A method to estimate sample sizes for DCE-MRI-based studies of heterogeneous tumors

C. J. Rose[1,2], J. P. O'Connor[1,2], C. Roberts[1,2], G. A. Buonaccorsi[1,2], Y. Watson[1,2], S. Cheung[1,2], G. Jayson[3], and G. J. Parker[1,2]
[1]Imaging Science and Biomedical Engineering, The University of Manchester, Manchester, United Kingdom, [2]The University of Manchester Biomedical Imaging Institute, Manchester, United Kingdom, [3]Cancer Research UK and University of Manchester Department of Medical Oncology, The Christie Hospital, Manchester, United Kingdom

INTRODUCTION This abstract presents a method for estimating the number of patients that should be recruited to dynamic contrast-enhanced (DCE) MRI-based clinical trials of novel cancer therapies; the main contribution we make is to consider the role that tumor heterogeneity plays. It is increasingly recognized that DCE-MRI tracer kinetic modeling[1] is of great utility in the study of the tumor microenvironment, particularly in the context of early phase clinical trials of novel therapies[2]. Until recently it has been common practice to treat tumors as homogenous—either by fitting a tracer kinetic model to an averaged time series (i.e., the time series formed by averaging the contrast agent concentrations at each dynamic time point over every tumor voxel), or by fitting a tracer kinetic model at each voxel, but then averaging the voxel-wise parameter values over the entire tumor. While these methods have proven useful, maps of tracer kinetic model parameters typically exhibit clear spatial structure in the tumor microenvironment (Fig. 1), and there is growing evidence that by considering tumor heterogeneity, better analyses can be performed[3,4]. Here, we focus on the planning of imaging-based trials and present a method to estimate the number of participants that should be recruited, by considering statistical power and tumor heterogeneity. Our work is important because only 14% of the DCE-MRI-based trials reviewed in reference 5 reported sample size considerations and none considered tumor heterogeneity.

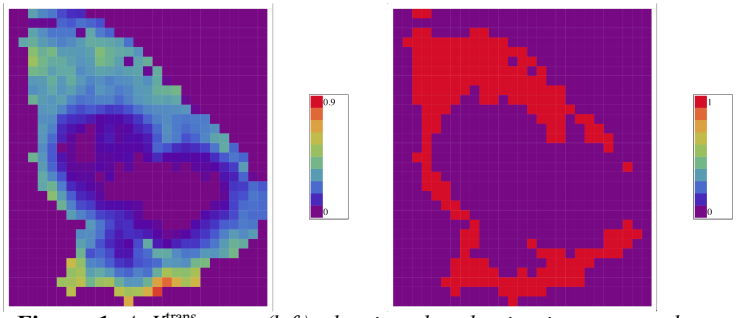

Figure 1 *A K^{trans} map (left) showing the classic rim, core and non-enhancing core partitioning, and the corresponding clustering result (right), showing the voxels identified as rim.*

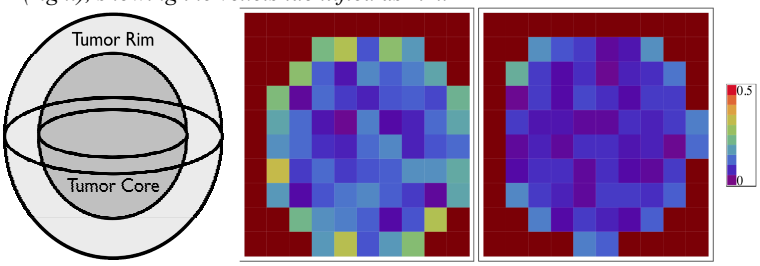

Figure 2 *The model of tumor heterogeneity (left), and a simulated pre- and post-treatment parameter map pair (right).*

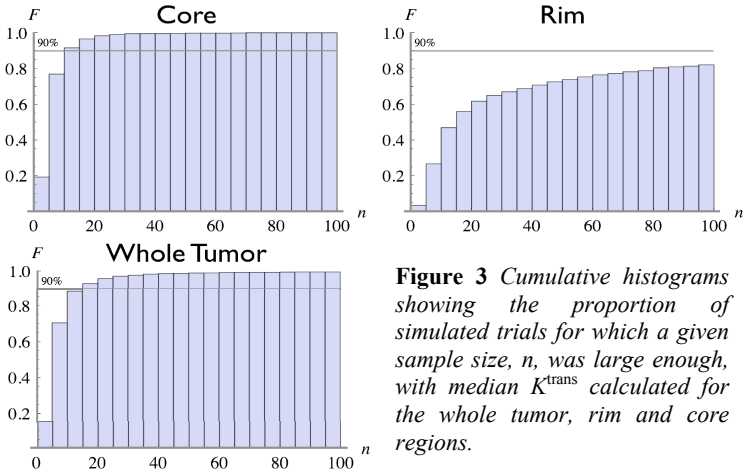

Figure 3 *Cumulative histograms showing the proportion of simulated trials for which a given sample size, n, was large enough, with median K^{trans} calculated for the whole tumor, rim and core regions.*

