Journal of Imaging Science and Technology® 51(2): 117–121, 2007.

Quantitative Medical Imaging

Edward A. Ashton
VirtualScopics, Inc., 350 Linden Oaks, Rochester, New York 14625
E-mail: ashton@virtualscopics.com

Abstract. *The field of medical imaging is undergoing a radical shift, from subjective interpretation to quantitative analysis and measurement. This transformation is well established in the clinical trials arena and is beginning to enter the diagnostic field as well. This article will consider the implications of this change in terms of instrumentation, procedure, and analysis. It will include a brief review of current imaging practices and standards, focusing primarily on the clinical trials arena, and will examine in detail the acquisition and analysis techniques that will be required to successfully complete the transition from qualitative to quantitative imaging.* © *2007 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.(2007)51:2(117)]

INTRODUCTION

In the past three decades, medical science has made great strides toward the understanding of cellular biochemistry and the mechanisms of disease. In concert with this change, the technologies available for medical imaging have proliferated. In particular, cross-sectional imaging techniques, which allow the precise reconstruction of three dimensional structures within the body, and functional imaging techniques, which allow the assessment of biological function as well as form, have become widely available. These imaging techniques can potentially provide a tremendous amount of information about disease state and response to treatment. However, the means to interpret that information has badly lagged behind the ability to acquire it.

A good example of this phenomenon is given by the response evaluation criteria in solid tumors (RECIST) standard,[1] which is the primary imaging endpoint for assessing disease progression or response to treatment in many types of cancer. This technique reduces the assessment of structural changes in tumors to a simple summation of longest diameters, limited to the axial imaging plane. The bidimensional World Health Organization standard from which RECIST is derived was originally developed with plain film x-ray imaging in mind and fails to take advantage of the far richer three dimensional information set available in a spiral computed tomography (CT) scan. The essence of the RECIST evaluation is the classification of cases into one of four categories: complete response—those whose tumors have disappeared completely, partial response—those whose tumors have shrunk by at least 30%, progressive disease—those whose tumors have grown by at least 20%, and stable disease—all cases that do not fit any other category. It is instructive to realize that the 30% reduction and 20% increase cutoff values between categories are derived not from any study of what changes in tumor size might be biologically meaningful, but rather from the differences in the sizes of balls placed under a foam mattress that could be accurately detected by manual palpation. It is unsurprising in light of this that the results of this classification are frequently poor correlates to harder endpoints such as patient survival.[2] Moreover, in some cases, the assessment gleaned from RECIST will correlate well with the changes seen in more sensitive measurements such as tumor volume or radio density. In others, however, it will not.[3] An example of this is given in Fig. 1.

The continued use of RECIST as a standard evaluation in cancer is not due to any strong argument that single diameters are better than, or even equivalent to, a full volumetric structural or functional assessment. Rather, it is due to the lack of availability of convenient and reliable tools for producing these sorts of measurements. Several groups, including VirtualScopics, are now in the process of producing, validating, and commercializing tools in this and other areas which will allow the use of precise, quantitative measurement in medical imaging, first in the clinical trials arena, and later in the clinic itself.

QUANTITATIVE MEDICAL IMAGING IN CLINICAL TRIALS

Recent scientific advances have brought about massive changes in the business of drug discovery and development. Large companies that once had relatively few compounds in development now juggle hundreds of potential candidate

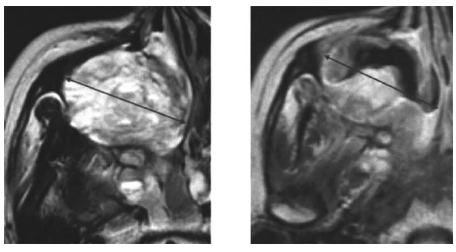

Figure 1. (Left) Baseline scan showing a large tumor filling much of the subject's right sinus cavity. Standard RECIST measurement is shown by black arrows. (Right) Follow-up scan showing the same tumor, which has cavitated, losing approximately 70% of its bulk. However, the standard RECIST measure is roughly unchanged.

Received Aug. 9, 2006; accepted for publication Sep. 22, 2006.
1062-3701/2007/51(2)/117/5/$20.00.

compounds. For the industry to sustain itself, companies must devise ways to quickly identify promising compounds and cull ineffective ones. Various estimates place the cost of developing a single compound from discovery to the pharmacy shelf at between $800 million and $1.7 billion. The Tufts Center for the Study of Drug Development in its Impact Report, Volume 4, Number 5, September/October 2002 estimated that it costs $808 million, on average, to develop and win market approval for a new drug in the United States. This study stated that better preclinical screens, to increase success rates from the current 21.5% to one in three, could reduce capitalized total cost per approved drug by $242 million.

