

# A Bayesian Hierarchical Model for the Analysis of a Longitudinal Dynamic Contrast-Enhanced MRI Oncology Study

Volker J. Schmid<sup>†</sup>      Brandon Whitcher\*      Anwar R. Padhani<sup>‡</sup>  
N. Jane Taylor<sup>‡</sup>      Guang-Zhong Yang<sup>†</sup>

\* Corresponding Author:

Clinical Imaging Centre  
GlaxoSmithKline  
Hammersmith Hospital  
Imperial College London  
Du Cane Road  
London W12 0NN  
United Kingdom  
+44 20 8008 6052; +44 20 8008 6491 (fax)  
*brandon.j.whitcher@gsk.com*

<sup>†</sup> Institute of Biomedical Engineering  
Imperial College London  
London SW7 2AZ  
United Kingdom

<sup>‡</sup> Paul Strickland Scanner Centre  
Mt. Vernon Cancer Centre  
Northwood HA6 2RN  
United Kingdom

*Key words.*

*Running head.* Bayesian Hierarchical Model for DCE-MRI in Oncology

*Approximate word count.* 5000

## Abstract

Imaging in clinical oncology trials provides a wealth of information that contributes to the drug development process, especially in early phase studies. This paper focuses on kinetic modeling in DCE-MRI, inspired by mixed-effects models that are frequently used in the analysis of clinical trials. Instead of summarizing each scanning session as a single kinetic parameter – such as median  $K^{\text{trans}}$  across all voxels in the tumor ROI – we propose to analyze all voxel time courses from all scans and across all subjects simultaneously in a single model. The kinetic parameters from the usual non-linear regression model are decomposed into unique components associated with factors from the longitudinal study; e.g., treatment, patient and voxel effects. A Bayesian hierarchical model provides the framework in order to construct a data model, a parameter model, as well as prior distributions. The posterior distribution of the kinetic parameters is estimated using Markov chain Monte Carlo (MCMC) methods. Hypothesis testing at the study level for an overall treatment effect is straightforward and the patient- and voxel-level parameters capture random effects that provide additional information at various levels of resolution to allow a thorough evaluation of the clinical trial. The proposed method is validated with a breast cancer study, where the subjects were imaged before and after two cycles of chemotherapy, demonstrating the clinical potential of this method to longitudinal oncology studies.

# 1 Introduction

Assessing the efficacy of cancer treatments using *in vivo* imaging is shifting from qualitative techniques to quantitative imaging methods that characterize biologically relevant properties of tumor tissue. The use of model-free or heuristic measures, such as the initial area under the Gadolinium curve (IAUGC), or fully quantitative measures, such as the kinetic parameters from a compartmental model, are relatively well understood in the analysis of dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) (1; 2; 3). Analysis of an oncology imaging trial is usually achieved by applying statistical summaries, such as the mean or median, to the parameters of interest derived from tissue regions of interest (ROIs). That is, enhancing (tumor) voxels are identified from the DCE-MRI data for each scan across all subjects and those voxels are represented by a single parameter; e.g.,  $K^{\text{trans}}$  from quantitative analysis and  $\text{IAUGC}_{90}$  from a heuristic analysis. Hypothesis testing, either parametric or non-parametric, may then be applied to the derived statistics in order to assess the effects of treatment.

Applying statistical summaries to the kinetic parameter maps from DCE-MRI however discards a substantial amount of information contained in the contrast agent concentration time curves (CTCs) at each voxel, essentially abstracting thousands of observations in space and time to a single number per scan per subject. We believe that there is a wealth of potential information by retaining the collection of CTCs across all subjects and scans, at the same time acknowledging the fact that not all CTCs are the same and not all patients are the same.

This paper proposes a Bayesian hierarchical model to analyze all tumor CTCs across all patients and scans in a given study simultaneously based on the concept of a mixed-effects model. Mixed-effects models are well established in the statistical community and have found widespread applications in, for example, agriculture, economics, geophysics and the analysis of clinical trials (4; 5). Previous examples of mixed-effects models in neuroimaging primarily exist for functional MRI studies (6; 7; 8). Mixed-effects models extend the concept of traditional linear or non-linear models by combining both fixed effects and random effects in the same model. More generally, mixed-effects models are most often used to describe relationships between the measured response and explanatory variables in data that are grouped according to one or more factors. Fixed effects denote parameters that are associated with an entire population and random effects denote parameters which are associated with random samples from a population. For example, the drug or radiation therapy given in a trial is a fixed effect, whereas patients are inherently random because they are sampled from the general population. By acknowledging the fact that some parameters are associated with random samples from a population, the results from a mixed-effects model may be generalized beyond the collection of subjects used in the specific clinical trial. R1.5

Bayesian methods are used here in the construction and estimation of the generalized additive model (9; 10) that is associated with each kinetic parameter in the non-linear model of the CTCs. Similar to mixed-effects models in a maximum likelihood setting the variances associated with the fixed effects are chosen to be constant, but the variance terms associated with the random effects are given prior distributions. This leads to a shrinkage estimation of the random effects so that they are pushed towards

zero (11). The fixed effects in the model explain as much variance as possible, whereas the random effects capture variability that cannot be explained by the fixed effects.

Formulation of a Bayesian hierarchical model is typically achieved in three stages: the data model, the parameter model and the prior distributions (12; 13). The data model reflects our knowledge of the CTCs at the voxel level using the class of compartmental models (14; 15; 16) with a standard arterial input function (AIF) taken from the literature (17; 18; 19). The parameter model describes how parameters from the compartmental model are associated with important factors of the clinical trial. At this stage we decompose the kinetic parameters into treatment, patient and voxel effects. We acknowledge that this decomposition will depend on the specific features unique to the clinical trial design. Following from Bayesian theory all parameters are regarded as random variables with pre-specified (prior) distributions. This includes the parameters for the fixed as well as the random effects in the model. Choices have been made in the construction of the Bayesian hierarchical model in order to utilize efficient sampling methods wherever possible, and therefore reduce the computational burden. R1.1

The output from the Bayesian hierarchical model not only answers the basic question from a DCE-MRI oncology study (i.e., did the treatment reduce  $K^{\text{trans}}$ ?) but also provides information about various aspects of the study through the decomposition of the kinetic parameters. Posterior estimates for  $K^{\text{trans}}$  and  $k_{\text{ep}}$  are available to inspect the treatment effect for each subject, averaged over the tumor region of interest, for potential comparison with clinical endpoints. Estimates at the voxel level are also available which allow one to separate spatial heterogeneity within a single region of interest with uncertainty in the kinetic parameters due to a poor model fit. Estimates at the patient level may be used to compare imaging biomarkers (kinetic parameters) with clinical endpoints, such as clinical response or clinical benefit. The breast cancer study analyzed here illustrates the potential of this methodology by interrogating the output from the hierarchical model at several levels.