METHOD The key insight that we build upon is the recognition that anti-vascular and vascular-disrupting agents tend to preferentially affect certain regions of tumors[3]. The *t*-test is commonly used in DCE-MRI-based studies of cancer therapies[5] to test for post-treatment changes in summaries of K^{trans}. These summaries are often the mean or median value of K^{trans}, with averaging performed over the tumor; however, the observed effect will be attenuated by tumor regions that are unaffected by the treatment (such as necrotic regions[3]). Such averaging will disguise genuine therapeutic action and may, in the worst case, result in type II errors. The general idea is to use imaging-derived knowledge of drug effect and heterogeneity from one phase of testing (e.g., phase I) to inform a subsequent phase (e.g., phase II): by simulating a large number of imaging-based trials using a tumor model that understands how heterogeneity changes after treatment, a distribution of required sample sizes for various tumor partitions (described below) can be computed. Compared to the conventional approach of computing a point estimate of the required sample size, computing distributions allows decisions about the tradeoff between trial cost and power to be made.

We model tumor heterogeneity in terms of a spherical core and a spherical shell-shaped rim (Fig. 2), which differ in K^{trans} values. Tumors are usually aspherical, but shape is much less important than the proportion of voxels in each partition. We estimate these proportions using *k*-means clustering[6], with k=2 to identify rim and core partitions (Fig. 1). We account for a possible non-enhancing region of the core (Fig. 1), by assuming it to be non-tumor.

Let us assume the set of tumors $\mathbf{T}=\{T_1,\ldots,T_N\}$. We denote the (real) pre- and post-treatment K^{trans} maps of tumor T_i as K_i^{Pre} and K_i^{Post}, respectively, where $T_i \in \mathbf{T}$. To simulate a pre-treatment K^{trans} map, a tumor $T_j \in \mathbf{T}$, is randomly chosen and each voxel of an appropriately-sized 3D array is assigned as being rim, core or non-tumor according to the model of tumor heterogeneity and the proportions determined by the clustering result (Fig. 2). A resampling method is then used to populate those rim and core voxels with K^{trans} values taken from the rim and core partitions in K_j^{Pre}. The result is a simulated pre-treatment K^{trans} map (Fig. 2). The corresponding post-treatment map is simulated similarly, except we use K_j^{Post} as the prototype map, modeling *intra*-tumor changes in heterogeneity, rather than *inter*-tumor changes. Repeating this process several times allows us to simulate an entire trial; we choose the prototypes K_j^{Pre} and K_j^{Post} using resampling to model the variation likely to be observed in the population.

We used data from 26 liver metastases from 10 colorectal cancer patients who underwent treatment with bevacizumab. DCE-MR imaging was performed before and after treatment using a 1.5 T Philips Intera scanner and pre- and post-treatment K^{trans} maps were computed; the reader is referred to reference 7 for full details of the study protocol. We simulated 1000 trials, each with 10 tumors. For each partition (rim, core, & whole tumour) and each trial, the observed effect size was computed and used to calculate the number of tumors, n, that would be required[8] to detect that effect at the α=0.05 significance level with statistical power 1-β=0.8. The distributions of these required sample sizes were then visualized using cumulative histograms. The simulations were performed using Mathematica version 7.

RESULTS Fig. 3 shows the sample size distributions. For this trial, distributions for the core and whole tumor partitions were similar, but computing median K^{trans} over the core results in more powerful tests: with $n \leq 15$ patients, there is a 92% probability of detecting drug effect, but up to 20 patients are required to detect drug effect with probability > 90% if median K^{trans} is computed for the whole tumor (Fig. 3). If averaging were performed over the rim, more than 100 patients would be required.

CONCLUSIONS We have developed a method to estimate sample size requirements for imaging-based trials that accounts for tumor heterogeneity. Rather than provide point estimates of sample size requirements, the method estimates distributions to allow decisions about the tradeoff between trial cost and power to be made.

REFERENCES 1 Tofts. JMRI-J Magn Reson Im 1997;7(1):91-101. **2** Leach et al. Br J Cancer 2005;92(10):1599-1610. **3** Berry et al. Magn Reson Med 2008;60(1):64–72. **4** Rose et al. Magn Res Med 2009;62(2);488–499. **5** Jackson et al. Clin Res Cancer 2007;13(12):3449–3459. **6** Hastie. Elements of statistical learning, 2nd Ed., 2009. **7** O'Connor et al. Clin Cancer Res 2009;15(21):6674–6682. **8** Machin. Sample size tables for clinical studies, 2008.