Quantitative imaging can reduce late-stage attrition dramatically by offering more accurate information about both safety and efficacy much earlier in the drug development process. As an example, dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) is able to provide information about blood flow and vascular permeability in tumors. This information allows a relatively small trial to quickly determine whether an antiangiogenic or vascular disruptive agent is effectively reducing blood flow or vascular permeability within targeted tumors.[4,5] Whereas phase I clinical trials normally focus on dosage and safety, studies that incorporate quantitative imaging can also test for drug efficacy, offering information that can save millions of dollars by allowing companies to better prioritize their drug pipelines and make go/no-go decisions earlier in the development process. In preclinical research, scientists can test method of action and lay the foundation for a more streamlined clinical trial process from phase I through phase III by obtaining critical information about efficacy.

The value of medical image analysis stems from quantification and automation. While the primary shortcoming of standard endpoints, such as pain or functionality scoring, is that they are largely subjective and difficult to reproduce, quantitative imaging allows the replacement of a subjective evaluation like pain scoring with an objective quantification such as cartilage volume or thickness. Automation in the image analysis process—using a computer algorithm to measure lesion size rather than a clinician with a pair of calipers—provides a degree of accuracy and reproducibility that cannot be duplicated by manual techniques. A good example of this phenomenon is provided by the measurement, using MRI, of lesion burden in multiple sclerosis (MS) patients. MS lesions generally are irregularly shaped, and tend to have fuzzy, indistinct boundaries (see Fig. 2). As a result, several studies have estimated the interoperator coefficient of variability (CV)—the standard deviation of repeated measurements divided by their mean—in white matter lesion burden measurement at 20% or more[6,7] and the intraoperator CV at ~7%. Introducing automation into this process can reduce this measurement variability to 2% or less.[6] Because the number of subjects required for a study to achieve statistical significance is directly related to measurement variability, this allows efficacy findings to be obtained in smaller trials earlier in the development process.

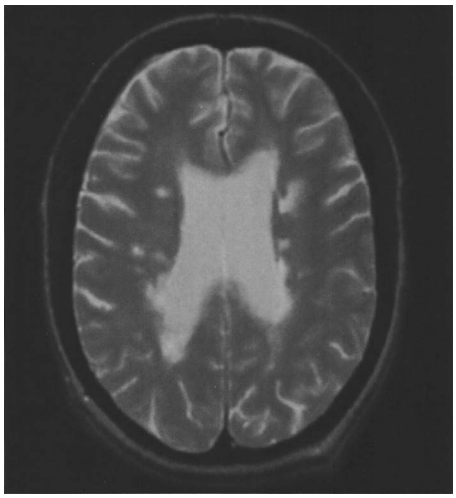

Figure 2. *T2* weighted MRI scan of the brain of a multiple sclerosis patient. The irregular bright areas surrounding the ventricles are white matter lesions. The small size and indistinct boundaries of these lesions make them very difficult to quantify manually.

IMPORTANCE OF PRECISION

An important corollary to the previous argument is that quantitative medical image analysis is valuable only insofar as it is done well. Quality in this case is defined primarily in terms of scan-rescan CV—the coefficient of variability measured for repeated imaging sessions in the absence of true biological change. If the results of a particular test are highly variable even in the absence of any biological change, that test has little value in detecting either disease progression or response to treatment. A good example of this is given by DCE-MRI. This technique, while extremely valuable in concept for determining the effectiveness of compounds designed to reduce blood flow or vascular permeability in tumors, has until recently been severely limited by poor reproducibility even in untreated subjects over very short time frames, with scan-rescan CVs ranging from 18% to 25%.[8]

There are several noise sources that contribute to this measurement variability. Because DCE-MRI involves repeated imaging of the same volume over the course of several minutes, subject motion during data acquisition is a serious issue. Typical solutions to this problem, such as respiratory gating or navigator pulses, are ruled out due to the need for acquisition speed—the requirements of the vascular modeling typically make it necessary to acquire a full volume every 8 s or less. Subject motion can most reasonably be addressed through the use of a robust image coregistration algorithm prior to analysis and modeling, in combination with an acquisition protocol that minimizes out-of-plane movement. For chest and abdominal imaging, for example, coronal plane imaging is recommended, because respiratory motion in the abdomen and chest can largely be confined to the coronal plane.