## 2 Bayesian Hierarchical Model

Bayesian methods rely on the specification of prior distributions  $p(\boldsymbol{\theta})$ , that express our information about the unknown parameters  $\boldsymbol{\theta}$  before any measurements are taken, and combine this information with empirical evidence from direct observations in order to estimate those parameters. To assess the model parameters after observing the data, the posterior distribution  $p(\boldsymbol{\theta} | \mathbf{Y})$  is computed, where  $\boldsymbol{\theta}$  is the vector of all unknown parameters and  $\mathbf{Y}$  is the vector of observations. The posterior distribution of the parameter vector  $\boldsymbol{\theta}$  is obtained by applying Bayes' theorem

$$p(\boldsymbol{\theta} | \mathbf{Y}) = \frac{p(\boldsymbol{\theta}) \ell(\mathbf{Y} | \boldsymbol{\theta})}{\int p(\boldsymbol{\theta}^*) \ell(\mathbf{Y} | \boldsymbol{\theta}^*) d\boldsymbol{\theta}^*}, \quad (1)$$

where  $\ell(\mathbf{Y} | \boldsymbol{\theta})$  denotes the likelihood function of  $\mathbf{Y}$  and  $p(\boldsymbol{\theta})$  the product of all *a priori* probability distribution functions. One can think of the posterior as an update to the prior distribution, our beliefs, on  $\boldsymbol{\theta}$  after measuring a process – producing a mixture of previous knowledge and experimental data. Examples of such Bayesian methodology in neuroimaging exists for functional MRI (20; 21) and also

diffusion tensor imaging (22).

The following sections introduce the key components in the Bayesian hierarchical model: the data model, the parameter model and the prior distributions. Each stage of the model development has been tailored to the analysis of a longitudinal cancer treatment study with two time points. Figure 1 provides a schematic overview of the proposed Bayesian Hierarchical Model (BHM). The three model stages are the rows and the columns represent the “resolution” of the parameters. The kinetic parameters of interest  $K^{\text{trans}}$  and  $k_{\text{ep}}$  are decomposed into global (study-wide), subject and voxel effects through the BHM, while  $v_p$  is simply estimated for each voxel without further decomposition. The measurement error term is independent of the specific parameter model and involves both prior and hyperprior distributions. A standard compartmental model is used to describe the concentration time curves observed at each voxel. A generalized additive model (23) is proposed to decompose the kinetic parameters into factors that are relevant to the design of the longitudinal study. Finally, the prior distributions, including necessary hyperparameters, are specified on all factors of the parameter model (these terms will be explained in detail subsequently). These prior distributions are relatively flat, reflecting a lack of knowledge concerning the parameter, but also incorporate biological knowledge, such as a transfer rate must be non-negative, or statistical knowledge, for example a variance must be non-negative.

## 2.1 Data Model

A hierarchical Bayesian framework is used to model the contrast agent concentration time curve (CTC) of all voxels (24). Let  $\mathbf{Y} = [Y(t_1), Y(t_2), \dots, Y(t_T)]^\top$  denote the CTC associated with a single voxel observed at  $T$  time points determined by the image acquisition protocol. The CTC is assumed to follow a standard compartment model (19)

$$C_t(t) = v_p C_p(t) + C_p(t) \otimes K^{\text{trans}} \exp(-t k_{\text{ep}}), \quad (2)$$

where  $\otimes$  denotes the convolution operator,  $K^{\text{trans}}$  represents the transfer rate from plasma to extracellular extravascular space (EES) per minute,  $k_{\text{ep}}$  the rate constant between EES and blood plasma per minute and  $v_p$  the vascular space fraction. The choice of model for the CTC depends on the scientific goals of the study. Replacing Eq. [2] with a more or less complicated model is straightforward in our model-building framework. The vector of observations  $\mathbf{Y}$  may therefore be thought of as noisy measurements of the true contrast agent concentration  $C_t(t)$ , sampled at the discrete values  $t_1, t_2, \dots, t_T$  to produce the vector  $\mathbf{C}_t = [C_t(t_1), C_t(t_2), \dots, C_t(t_T)]^\top$ , given as a realization of a multivariate Normal distribution

$$\mathbf{Y} \sim N_T(\mathbf{C}_t, \sigma^2 I_T), \quad (3)$$

where the notation  $Y \sim N_T(\boldsymbol{\mu}, \Sigma)$  means that the random variable  $Y$  comes from a multivariate Normal distribution of dimension  $T$  with a mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\Sigma$ .

We assume a common arterial input function (AIF), taken from the literature for all patients in the study, and use a bi-exponential function (15)

$$C_p(t) = D[a_1 \exp(m_1 t) + a_2 \exp(m_2 t)], \quad (4)$$

where  $a_1 = 24.0$  kg/l,  $a_2 = 6.20$  kg/l,  $m_1 = 3.00$  min<sup>-1</sup> and  $m_2 = 0.016$  min<sup>-1</sup> are inspired by the work of Fritz-Hansen *et al.* (18). A Bayesian implementation of the compartmental model above has been previously proposed by Schmid *et al.* (25).

## 2.2 Parameter Model

The pharmacokinetic parameters from the data model are defined at every tumor voxel across all subjects and scans. We assume *a priori* that the distribution of the random variables  $K^{\text{trans}}$  and  $k_{\text{ep}}$  in the tumor are patient-specific and are changed by treatment in a similar way. Therefore a generalized additive model is used where the log-transformed kinetic parameters  $\psi_1 = \ln(K^{\text{trans}})$  and  $\psi_2 = \ln(k_{\text{ep}})$  are expressed as a linear combination of fixed- and random-effects associated with identifiable factors in the study. In addition to restricting both parameters to be positive, the log transform is appealing in this context because individual terms in the additive model may be interpreted as a percentage change from baseline. We assume that the distribution of the vascular fraction  $v_p$  will not be modified by the treatment, but individual  $v_p$  values are allowed to change at the voxel level. Let  $i = 1, \dots, I$  denote the scans acquired and let  $j = 1, \dots, J$  denote the patients, so that  $n_{ij}$  denotes the number of tumor voxels for patient  $j$  at scan  $i$ , measured at  $T$  time points.

