typical phantom study using an eight-channel torso phased array coil is shown in Figure 5. Comparison to similar plots shown in Figure 2, which were obtained in both cases using the inherent body coil, shows a significantly more nonlinear (and indeed nonmonotonic) relationship between signal delta and gadolinium concentration as well as a drastically increased dependence on initial T1. Both these effects are the result of spatially varying sensitivity in the phased array coil. This variation can be mitigated by thorough maintenance and calibration of the phased array coil. However, not all clinical sites routinely carry out this sort of maintenance and calibration, making the choice to use a phased array coil in a study of this sort a potentially risky one. The inherent body coil, while less sensitive overall, will often have a far more uniform sensitivity profile across the field of view.

This difference can be critical to DCE-MRI analysis, because the AIF and tumor uptake curves will often be taken from anatomically distant locations. As a result, DCE-MRI studies which attempt to derive the AIF directly from the data (as opposed to those that use a generic population AIF) (59,60) will generally show significantly greater variability if the data are acquired with a phased array coil rather than the inherent body coil. Problems of this sort are not seen in similar techniques using other imaging platforms, such as dynamic contrast enhanced CT (61).

To take advantage of the flexibility and power of quantitative MR imaging, therefore, it is necessary to pay meticulous attention to each stage of the image acquisition and analysis chain: (i) protocol development and testing, (ii) site qualification and training, (iii) ongoing site monitoring and quality control, and (iv) analysis software and procedures.

At each link in the chain for a given trial it is vital to identify possible failure points, which may be more numerous for MR than for other potential imaging modalities, and to ensure that they are avoided or ameliorated to the greatest possible extent. If this process is followed with both rigor and care, however, the quality and quantity of information that can be generated can provide a development team with critical insight into drug mechanism of action and clinical efficacy, allowing them to either accelerate or terminate the compound development with confidence.