A second major noise source in DCE-MRI analysis is inaccuracies in the measurement of the arterial input function (AIF), which is the tracer time-concentration curve in arterial plasma. This function serves as the input to the lin-

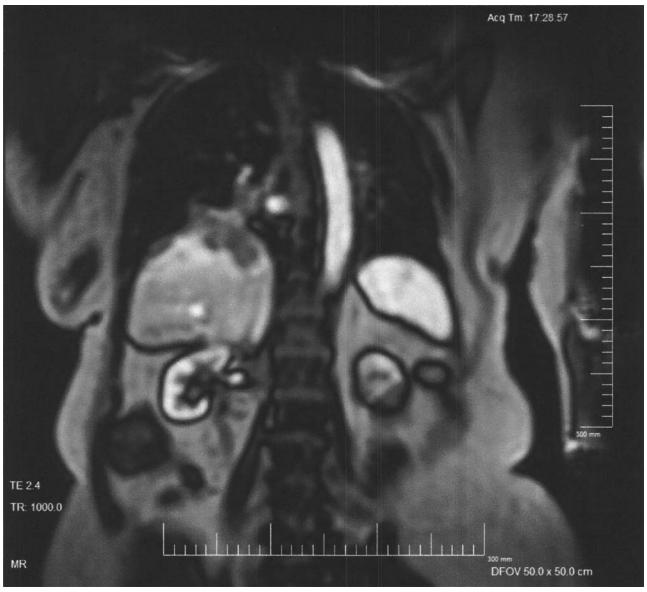

Figure 3. One image from a coronal DCE-MRI scan. Note the high enhancement in the aorta, kidney, and liver, and the well defined tumors in the liver.

ear system model defining the vascular bed, so accuracy in its definition is an absolute necessity. This is a difficult problem, because the continuous motion of the blood introduces significant artifacts into the acquired images, making it very difficult to accurately assess the concentration of tracer within a large artery such as the aorta. Most work in the past[8,9] has assumed that a model AIF rather than an image-derived one is adequate for vascular parameter calculation. However, several recent studies[5,10,11] have shown that using an accurate image-derived AIF can greatly reduce measurement noise.

A third major source of measurement noise in DCE-MRI analysis is a lack of precision in the identification of the regions of interest (typically tumor boundaries) which are to be analyzed. Hand drawing of these boundaries can lead to significant variability both among different analysts and across time points, which can be greatly reduced through the introduction of automated boundary finding techniques.[6,12]

VirtualScopics has designed and deployed a DCE-MRI analysis system incorporating robust time point co-registration, automated data-derived AIF calculation, and semiautomated tumor boundary identification. In conjunction with MediciNova (San Diego, CA), we have recently tested the scan-rescan variability of this system using data from a phase I human clinical trial. This experiment involved ten subjects, each of whom was imaged twice, with minimum and maximum separation between scans of 24 h and 20 days, respectively. These subjects did receive study drug, but at dose levels low enough that no measurable biological effects were expected. A five slice, 5 cm slab was imaged in each case, with spatial resolution of 2 mm in-plane and 10 mm between images. Images were acquired in the coronal plane using a semikeyhole technique, with TE/TR/TI/FA of 2.42/1000/340/16 and temporal resolution of approximately 3 s per slab. A sample image from this experiment is shown in Fig. 3.

The primary parameter of interest in this experiment was K^{Trans}, defined as the volume transfer constant between arterial plasma and extracellular extravascular space. This parameter is highly correlated with both blood flow and vascular permeability.[13] Analysis of these subjects' data yielded a scan-rescan CV of 6.8%. Moreover, the observed change from baseline to follow-up was correlated with the amount of time elapsed between the two scans ($R^2=0.45$), indicating that some of the observed variability was likely due to true biological change—in particular, to an increase in vascular perfusion within the tumors over time. This phenomenon was particularly apparent in the three cases with separation between base line and follow-up of greater than 7 days. Removing these three cases from the analysis yielded a coefficient of variability for the remaining 7 cases of 3.6%. Measurement with this level of precision enables not only the detection of very large changes in vascular perfusion, such as those induced by effective compounds at maximum tolerated dose, but also the differences among small cohorts at varying dose levels. This makes it possible to establish dose-response relationships early in a development program, and potentially enables the detection of a biologically optimal dose at a different level than the MTD through the precise definition of the compound's dose-response curve.