The factor of interest when measuring a change in the kinetic parameters ( $K^{\text{trans}}$  or  $k_{\text{ep}}$ ) is the treatment effect, or the difference between pre- and post-treatment visits in the study investigated here. It is necessary to model the fact that substantial variability exists across patients in the study and between the voxels in each region of interest (ROI) that describes the enhancing region in the dynamic acquisition. Hence, the model for  $\ln(K^{\text{trans}})$  is given by

$$\psi_{ijk1} = \alpha_1 + \beta_1 x_i + \gamma_{j1} + \delta_{j1} x_i + \epsilon_{ijk1}, \quad \text{for all } i, j, k, \quad (5)$$

where  $k$  denotes a unique voxel in the tumor ROI and

$$x_i = \begin{cases} 0 & \text{scan } i = 1 \text{ (pre-treatment);} \\ 1 & \text{scan } i = 2 \text{ (post-treatment).} \end{cases} \quad (6)$$

The parameter  $\alpha_1$  is the factor of  $\ln(K^{\text{trans}})$  associated with the baseline scan and  $\beta_1$  is the treatment effect (since it is only associated with the post-treatment acquisition). For the clinical study analyzed here, the treatment effect for  $K^{\text{trans}}$ , as quantified via  $\beta_1$ , is the key parameter of interest. A formal hypothesis test will be performed on  $\beta_1$  in order to detect a significant reduction in  $K^{\text{trans}}$  between the pre- and post-treatment scans.

The parameters  $\alpha_1$  and  $\beta_1$  are regarded as fixed effects (the global column of Figure 1), and thus do not vary between patients in the study. In the Bayesian framework, a marginal posterior distribution will be available for each parameter. The parameter  $\gamma_{j1}$  is the effect of patient  $j$  on  $\ln(K^{\text{trans}})$  and  $\delta_{j1}$  is the interaction between patient  $j$  and treatment. Although the parameters  $\alpha_1$  and  $\beta_1$  summarize information for the baseline and treatment effects of  $K^{\text{trans}}$  across all subjects in the study, the parameters  $\gamma_{j1}$  and  $\delta_{j1}$  allow the estimate of  $K^{\text{trans}}$  to adapt to the baseline and treatment effects observed for each subject.

These parameters are random effects since each patient is assumed to be drawn from a larger population of patients suffering from this condition (the subject column of Figure 1). Finally, the parameter  $\epsilon_{ijk1}$  is the random effect of voxel  $k$  in scan  $i$  of patient  $j$  on  $\ln(K^{\text{trans}})$ . The voxel effect models the fact that each voxel in the tumor volume is drawn from a distribution that describes the ideal tumor voxel (the voxel column of Figure 1). The combination of fixed and random effects in a single model is commonly referred to as a mixed-effects model (9).

Using matrix notation, we can combine the generalized additive model across both kinetic parameters,  $\ln K^{\text{trans}}$  and  $\ln k_{\text{ep}}$ , such that

$$\psi_{ijk} = \mathbf{Z}_i \begin{bmatrix} \phi \\ \boldsymbol{\theta}_j \end{bmatrix} + \epsilon_{ijk} \quad (7)$$

$$\mathbf{X}_i = \begin{bmatrix} 1 & x_i & 0 & 0 \\ 0 & 0 & 1 & x_i \end{bmatrix}; \quad \mathbf{Z}_i = [\mathbf{X}_i \ \mathbf{X}_i]; \quad \phi = \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \end{bmatrix}; \quad \boldsymbol{\theta}_j = \begin{bmatrix} \gamma_{j1} \\ \delta_{j1} \\ \gamma_{j2} \\ \delta_{j2} \end{bmatrix}; \quad \epsilon_{ijk} = \begin{bmatrix} \epsilon_{ijk1} \\ \epsilon_{ijk2} \end{bmatrix} \quad (8)$$

The scan-specific covariates in the model are captured in  $\mathbf{Z}_i$ , the fixed effects are in  $\phi$ , the patient-specific random effects are in  $\boldsymbol{\theta}_j$  and the voxel-specific random effects are in  $\epsilon_{ijk}$ . The model formulation in Eq. [7] can be adapted in order to incorporate additional covariates or a greater number of scans (e.g., in a longitudinal study).

### 2.3 Prior Distributions

In the Bayesian framework prior distributions with unknown variances are used to model the random effects (26). We use vector notation to denote the patient-specific random effects such that  $\boldsymbol{\gamma} = [\gamma_{11}, \gamma_{21}, \dots, \gamma_{J1}, \gamma_{12}, \dots, \gamma_{J2}]^T$  and  $\boldsymbol{\delta} = [\delta_{11}, \delta_{21}, \dots, \delta_{J1}, \delta_{12}, \dots, \delta_{J2}]^T$ , where we have dropped the kinetic parameter subscript to simplify the notation. We use multivariate Gaussian distributions to characterize the prior distributions of the patient-specific random effects; i.e.,

$$\boldsymbol{\gamma} \sim N_{2J}(\mathbf{0}, \text{diag}(\boldsymbol{\tau}_\gamma^2)), \quad (9)$$

$$\boldsymbol{\delta} \sim N_{2J}(\mathbf{0}, \text{diag}(\boldsymbol{\tau}_\delta^2)), \quad (10)$$

where  $\boldsymbol{\tau}_\gamma^2$  and  $\boldsymbol{\tau}_\delta^2$  are vectors of the same length and indexed as  $\boldsymbol{\gamma}$  and  $\boldsymbol{\delta}$ , respectively. The voxel-specific random-effect vectors are given unique prior distributions by scan, patient and parameter, so that each vector is given by  $\boldsymbol{\epsilon}_{ijl} = [\epsilon_{ij1l}, \epsilon_{ij2l}, \dots, \epsilon_{ijn_{ij}l}]^T$ . We use a multivariate Gaussian prior distribution

$$\boldsymbol{\epsilon}_{ijl} \sim N_{n_{ij}}(\mathbf{0}, \tau_{\epsilon;ijl}^2 I_{n_{ij}}), \quad (11)$$

where  $n_{ij}$  is the number of voxels in the region of interest of scan  $i$  of patient  $j$ , and  $\tau_{\epsilon;ijl}^2$  is an unknown variance associated with scan  $i$ , patient  $j$  and kinetic parameter  $l$ . Since the variances are unknown

parameters, they must have their own prior distributions which are given by

$$\tau_\gamma^2 \stackrel{\text{iid}}{\sim} \text{IG}(a, b), \quad (12)$$

$$\tau_\delta^2 \stackrel{\text{iid}}{\sim} \text{IG}(a, b), \quad (13)$$

$$\tau_\epsilon^2 \stackrel{\text{iid}}{\sim} \text{IG}(c, d), \quad (14)$$

where  $a = b = c = 1$ ,  $d = 10^{-5}$ , and  $\text{IG}(\cdot, \cdot)$  denotes the inverse Gamma distribution (27). Note, the notation  $\mathbf{X} \stackrel{\text{iid}}{\sim} F$  means that all random variables  $X_i$  in the vector  $\mathbf{X}$  are independently and identically distributed from the distribution  $F$ . The inverse Gamma distribution is known as a conjugate prior for the Normal distribution (12); i.e., the posterior distribution of  $\gamma$  given the variance  $\tau_\gamma^2$  and all other parameters (known as the full conditional distribution) is again a multivariate Gaussian distribution (see the Appendix for further details). For the fixed effects we use uniform priors so that the prior distribution does not contain any relevant information, denoted by

$$p(\alpha_l) = p(\beta_l) = \text{constant} \quad \text{for } l = 1, 2. \quad (15)$$

Note, the prior distributions on  $\alpha$  and  $\beta$  are improper (not a valid probability distribution), but this is not a problem since the full conditional distributions for both parameters are valid distributions (see the Appendix).