REFERENCES

1. Harris J, Frank J, Patronas N, McFarlin D, McFarland H. Serial gadolinium-enhanced magnetic resonance imaging scans in patients with early, relapsing-remitting multiple sclerosis: implications for clinical trials and natural history. *Ann Neurol* 1991; 29:548–555.
2. Paty D, Li D. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. II: MRI analysis results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology* 1993;43:662–667.
3. Marckman P, Skov L, Rossen K, et al. Nephrogenic systemic fibrosis: suspected causative role of gadodiamide used for contrast-enhanced magnetic resonance imaging. *J Am Soc Nephrol* 2006; 17:2359–2362.
4. Daffertshofer M, Gass A, Ringleb P, et al. Transcranial low-frequency ultrasound-mediated thrombolysis in brain ischemia: increased risk of hemorrhage with combined ultrasound and tissue plasminogen activator: results of a Phase II clinical trial. *Stroke* 2005;36:1441–1446.
5. Warach S. Use of diffusion and perfusion magnetic resonance imaging as a tool in acute stroke clinical trials. *Curr Control Trials Cardiovasc Med* 2001;2:38–44.
6. Liu G, Rugo H, Wilding G, et al. Dynamic contrast-enhanced magnetic resonance imaging as a pharmacodynamic measure of response after acute dosing of AG-013736, an oral angiogenesis inhibitor, in patients with advanced solid tumors: results from a Phase I study. *J Clin Oncol* 2005;23:5464–5473.
7. Tuszynski M, Thal L, Pay M, et al. A phase I clinical trial of nerve growth factor gene therapy for Alzheimer disease. *Nat Med* 2005; 11:551–555.
8. Hendrikx M, Hensen K, Clijsters C, et al. Recovery of regional but not global contractile function by the direct intramyocardial autologous bone marrow transplantation: results from a randomized controlled clinical trial. *Circulation* 2006;114:I-101–I-107.
9. Therasse P, Arbuck E, Eisenhauer E, et al. New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst* 2000;92:205–216.
10. Sharp J, Lidsky M, Collins L, Moreland J. Methods of scoring the progression of radiologic changes in rheumatoid arthritis. Correlation of radiologic, clinical and laboratory abnormalities. *Arthritis Rheum* 1971;14:706–720.
11. Miller D, Khan O, Sheremata W, et al. A controlled trial of natalizumab for relapsing multiple sclerosis. *N Engl J Med* 2003;348:15–23.
12. Mross K, Dreves J, Muller M, et al. Phase I clinical and pharmacokinetic study of PTK/ZK, a multiple VEGF receptor inhibitor, in patients with liver metastases from solid tumors. *Eur J Cancer* 2005;41:1291–1299.
13. De Vita S, Zaja F, Sacco S, De Candia A, Fanin R, Ferraccioli G. Efficacy of selective B cell blockade in the treatment of rheumatoid arthritis: evidence for a pathogenetic role of B cells. *Arthritis Rheum* 2002;46:2029–2033.
14. Laham R. Intracoronary basic fibroblast growth factor (FGF-2) in patients with severe ischemic heart disease: results of a Phase I open-label dose escalation study. *J Am Coll Cardiol* 2000;36: 2132–2139.
15. Herbst R, Onn A, Sandler A. Angiogenesis and lung cancer: prognostic and therapeutic implications. *J Clin Oncol* 2005;23:3243–3256.
16. Anderson H, Yap J, Miller M, Robbins A, Jones T, Price P. Assessment of pharmacodynamic vascular response in a phase I trial of combretastatin A4 phosphate. *J Clin Oncol* 2003;21:2823–2830.
17. Jackson A, Haroon H, Zhu X, Li K, Thacker N, Jayson G. Breath-hold perfusion and permeability mapping of hepatic malignancies using magnetic resonance imaging and a first-pass leakage profile model. *NMR Biomed* 2002;15:164–173.
18. Tofts P. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. *J Magn Reson Imag* 1997;7:91–101.
19. Shi W, Siemann D. Preclinical studies of the novel vascular disrupting agent MN-029. *Anticancer Res* 2005;25:3899–3904.
20. Raynaud J, Pelletier J, Berthiaume M, et al. Quantitative magnetic resonance imaging evaluation of knee osteoarthritis progression over two years and correlation with clinical symptoms and radiologic changes. *Arthritis Rheum* 2004;50:476–487.
21. Ashton E, Takahashi C, Berg M, Goodman A, Totterman S, Ekholm S. Accuracy and reproducibility of manual and semiautomated quantification of MS lesions by MRI. *J Magn Reson Imag* 2003;17:300–308.
22. Takaya N, Cai J, Ferguson M, et al. Intra- and interreader reproducibility of magnetic resonance imaging for quantifying the lipid-rich necrotic core is improved with gadolinium contrast enhancement. *J Magn Reson Imaging* 2006;24:203–210.
23. Doria A, Castro C, Kiss M, et al. Inter- and intrareader variability in the interpretation of two radiographic classification systems for juvenile rheumatoid arthritis. *Pediatr Radiol* 2003;33: 673–681.
24. Lurie J, Tosteson A, Tosteson T, et al. Reliability of readings of magnetic resonance imaging features of lumbar spinal stenosis. *Spine* 2008;33:1605–1610.
25. Fu L, Fonov V, Pike B, Evans A, Collins D. Automated analysis of multi site MRI phantom data for the NIHDP project. *Lect Notes Comput Sci* 2006;4191:144–151.
26. Collins J. Imaging and other biomarkers in early clinical studies: ones step at a time or re-engineering drug development? *J Clin Oncol* 2005;23:5417–5419.
27. Jack C, Bernstein M, Fox N, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Magn Reson Imaging* 2008;27:685–691.
28. Eckstein F, Maschek S, Wirth W, et al. One year change of knee cartilage morphology in the first release of participants from the Osteoarthritis Initiative progression subcohort: association with sex, body mass index, symptoms and radiographic osteoarthritis status. *Ann Rheum Dis* 2009;68:674–679.
29. Hussain H, Chenevert T, Londy F, et al. Hepatic fat fraction: MR imaging for quantitative measurement and display—early experience. *Radiology* 2005;237:1048–1055.
30. Ma J. Dixon techniques for water and fat imaging. *J Magn Reson Imaging* 2008;28:543–558.
31. Reeder S, Pineda A, Wen Z, et al. Iterative decomposition of water and fat with echo asymmetry and least-squares estimation (IDEAL): application with fast spin-echo imaging. *Magn Reson Med* 2005;54:636–644.
32. Rakow-Penner R, Daniel B, Yu H, Sawyer-Glover A, Glover G. Relaxation times of breast tissue at 1.5T and 3T measured using IDEAL. *J Magn Reson Imaging* 2006;23:87–91.
33. O'Regan D, Callaghan M, Wylezinska-Arridge M, et al. Liver fat content and T2*: simultaneous measurement by using breath-hold multiecho MR imaging at 3.0T—feasibility. *Radiology* 2008; 247:550–557.
34. Bernard C, Liney G, Manton D, Turnbull L, Langton C. Comparison of fat quantification methods: a phantom study at 3.0T. *J Magn Reson Imaging* 2008;27:192–197.
35. Yu H, Shimakawa A, McKenzie C, Brodsky E, Brittain J, Reeder S. Multiecho Water-fat separation and simultaneous R2* estimation with multifrequency fat spectrum modeling. *Magn Reson Med* 2008;60:1122–1134.
36. Ashton E, Durkin E, Kwok E, Evelhoch J. Conversion from signal intensity to Gd concentration may be unnecessary for perfusion assessment of tumors using DCE-MRI. In: Proceedings of the 15th Annual Meeting of ISMRM, Berlin, 2007. (abstract 2813).
37. Chu B, Zhao X, Saam T, et al. Feasibility of in vivo, multicontrast-weighted MR imaging of carotid atherosclerosis for multicenter studies. *J Magn Reson Imaging* 2005;21:809–817.
38. Lieberman J, Tollefson G, Charles C, et al. Antipsychotic drug effects on brain morphology in first-episode psychosis. *Arch Gen Psychiatry* 2005;62:361–370.
39. Bakshi R, Benedict R, Bermel R, et al. T2 Hypointensity in the deep gray matter of patients with multiple sclerosis. *Arch Neurol* 2002;59:62–68.
40. Ashton E. Widespread functional and molecular imaging in drug development. *Expert Opin Drug Discov* 2007;2:1477–1485.
41. Ashton E. Quantitative medical imaging. *J Imaging Sci Technol* 2007;51:117–121.
42. Eckstein F, Westhoff J, Sittek H, et al. In vivo reproducibility of three-dimensional cartilage volume and thickness measurements with MR imaging. *AJR Am J Roentgenol* 1998;170:593–597.
43. Wluka A, Stuckey S, Snaddon J, Cicuttini F. The determinants of change in tibial cartilage volume in osteoarthritic knees. *Arthritis Rheum* 2002;46:2065–2072.
44. Qvist P, Bay-Jensen A, Christiansen C, Dam E, Pastoureaux P, Karsdal M. The disease modifying osteoarthritis drug (DMOAD): is it in the horizon? *Pharmacol Res* 2008;58:1–7.
45. Temenoff J, Mikos A. Review: tissue engineering for regeneration of articular cartilage. *Biomaterials* 2000;21:431–440.