Precise, automated measurement brings another critical benefit: it enables the detection of small changes in structure and function over shorter periods of time than would otherwise be possible. In the evaluation of osteoarthritis, for example, MRI of the cartilage in the knee coupled with automated measurement of volume and chemical composition shows disease changes in months; these changes would not be apparent using standard x-ray evaluation for years. With this quality of information, researchers can more confidently make the go/no-go decision for a compound early in the evaluation process, allowing scarce resources to be allocated to the most promising candidates.

Reproducible medical image analysis is driven by algorithms that enable quantitative, volumetric measurement of structures and metabolic functions. Guided by the information present in the images, as well as embedded anatomical knowledge, the algorithms enable segmentation of different tissue types such as bone, muscle, fat, and fluid. From an MRI knee scan, for instance, it is possible to produce a three-dimensional reconstruction that graphically distinguishes cartilage from underlying bone, as well as from ligaments, fluid, degenerated menisci, or inflamed synovium (see Fig. 4). This capability provides a valuable assessment tool for clinical research in osteoarthritis—a disease with multiple endpoints—because it allows the very sensitive and specific measurement of all the components of the knee joint and the detection of small changes in any of those components over time.

In addition to structural measurements such as size, thickness, or shape, properly utilized medical imaging can allow the assessment of functional parameters. Functional imaging encompasses a wide variety of imaging techniques, including functional MRI (fMRI), DCE-MRI, dynamic

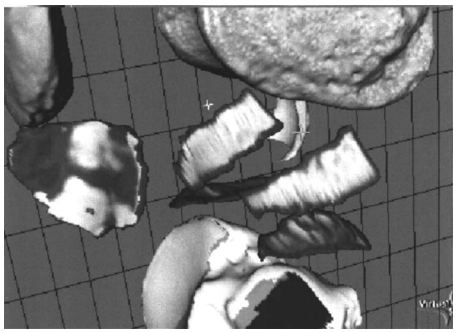

Figure 4. Three dimensional rendering of the knee joint of an osteoarthritis subject, showing tibial, femoral, and patellar cartilage components.

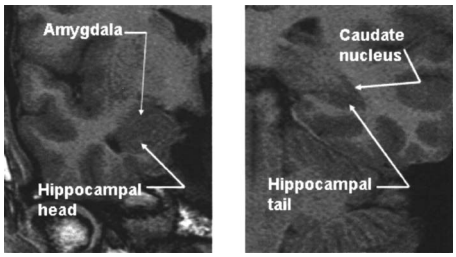

Figure 5. Hippocampus, as seen in *T1*-weighted MRI brain scans. (Left) The boundary between the hippocampal head and the amygdala. (Right) The boundary between the hippocampal tail and the caudate nucleus.

contrast-enhanced computed tomography (DCE-CT), and positron emission tomography (PET). These methods allow the assessment of the metabolic activity of an organ or lesion through the measurement of markers such as tissue blood volume, blood flow, oxygen utilization, or glucose metabolism.

In the clinic, functional imaging makes it possible to distinguish between scar tissue and viable tumor, and in some cases between benign lesions and malignant ones. In drug development, functional imaging allows the direct measurement of drug effects that otherwise would only be observable indirectly, through their influence on patient survival or well being. As discussed previously, several functional imaging modalities, including DCE-MRI, DCE-CT, and dynamic PET, allow the direct measurement of parameters such as blood flow, blood volume, and vessel permeability. However, functional imaging imposes an additional burden on the radiologist. From the previous section, it is clear that obtaining a measurement of tissue blood flow from a DCE-MRI scan requires fairly sophisticated modeling software and currently there is no standardized commercial package available to accomplish this. Acquisition techniques are also not yet standardized, with the result that scientists and clinicians wishing to make use of DCE-MRI are sometimes given conflicting advice by various experts.

These problems are less pronounced in PET imaging, simply because the data acquisition and modeling are less complex. However, there are questions that must be answered before designing a clinical study using this modality as well. Should blood glucose levels be accounted for in the modeling and analysis process? Should activity measurements be normalized by subject weight, lean body mass, or body surface area? What size of lesion can be reliably modeled? The answers to these questions will be dependent on the structure and goals of each study, and failure to adequately address them may lead to ambiguous or erroneous conclusions.