The prior distributions on the coefficients in the generalized additive model have been chosen so that as much variance in the data as possible is explained by the fixed effects  $\alpha$  and  $\beta$  – as no prior information is used for those parameters. Variability which cannot be explained by the fixed effects will be covered by the random effects  $\gamma$  and  $\delta$ . For these parameters an additional prior distribution (hyperprior) on the variance of the parameters is defined which leads to a shrinkage of those effects; that is, the parameters are pushed towards zero and therefore do not cover variance explained by the fixed effects. Any additional variance is explained by the voxel effects. Fig. 2 shows the prior probability density function for the patient-specific effects,  $\exp(\gamma)$  and  $\exp(\delta)$ , and for the voxel effect  $\exp(\epsilon)$ . The figure shows that the prior distribution on the voxel effects is quite informative and does *a priori* not allow the voxel effect to deviate much from baseline and patient specific effect. The patient-specific effects are less informative, being stochastically restricted to the typical range of values for  $K^{\text{trans}}$  and  $k_{\text{ep}}$  in breast cancer (28; 25). The prior probability density function of  $\alpha$  and  $\beta$  however can not be drawn, as both priors are uniform. It is worth noting, that the prior distribution of  $\psi$  is also flat; i.e., there is no prior information on  $\log(K^{\text{trans}})$  and  $\log(k_{\text{ep}})$ . R1.7

For the vascular space fraction we impose a relatively flat prior

$$v_{p;ijk} \stackrel{\text{iid}}{\sim} \text{B}(e, f), \quad \text{for all } i, j, k, \quad (16)$$

with  $e = 1$  and  $f = 19$ , where  $\text{B}(e, f)$  denotes the Beta distribution (29), so that the *a priori* expected value of  $v_p$  is  $e/(e + f) = 0.05$ . The Bayesian hierarchical model is complete by specifying a prior distribution for the variance of the observational error in Eq. [3], with one variance parameter per scan per patient,

$$\sigma_{ij}^2 \stackrel{\text{iid}}{\sim} \text{IG}(g, h) \quad \text{for all } i, j, \quad (17)$$



where  $g = 1$  and  $h = 10^{-2}$ . Note, the prior specification depends on the patient and scanning session but is assumed to be the same across voxels.

### 3 Materials and Methods

#### 3.1 Data acquisition

The first twelve patients from a previously reported breast cancer study are included in the analysis (30; 25). Data were provided by the Paul Strickland Scanner Centre (PSSC) at Mount Vernon Hospital, Northwood, UK. Each patient underwent a DCE-MRI study before and after two cycles of chemotherapy (5-fluorouracil, epirubicin and cyclophosphamide). Six of these patients were identified as pathological responders after receiving six cycles of chemotherapy, the others were non-responders.

For the calculation of  $T_1$  values, we used a two-point measurement with calibration curves as described in (31; 32). The  $T_1$  values are computed as ratio of a  $T_1$ -weighted fast low-angle shot (FLASH) image and a proton density weighted (PDw) FLASH image. The imaging parameters of the  $T_1$ -weighted FLASH images were TR = 11 ms, TE = 4.7 ms,  $\alpha = 35^\circ$ , and the parameters of the PDw FLASH images were TR = 350 ms, TE = 4.7 ms,  $\alpha = 6^\circ$ . Field of view was the same for all scans,  $260 \times 260 \times 8$  mm per slice, so voxel dimensions were  $1.016 \times 1.016 \times 8$  mm. A scan consists of three sequential slices of  $256 \times 256$  voxels and one slice placed in the contra lateral breast as control, which we do not use for our analysis. A total of 40 to 50 acquisitions were acquired, with one acquisition each 11.9 seconds. A dose of  $D = 0.1$  mmol per kg body weight of Gd-DTPA was injected after the fourth scan using a power injector with 4 ml/s with a 20 ml saline flush also at 4 ml/s. The first four scans, before contrast, were used to compute  $T_{10}$  as the average of the  $T_1$  values of these images. Data from this study were acquired in accordance with the recommendation given by (33). Informed consent was obtained from all patients.

Regions of interest (ROIs) were drawn manually by an expert radiologist on a scan-by-scan basis using anatomical images and subtraction images from the dynamic data to define tumor voxels in pre- and post-treatment scans.

#### 3.2 Parameter Estimation via MCMC

The proposed Bayesian hierarchical model produces a joint posterior distribution of all parameters by combining the data model (also known as the likelihood function), the parameter model and the prior models (also known as the prior distributions), via Bayes theorem, Eq. [1]. Samples from the posterior distribution may be assessed via Markov chain Monte Carlo (MCMC) (34). Samples from the joint posterior distribution are used to obtain additional information on the accuracy and precision of the estimates. For example, the standard error of the posterior distribution is the observational error. Statistics of interest (e.g., the mean, median, quantiles, etc.) may be derived from the posterior distribution so that not only point estimates but also confidence intervals are readily available for all parameters.

Implementation of the MCMC algorithm involved 25,000 iterations being drawn from the posterior

distribution after an initial burn-in phase. To reduce computer storage, every 10th sample was used, producing a total of 2500 samples to describe the posterior distribution. The sampling rate was determined by visual inspection of the sample autocorrelation function. The burn-in phase involved 10,000 iterations; after this the algorithm was judged to have converged to the invariant distribution and produce samples from the posterior distribution that are independent from the initial starting conditions. The length of the burn-in phase was determined by visual inspection of the sampled parameters. The analysis of the full study took approximately 48 hours. R1.4  
R1.11

The global parameter  $\phi$  and patient-specific parameters  $\theta_j$  were drawn *en bloc* in Gaussian Gibbs steps (35), and hyperparameters were drawn in independent Gamma Gibbs steps (technical details may be found in the Appendix). Metropolis-Hastings steps with random walk proposals were necessary for the voxel-specific random effects and vascular space fraction. The algorithm was tuned to an acceptance rate of 30-50% (36). Summary statistics were computed from the samples of the posterior distribution to provide point estimates of the parameters from the generalized additive model. Empirical standard errors, along with sample quantiles, were used to characterize the precision of the parameter estimates.