46. Dardzinski B, Mosher T, Li S, Van Slyke M, Smithe M. Spatial variation of T2 in human articular cartilage. *Radiology* 1997;205:546-550.
47. David-Vaudey E, Ghosh S, Ries M, Majumdar S. T2 relaxation time measurements in osteoarthritis. *Magn Reson Imaging* 2004;22:673-682.
48. Bashir A, Gray M, Hartke J, Burstein D. Nondestructive imaging of human cartilage glycosaminoglycan concentration by MRI. *Magn Reson Med* 1999;41:857-865.
49. Tiderius J, Olsson L, Leander P, Ekberg O, Dahlberg L. Delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) in early knee osteoarthritis. *Magn Reson Med* 2003;49:488-492.
50. Maier C, Tan S, Hariharan H, Potter H. T2 quantitation of articular cartilage at 1.5 T. *J Magn Reson Imaging* 2003;17:358-364.
51. Riek J, Ashton E. Use of a dual-echo fast-spin-echo sequence for T2 mapping of cartilage within a clinical trial. In: Proceedings of the 17th Annual Meeting of ISMRM, Honolulu, 2009. (abstract 3988).
52. Trattinig S, Marlovits S, Gebetsroither S, et al. Three-dimensional delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) for in vivo evaluation of reparative cartilage after matrix-associated autologous chondrocyte transplantation at 3.0T: preliminary results. *J Magn Reson Imaging* 2007;26:974-982.
53. Kimelman T, Vu A, Storey P, McKenzie C, Burstein D, Prasad P. Three-dimensional t1 mapping for dGEMRIC at 3.0T using the look locker method. *Invest Radiol* 2006;41:198-203.
54. Evelhoch J. Key factors in the acquisition of contrast kinetic data for oncology. *J Magn Reson Imaging* 1999;10:254-259.
55. Murase K. Efficient method for calculating kinetic parameters using T1-weighted dynamic contrast-enhanced magnetic resonance imaging. *Magn Reson Med* 2004;51:858-862.
56. Landau H. Sampling, data transmission, and the Nyquist rate. *Proc IEEE* 1967;55:1701-1706.
57. Roberts H, Roberts T, Lee T, Dillon W. Dynamic, contrast-enhanced CT in human brain tumors: quantitative assessment of blood volume, blood flow, and microvascular permeability. *AJNR Am J Neuroradiol* 2002;23:828-832.
58. Jayson G, Parker G, Mullamitha S, et al. Blockade of platelet-derived growth factor receptor-beta by CDP860, a humanized, PEGylated di-Fab', leads to fluid accumulation and is associated with increased tumor vascularized volume. *J Clin Oncol* 2005;23:973-981.
59. Weinmann H, Laniado M, Mutzel W. Pharmacokinetics of Gd-DTPA/dimeglumine after intravenous injection into healthy volunteers. *Physiol Chem Phys Med NMR* 1984;16:167-172.
60. Padhani A, Hayes C, Landau S, Leach M. Reproducibility of quantitative dynamic MRI of normal human tissues. *NMR Biomed* 2002;15:143-153.
61. Cenic A, Nabavi D, Craen R, Gelb A, Lee T. A dynamic CT method to measure cerebral blood flow in brain tumors: validation and application to CBF maps. *AJNR Am J Neuroradiol* 2000;21:462-470.