APPLICATIONS IN NEUROLOGY AND CARDIOLOGY

Automated medical image analysis is particularly useful in evaluating diseases of the brain. Because the brain has no moving parts and has a fairly consistent structure from person to person, it is possible to generate a generalized map, or anatomic atlas, of the location and shape of many of its important structures.[14] Under the supervision of a neuroradiologist, this map can then be applied to a series of patient scans to provide a consistent measurement of neural structures that frequently have unclear or even arbitrary boundaries.

A good example of an important but difficult to measure neural structure is the hippocampus, a gray matter structure of the brain that is involved in a number of functions, including the formation of long-term memory. Changes in the hippocampus are implicated in a number of diseases, including intractable temporal lobe epilepsy and Alzheimer's disease. The hippocampus is adjacent to and difficult to separate from other gray matter structures, including the amygdala and the caudate nucleus (see Fig. 5). Manual measurements of the hippocampus are difficult to reproduce because there is no clearly visible boundary with these structures. An automated measurement technique using an anatomical atlas might not always agree with any particular radiologist. However, experiments have shown that any given radiologist is unlikely to precisely agree even with himself if asked to measure the same scan a week or two later.[6,15] When tracking changes over time the key factor is reproducibility and in that area automated methods provide a significant advantage.

Medical image analytics also have shown great promise in cardiology and angiography. It was long thought that the key danger sign in the assessment of arterial disease was vascular occlusion—the narrowing or blockage of the coronary or carotid arteries. However, it is now known that large vascular plaques can form in key arteries without narrowing the lumen at all, by pushing the outer wall of the vessel outward rather than pushing the inner wall inward. Such plaques can be either relatively harmless or deadly, depending on what is inside them and how likely they are to break open, spilling their contents into the blood stream. The most dangerous plaques have thin walls and large liquid cores filled with lipids and other substances. When these plaques burst, the contents quickly form clots, which can then lodge in the brain, heart or lungs.

With the proper acquisition parameters, MRI can allow the identification and measurement of vascular plaques. More importantly, it also can allow a determination of plaque composition. This makes it possible for surgeons to distinguish between a relatively benign plaque that can be left for future observation, and a potentially deadly one that requires immediate surgery.

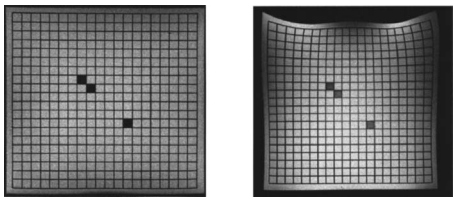

Figure 6. (Left) MRI scan of a volume and linearity phantom obtained using a well-maintained system. (Right) MRI scan of the same phantom obtained using a clinical scanner that was in routine use at the site. This scanner eventually required radio frequency coil replacement.

PRACTICAL CONSIDERATIONS

Although the potential benefits of the techniques discussed here are many and varied, it is important to bear in mind that there are several obstacles that must be overcome before these methods can be fully exploited in either the clinical trial or diagnostic arenas. Primary among these is that the personnel and systems presently deployed at radiology sites have been trained and designed around qualitative rather than quantitative imaging. This means that many of the imaging techniques necessary for quantitative imaging are not routine for these sites, and must be learned.

Equally important, the equipment maintenance standards currently in place at most imaging sites are not sufficient to support precise quantitative measurement. Figure 6 shows two MRI scans of the same linearity and volume phantom. This phantom is a device, roughly the size of a small suitcase, which contains a regular grid of known geometry. A scan of this phantom can be compared to the expected results in order to highlight problems with the field homogeneity, coil function, etc., of the scanner under examination. The scan on the right was obtained from a clinical scanner that was in routine use at the imaging center at the time of acquisition.

The distortion which is apparent in the phantom scan was present in all clinical scans obtained using this machine, and had been for some time prior to this quality check. It seems fairly obvious to ask why this problem was not noted by the site technicians or radiologists. The answer is simple: This sort of distortion can be backed out by the human visual system as long as it is consistent over time, and is therefore not a major impediment to qualitative interpretation of the images generated by the system. The requirements for quantitative imaging are fundamentally different, and in fact significantly more stringent. Failure to take this into account can result in measurements that are meaningless or misleading.