## 4 Results

All parameter estimates are derived from the posterior distribution using Bayes theorem. Hence, a sampling distribution for each parameter value has been built up from which we produce a point estimate via the median of the sample and also credible intervals (Bayesian confidence intervals) by using the quantiles from their sampling distributions.

How the individual parameters from the generalized additive model coalesce to fit the observed contrast agent concentration time curve is illustrated, at the voxel level, in Figure 3. The observed CTCs for two voxels from three subjects, one voxel at baseline and one voxel after treatment, are plotted along with three fitted curves. The best estimate from the Bayesian hierarchical model at a specific voxel is provided by the solid lines in each plot. That is, all parameters from the generalized additive model in Eq. [7] are used in the parameter model in order to fit the data model. These curves are very similar to, but not exactly the same as, model fits from the standard non-linear regression method used in the quantitative analysis of DCE-MRI data (25). Removing the voxel-specific term from the model produces a fitted curve that is associated with patient and treatment effects, but not the specific voxel, and are plotted as dashed lines in Figure 3. Given the presence of inter-voxel heterogeneity in the tumor ROI, the dashed lines may or may not fit the observed data at a given voxel very well but they do represent the best (in the sense of a posterior median) fit to all voxels in the tumor ROI for a given patient at a single scan time point. Going back one more level in the generalized additive model and removing the patient effect leaves a fitted curve associated with the baseline and post-treatment scans (i.e., two curves that summarize the overall treatment effect) given by the dotted lines. The top row of Figure 3 contains voxels from three subjects before treatment so the dotted lines are identical and represent the best (in the sense of a posterior median) fit to all pre-treatment voxels across all subjects. The bottom row contains

voxels from the same subjects after treatment and the dotted line is the best fit to all post-treatment voxels.

Figure 4 shows the posterior distributions of pre-treatment (baseline)  $K^{\text{trans}}$  and post-treatment  $K^{\text{trans}}$ . That is, the posterior samples were transformed via  $\exp(\alpha_1)$  and  $\exp(\alpha_1 + \beta_1)$ , respectively. For ease of comparison between the two posterior distributions a smoothed version of each histogram, known as a kernel density estimate (37) is displayed. The posterior median of  $K^{\text{trans}}$  at baseline is 0.205, and the posterior median of  $K^{\text{trans}}$  after treatment is 0.156. Credible intervals for  $K^{\text{trans}}$ , that cover 95% of the posterior distribution, are  $[0.186, 0.234]$  at baseline and  $[0.121, 0.198]$  after treatment. That is, the true value of  $K^{\text{trans}}$  lies in the interval  $[0.186, 0.234]$  with posterior probability 0.95 at baseline and in  $[0.121, 0.198]$  with posterior probability 0.95 after treatment.

The density estimates in Figure 4 are unimodal and indicate an overall decrease in  $K^{\text{trans}}$  after treatment. In order to test for a treatment effect on  $K^{\text{trans}}$ , specifically a reduction in  $K^{\text{trans}}$  in the second acquisition compared to the first, we construct the hypothesis

$$H_0 : \beta_1 > 0 \quad \text{versus} \quad H_1 : \beta_1 \leq 0, \quad (18)$$

using the treatment effect from the parameter model in Eq. [7] and calculate the posterior probability of  $\beta_1$  exceeding zero. From the results of the MCMC simulation, the null hypothesis is rejected with  $p = 0.001$ .

When the generalized additive model was introduced previously the fact that the parameter  $K^{\text{trans}}$  and the covariates are linked through a logarithmic transform leads to the interpretation of individual covariates in the parameter model as percentage changes from baseline instead of absolute changes. For the treatment effect this translates into a  $100\% \cdot |0.7659 - 1| = 23.3\%$  median reduction in  $K^{\text{trans}}$  from baseline, where the sign determines whether the change is associated with an increase or decrease.

Figure 5 shows the patient-specific posterior distributions for pre-treatment  $K^{\text{trans}}$ , given by  $\exp(\alpha_1 + \gamma_{j1})$  for  $j = 1, \dots, 12$ , and post-treatment  $K^{\text{trans}}$ , given by  $\exp(\alpha_1 + \beta_1 + \gamma_{j1} + \delta_{j1})$  for  $j = 1, \dots, 12$ . The clinical responders are grouped in the first two columns of Figure 5 and the clinical non-responders are in the third and fourth columns. The same range for  $x$ -axis  $[0, 0.45]$  was used in all plots of  $K^{\text{trans}}$  for comparison. In general the decrease in  $K^{\text{trans}}$  observed in the clinical responders is greater than the clinical non-responders, but this is not absolute. For example, patient 12 shows only a small decrease in  $K^{\text{trans}}$  post-treatment and patient 6 shows an increase in  $K^{\text{trans}}$  after treatment, but both are clinical responders after additional chemotherapy. The interpretation of the treatment effect as a percentage change from baseline helps to quantify the results in Figure 5. The median percentage change in  $K^{\text{trans}}$  for subject  $j$  is obtained via  $100\% \cdot |\exp(\hat{\beta}_1 + \hat{\delta}_{j1}) - 1|$ , where the sign determines whether an increase or decrease occurred. For example, patient 1 (pathological responder) experienced a  $100\% \cdot |0.7684 - 1| = 23.2\%$  median reduction in  $K^{\text{trans}}$  which is very similar to the overall treatment effect. This is definitely not the norm as patient 9 experienced a  $100\% \cdot |0.4285 - 1| = 57.2\%$  median reduction in  $K^{\text{trans}}$  and patient 6 experienced a  $100\% \cdot |1.0817 - 1| = 8.17\%$  median *increase* in  $K^{\text{trans}}$ , both were pathological responders.

Figure 6 shows the voxel-specific median posterior for pre- and post-treatment  $K^{\text{trans}}$ . The clinical

responders are grouped in the first two columns and the clinical non-responders are in the third and fourth columns (identical to Fig. 5). The range for the  $x$ -axis was restricted to  $[0, 1]$  in all plots for comparison. Given the number of samples from the posterior distribution across all voxels, the median value of  $\exp(\alpha_1 + \gamma_{j1} + \epsilon_{1jk1})$  for  $j = 1, \dots, 12; k = 1, \dots, n_{1j}$  and  $\exp(\alpha_1 + \beta_1 + \gamma_{j1} + \delta_{j1} + \epsilon_{2jk1})$  for  $j = 1, \dots, 12; k = 1, \dots, n_{2j}$  across the 2500 samples, for each voxel  $k$ , was computed to summarize the voxel effect. The resulting histograms for the voxel effect have been summarized by a kernel density estimate. Most voxel-level distributions of median  $K^{\text{trans}}$  show a substantial change in shape after treatment, although this is more apparent in the responders compared to the non-responders.

It is interesting to note the extent of changes in the shape of these distributions across the different subjects. For example, patient 11 is characterized by a tumor with two distinct modes in estimated  $K^{\text{trans}}$  at baseline and a single mode after treatment. Looking at the statistical images of the median posterior  $K^{\text{trans}}$  at baseline the tumor ROI for patient 11 includes a substantial number of non-enhancing voxels, in addition to those with reasonable  $K^{\text{trans}}$  values, that are contributing to the bi-modal appearance of the posterior distribution. The post-treatment tumor ROI is similar in size (i.e., number of voxels), but the vast majority of voxels do not appear to be enhancing as measured by  $K^{\text{trans}}$ . Hence, a single mode is present in the posterior distribution. This variability, due to ROI definition, has been previously documented (38). The distributions of median  $K^{\text{trans}}$  for patient 8 show the reverse effect, albeit much more subtle than patient 11, where the post-treatment distribution of median  $K^{\text{trans}}$  appears to be bimodal but still spans a similar range of values. The statistical images of median  $K^{\text{trans}}$  indicate a tumor ROI with a highly perfused rim and substantial non-enhancing core at baseline, contributing to the relatively wide histogram in Fig. 6. The post-treatment images have a greatly reduced tumor ROI with a mixture of enhancing and non-enhancing voxels. Looking at the statistical images of the median posterior  $K^{\text{trans}}$  for this patient, the difference between pre- and post-treatment is likely to be due to tumor compaction and the removal of dead cells caused by chemotherapy. Such large effects in tumor size are well documented in chemotherapy of breast cancer. In addition to the biological explanation, some of the perceived difference may be attributed to the visualization technique used. The close proximity of multiple peaks in the smoothed histogram of  $K^{\text{trans}}$  makes it difficult to draw firm conclusions since a slightly wider smoothing kernel could eliminate this apparent feature and produce a broad single peak.

## 5 Discussion

Information is obtained at multiple levels during an imaging study in the clinical trial setting. The main scientific question of interest is usually, was there a treatment effect? This key hypothesis test drives study design by influencing critical experimental design parameters such as power and sample size. However, information at other levels, such as the patient or voxel level, can provide insight into much more subtle features concerning patients, tumors and the treatment effect. Patient variability with application to predicting clinical response and tumor heterogeneity, as measured by voxel-wise properties of the pharmacokinetic model, are just two examples of so-called secondary endpoints.

The Bayesian hierarchical model presented here was developed to test the hypothesis of a treatment effect for an imaging study while acknowledging known sources of uncertainty; e.g., patients and voxels. This is similar to the approach taken in standard analysis methods for clinical trials where fixed and random effects are identified in the model (5). The specification of such fixed and random effects allows the results from the study to be applicable beyond the patient population recruited for a specific study.

The results from all levels of the Bayesian hierarchical model have been presented using smoothed histograms in order to convey information regarding the posterior distributions of the parameter estimates. When interpreting the voxel-level results, spatial information (such as the spatial pattern of the parameter estimate and the size of the tumor ROI) are lost in such a statistical summary. It is important to also view the statistical images overlaid on suitable anatomical image in order to provide physiological explanations for the observed shapes of the voxel-level posterior distributions. Even when drawn by an experienced radiologist the tumor ROIs may substantially influence the quantitative results.

## 5.1 Comparison with Nonlinear Regression

A standard analysis was performed on the ROIs, using nonlinear regression to estimate the kinetic parameters in the compartmental model over all voxels independently, and the median  $K^{\text{trans}}$  values have been summarized in Table 1. A non-parametric test (one-sided Wilcoxon signed rank test) was performed to test that the difference between the median values was greater than zero; i.e., the treatment did not reduce  $K^{\text{trans}}$  across all subjects. The null hypothesis was rejected at a borderline significance level ( $p = 0.055$ ). Given the small sample size,  $N_1 = 6$  responders and  $N_2 = 6$  non-responders, this is an impressive result and there is obviously a reasonable difference in  $K^{\text{trans}}$  between the two groups. Fig. 7 shows the kernel density estimates of  $K^{\text{trans}}$  for each ROI, before and after treatment, using a voxel-wise nonlinear regression analysis. That is, the compartmental model in Eq. [2] was fitted to each voxel independently using the Levenberg-Marquardt optimization procedure. The empirical distributions observed for each patient are extremely similar to those obtained in the BHM, Fig. 6. This is to be expected given the relatively flat priors that were imposed on the kinetic parameters (25). The only difference between Figs. 6 and 7 is the method of kinetic parameter estimation – Bayesian versus non-linear regression.

R1.10

While the voxel-wise results from the Bayesian and regression methods are very similar, and thus provide a check on the consistency of the Bayesian model fitting procedure, the advantages of the Bayesian hierarchical model are clear through the coefficients from the generalized additive model in Eq. [7]. The regression analysis can only summarize the study through Table 1, but the BHM allows one to isolate and interrogate specific effects, at the study or patient or voxel level, through the generalized additive model. Examples of such interrogations have been presented here in Figures 4 and 5, but the possibilities for such model summaries are only limited by the construction of the parameter model and design of the clinical trial.

## 5.2 Prior assumptions

Bayesian models rely on *a priori* beliefs about the model and parameters, expressed as prior distributions. When combined with observed data, these beliefs are updated to reflect both sources of information. In general, the use of relatively flat (uninformative) prior distributions produces a parameter estimates with similar characteristics to those obtained from maximum-likelihood estimation, and hence similar results overall. In the proposed Bayesian hierarchical model our prior beliefs are based on a mixed-effect model and follow suggestions in the Bayesian mixed-effect literature (26) and in kinetic modelling in DCE-MRI (28; 25). The baseline and treatment effect are modelled as fixed effects, and thus uniform priors are used for those parameters. The patient-specific effects and the voxel-specific effect are modelled as mixed effects, where a Gaussian distribution with zero mean and unknown variance is used. This is also known as shrinkage prior, as it shifts the parameters towards zero. The so-called hyperprior on the unknown variance terms in these prior distributions determines how much variability is accumulated by the fixed effects versus the random effects. Hence, the hyperpriors are chosen so that the baseline and treatment effects explain as much variability as possible. The patient-specific effects accumulate variability not explained by baseline and treatment and the voxel effect accumulates variability not explained by baseline, treatment or the patient-specific effects. The prior on the voxel effects is quite R1.10  
informative and ensures that these effects are only used when the fixed effects, and the patient-specific effects, are unable to model the observed concentration time curve; i.e., the information contained in the MR measurements must overcome the relatively strong prior specification. Fig. 8 depicts prior and R1.9  
posterior distribution of patient-specific and voxel effects. As previously mentioned the prior distribution on the voxel effects is informative, but the posterior distributions differ substantially from the prior and hence are driven by the observed data. For the patient-specific effects, the prior is less informative and again the information contained in the observed data produce posterior distributions with low variance that differ from the prior.