CONCLUSIONS

Quantitative medical imaging promises immense benefits for both diagnostic and clinical trial applications. In the clinical trials arena, quantitative imaging allows precise direct measurement of the biological activity of a compound. This replaces inference of drug effect through observation of secondary effects such as changes in reported pain. Because the error bar placed on quantitative measurement is much smaller than that placed on more subjective measures, statistically significant results relating to compound efficacy can be obtained in smaller, earlier phase clinical trials. This allows the decision point on further compound development to be pushed back from late phase II or even phase III to phase I trials, where failure is far less costly. In the diagnostic arena, precise quantitative measures may help facilitate the personalization of medicine, by assessing quickly and accurately whether a particular subject is responding favorably to a given targeted compound. A great deal of development and validation work is still required to bring all of these potentialities to full fruition, but the next 10 years should see many of these techniques making their way into routine clinical use.

REFERENCES

[1] P. Therasse, S. Arbuck, E. Eisenhauer, J. Wanders, R. Kaplan, L. Rubinstein, J. Verweij, M. Van Glabbeke, A. van Oosterom, M. Christian, and S. Gwyther, "New guidelines to evaluate the response to treatment in solid tumors", J. Natl. Cancer Inst. **92**, 205 (2000).

[2] R. Twombly, "Criticism of tumor response criteria raises trial design questions", J. Natl. Cancer Inst. **98**, 232 (2006).

[3] R. S. Benjamin, H. Choi, H. Macapinlac, M. Burgess, S. Patel, L. Chen, D. Podoloff, and C. Chuslip, "Response of gastrointestinal stromal tumors (GISTs) to imatinib by Choi criteria and response evaluation criteria in solid tumors (RECIST) as surrogates for survival and time to progression", 2006 ASCO Annual Meeting Proceedings Part I, 9506 (2006).

[4] G. Liu, H. Rugo, G. Wilding, T. McShane, J. Evelhoch, C. Ng, E. Jackson, F. Kelcz, B. Yeh, F. Lee, C. Charnsangavej, J. Park, E. Ashton, H. Steinfeldt, Y. Pithavala, S. Reich, and R. Herbst, "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. **23**, 5464 (2005).

[5] E. Ashton, T. McShane, and J. Evelhoch, "Inter-operator variability in perfusion assessment of tumors in MRI using automated AIF detection", LNCS **3749**, 451 (2005).

[6] E. Ashton, C. Takahashi, M. Berg, A. Goodman, S. Totterman, and S. Ekholm, "Accuracy and reproducibility of manual and semiautomated quantification of MS lesions by MRI", J. Magn. Reson. Imaging **17**, 300 (2003).

[7] K. Van Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens, "Automated segmentation of multiple sclerosis lesions by model outlier detection", IEEE Trans. Med. Imaging **20**, 677 (2001).

[8] S. Galbraith, M. Lodge, N. Taylor, G. Rustin, S. Bentzen, J. Stirling, and A. Padhani, "Reproducibility of dynamic contrast-enhanced MRI in human muscle and tumours: Comparison of quantitative and semi-quantitative analysis", NMR Biomed. **15**, 132 (2002).

[9] A. Padhani, C. Hayes, S. Landau, and M. Leach, "Reproducibility of quantitative dynamic MRI of normal human tissues", NMR Biomed. **15**, 143 (2002).

[10] M. Rijpkema, J. Kaanders, F. Joosten, A. van der Kogel, and A. Heerschap, "Method for quantitative mapping of dynamic MRI contrast agent enhancement in human tumors", J. Magn. Reson. Imaging **14**, 457 (2001).

[11] K. Murase, K. Kikuchi, H. Miki, T. Shimuzu, and J. Ikezoe, "Determination of arterial input function using fuzzy clustering for quantification of cerebral blood flow with dynamic susceptibility contrast-enhanced MR imaging", J. Magn. Reson. Imaging **13**, 797 (2001).

[12] E. Ashton, L. Molinelli, S. Totterman, and K. Parker, "Evaluation of reproducibility for manual and semi-automated feature extraction in CT and MR images", *Proc. IEEE ICIP* (IEEE, Piscataway, NJ, 2002) Vol. 3, p. 161.

[13] P. Tofts, "Modeling tracer kinetics in dynamic Gd-DTPA MR imaging", J. Magn. Reson Imaging **7**, 91 (1997).

[14] E. Ashton, J. Riek, L. Molinelli, M. Berg, and K. Parker, "A method for fully automated measurement of neurological structures in MRI", *Proc. SPIE* **5032**, 1125 (2003).

[15] E. Ashton, K. Parker, M. Berg, and C. Chen, "A novel volumetric feature extraction technique, with applications to MR images", J. Magn. Reson. Imaging **16**, 365 (1997).