Specification of the hyperpriors was determined by a sensitivity analysis. For example, with an inverse Gamma prior, with parameters  $a = b = 0.01$ , on the variance of the  $\gamma$  and  $\delta$  parameters the algorithm does not show any treatment effect. The patient-treatment interaction term accumulates too much of the variability, as the hyperprior is too conservative. The model was found to be insensitive to changes in the specification of the hyperprior on  $\epsilon$ , where values of  $d$  between  $10^{-3}$  and  $10^{-6}$  were found not to impact the results.

The mixed-effect model is built on top of a standard compartmental model used in DCE-MRI, and we use the logarithms of  $K^{\text{trans}}$  and  $k_{\text{ep}}$  in the specification. This transformation not only ensures positive values for both parameters, but also allows the assumption that the transformed values are Gaussian distributed (25). Hence, we can use a generalized additive model on those parameters. The vascular fraction  $v_p$  takes on values between zero and one, and hence we use a Beta distribution, with hyperparameters  $e$  and  $f$ , such that the expected value of  $v_p$  is 0.05. This value is appropriate for the data involved in this study and in fact the model is relatively insensitive to the specification of this particular prior distribution. A Beta distribution with  $e = f = 1$  (i.e., a uniform distribution on  $[0, 1]$ )

does not change the results of the model. For the observational model, we use a Gaussian distribution with unknown variance.

We did not use prior information to impose a specific relationship between  $K^{\text{trans}}$  and  $k_{\text{ep}}$ ; e.g.,  $K^{\text{trans}} < k_{\text{ep}}$ . From a modelling point of view, the compartmental model is a rough approximation to the physical system that is being investigated via DCE-MRI. At least two layers of abstraction sit between the biological truth and what is measured in DCE-MRI as inferred via a parametric model: the data acquired via MRI are imperfect measurements of the biological system because of the interaction between physics and biology in the measurement process; and the fact that the “extended Kety model” is a gross approximation to the underlying system. Hence, the kinetic parameters obtained from the extended Kety model are relatively far removed from the biological processes and we feel it would be too restrictive to enforce strict relationships between them in the BHM. In the results from the BHM on 12 subjects presented here, we found that  $K^{\text{trans}}$  was larger than  $k_{\text{ep}}$  in less than 1.5% of all voxels analyzed. These were typically located at the edge of the tumor ROI, where perfusion characteristics often cause errors in parameter estimation (2). R1.12

The signal intensity in magnitude MR images is known to be Rician, and approximately Gaussian for moderate signal-to-noise values, but the distribution of the error of the contrast agent concentration is a non-linear transformation of the observed signal intensities and, to the best of our knowledge, is not known analytically. However, for positive values of contrast agent concentration in tissue a Gaussian distribution seems appropriate. The hyperprior was determined by sensitivity analysis, and the model was not found to be sensitive to changes in the specification of this hyperprior.

### 5.3 Modifications and Extensions

In this paper a generalized additive model was constructed for the kinetic parameters ( $K^{\text{trans}}$  and  $k_{\text{ep}}$ ) in a compartmental model of a DCE-MRI oncology study. The model can easily be modified for other kinetic parameters of interest; e.g., if the variable of interest in a study is the volume fraction of the EES ( $v_e = K^{\text{trans}}/k_{\text{ep}}$ ) or  $v_p$ , which was handled as nuisance parameter in the study investigated here. The framework can also be applied to other kinetic models (39; 40). R1.6

This model incorporated two scanning sessions, and all subjects, to assess the treatment effect. The modeling framework is easily extended to handle additional covariates or scanning sessions. For example, a dose-ranging study design could be incorporated into the additive model where the treatment effect can be expressed as a function of the dose. Additional scans over time would enable the assessment of temporal dependence on treatment and provide information about the reliability of the data by potentially reducing the amount of uncertainty in the parameter estimates.

Another possible extension of this model would be to include the spatial information of adjacent voxels. In the current implementation of the Bayesian hierarchical model all voxels from one region of interest (tumor) were treated as spatially independent. Since voxel borders are arbitrary and do not represent physiological boundaries between different tissue types, it is likely that neighboring voxels share similar perfusion characteristics. This fact has been taken advantage of in the context of Bayesian modeling of

individual scans from a DCE-MRI study (25). The inclusion of a neighborhood structure in the modeling process would reduce the uncertainty in estimation and provide more reliable estimates of the kinetic parameters.

## 6 Acknowledgements

Support for VJ Schmid was financed through a research grant from GlaxoSmithKline and a platform grant (GR/T06735/01) from the UK Engineering and Physical Sciences Research Council (EPSRC).

## A Appendix

### A.1 Full conditional distributions

In each iteration of the MCMC (Markov chain Monte Carlo) algorithm, a random sample from the marginal posterior distribution for all parameters is drawn. This is performed by sampling from the conditional posterior distribution of one or more parameters given all other parameters and the data. Hence, the full conditional distributions must be computed. The full conditional is denoted by  $\theta | \cdot$ , where  $\theta$  is the parameter and  $\cdot$  denotes all other parameters and the data. If the full conditional takes the form of a standard distribution, one can sample directly from this distribution; this is known as the Gibbs sampler (34). If the full conditional is not a standard distribution, then a Metropolis-Hastings sampler must be constructed.

In the proposed Bayesian hierarchical model all full conditionals are from standard distributions due to the use of conjugate prior distributions, except for the voxel effect and  $v_p$ . Let  $\boldsymbol{\xi}_l = (\alpha_l, \beta_l, \gamma_l, \boldsymbol{\delta}_l)$  denote the vector of length  $P = I(J+1)$  associated with all parameters in the generalized additive model, except the voxel effect, for a specific kinetic parameter. The full conditional of  $\boldsymbol{\xi}_l$  is a multivariate Normal distribution given by

$$\begin{aligned} \boldsymbol{\xi}_l | \cdot &\sim \text{N}_P(\mathbf{V}^{-1}\mathbf{m}, \mathbf{V}^{-1}), \\ \mathbf{m} &= [m_1, \dots, m_P]^\top, \\ m_p &= \sum_{i=1}^2 \sum_{j=1}^J \left( \tau_{\epsilon;ij}^{-2} \sum_{k=1}^{n_{ij}} w_{ijp} \psi_{ijkl} \right) \quad \text{for } p = 1, \dots, P, \\ \mathbf{V} &= \mathbf{W}^\top \Lambda \mathbf{W} + \text{diag}(0, 0, \tau_{\gamma;1l}^{-2}, \dots, \tau_{\gamma;Jl}^{-2}, \tau_{\delta;1l}^{-2}, \dots, \tau_{\delta;Jl}^{-2}), \end{aligned}$$

where  $\mathbf{W}$  is a  $I(J+1) \times P$  matrix indicating which covariate should be included in the parameter model, Eq. [7], and  $\Lambda$  is a diagonal matrix with elements  $n_{ij}\tau_{\epsilon;ij}$ . The vector  $\boldsymbol{\xi}_l$  is drawn in one block from a multivariate Normal distribution with an efficient block-sampling algorithm (41).

The full conditional distribution of the voxel effect  $\epsilon_{ijkl}$  is a non-standard distribution. For computational reasons it is more convenient to sample from  $\psi_{ijkl}$  rather than from  $\epsilon_{ijkl}$ , where the full conditional



distribution of  $\psi_{ijkl}$  is given by

$$p(\psi_{ijkl} | \cdot) \propto \exp \left( -\frac{1}{2\tau_{\epsilon;ijk}^2} \psi_{ijkl}^2 - \frac{1}{2\sigma_{ij}^2} (Y_{ijkl} - \widehat{Y}_{ijkl})^2 \right). \quad (19)$$

Note,  $\widehat{Y}_{ijkl}$  is the estimated contrast agent concentration curve given by the estimated model parameters in  $\widehat{\psi}_{ijkl}$ . Samples from this distribution are obtained using a Metropolis-Hastings step.

The full conditionals of all variance parameters are inverse Gamma distributions, which are given by

$$\tau_{\gamma}^2 | \cdot \stackrel{\text{iid}}{\sim} \text{IG} (a + 1/2, b + \gamma_{jl}^2), \quad (20)$$

$$\tau_{\delta}^2 | \cdot \stackrel{\text{iid}}{\sim} \text{IG} (a + 1/2, b + \delta_{jl}^2), \quad (21)$$

$$\tau_{\epsilon}^2 | \cdot \stackrel{\text{iid}}{\sim} \text{IG} \left\{ c + \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J n_{ij}, d + \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^{n_{ij}} \left( \mathbf{z}_i \begin{bmatrix} \phi \\ \boldsymbol{\theta}_j \end{bmatrix} - \psi_{ijkl} \right)^2 \right\}. \quad (22)$$

Hence, the variance parameters can be drawn independently using a Gibbs sampler.

## References

- [1] Padhani AR, Yarnold J, Regan J, Husband JE. Dynamic MRI of breast hardness following radiation treatment. *Journal of Magnetic Resonance Imaging* 2003;17:427–434.
- [2] Collins DJ, Padhani AR. Dynamic magnetic resonance imaging of tumor perfusion. *IEEE Engineering in Biology and Medicine Magazine* 2004;65–83.
- [3] Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, Larsson HBW, Lee TY, Mayr NA, Parker GJM, Port RE, Taylor J, Weiskoff R. Estimating kinetic parameters from dynamic contrast-enhanced T<sub>1</sub>-weighted MRI of a diffusable tracer: Standardized quantities and symbols. *Journal of Magnetic Resonance Imaging* 1999;10:223–232.
- [4] Pinheiro JC, Bates DM. *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag, 2000.
- [5] Brown H, Prescott R. *Applied Mixed Models in Medicine*. Chichester, UK: John Wiley & Sons, 1999.
- [6] Beckmann CF, Jenkinson M, Smith SM. General multilevel linear modeling for group analysis of fMRI. *NeuroImage* 2003;20(2):1052–1063.
- [7] Friston KJ, Stephan KE, Lund TE, Morcom A, Kiebel S. Mixed-effects and fMRI studies. *NeuroImage* 2005;24:244–252.
- [8] Roche A, Mériaux S, Keller M, Thirion B. Mixed-effect statistics for group analysis in fMRI: A non-parametric maximum likelihood approach. *NeuroImage* 2007;38(3):501–510.
- [9] Fahrmeir L, Lang S. Bayesian inference for generalized additive mixed models based on Markov random field priors. *Applied Statistics* 2001;50:201–220.
- [10] Fahrmeir L, Tutz G. *Multivariate Statistical Modelling Based on Generalized Linear Models*. New York: Springer, 2nd edition, 2001.

- [11] Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B* 1996;58:267–288.
- [12] Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. Boca Raton, Florida: Chapman & Hall/CRC, 2nd edition, 2003.
- [13] Wikle CK. Hierarchical models in environmental science. *International Statistical Review* 2002; 71:181–199.
- [14] Kety S. Blood–tissue exchange methods. *Theory of blood-tissue exchange and its application to measurement of blood flow. Methods in Medical Research* 1960;8:223–227.
- [15] Tofts PS, Kermode AG. Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts. *Magnetic Resonance in Medicine* 1984; 17(2):357–367.
- [16] Larsson HB, Tofts PS. Measurement of the blood-brain barrier permeability and leakage space using dynamic Gd-DTPA scanning—a comparison of methods. *Magnetic Resonance in Medicine* 1992; 24(1):174–176.
- [17] Weinmann HJ, Laniado M, Mutzel W. Pharmacokinetics of Gd-DTPA/dimeglumine after intravenous injection into healthy volunteers. *Physiological Chemistry and Physics and Medical NMR* 1984;16:167–172.
- [18] Fritz-Hansen T, Rostrup E, Larsson HBW, Søndergaard L, Ring P, Henriksen O. Measurement of the arterial concentration of Gd-DTPA using MRI: A step toward quantitative perfusion imaging. *Magnetic Resonance in Medicine* 1996;36:225–231.
- [19] Buckley DL, Parker GJM. Measuring contrast agent concentration in  $T_1$ -weighted dynamic contrast-enhanced MRI. In A Jackson, DL Buckley, GJM Parker, eds., *Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology*. Berlin: Springer, 2005; 69–80.
- [20] Genovese CR. A Bayesian time-course model for functional magnetic resonance imaging data (with discussion). *Journal of the American Statistical Association* 2000;95(451):691–719.
- [21] Friston KJ, Penny W, Phillips C, Hinton G, Ashburner J. Classical and Bayesian inference in neuroimaging: Theory. *NeuroImage* 2002;16:465–483.
- [22] Behrens TEJ, Woolrich MW, Jenkinson M, Johansen-Berg H, Nunes RG, Clare S, Matthews PM, Brady JM, Smith SM. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine* 2003;50:1077–1088.
- [23] Hastie TJ, Tibshirani RJ. *Generalized Additive Models*. Chapman & Hall/CRC, 1990.
- [24] Gelman A, Stern H, Carlin J, Rubin D. *Bayesian Data Analysis*. Chapman & Hall/CRC Press, London/Boca Raton, 2003.
- [25] Schmid VJ, Whitcher B, Padhani AR, Taylor NJ, Yang GZ. Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging. *IEEE Transactions on Medical Imaging* 2006;25(12):1627–1636.

- [26] Clayton DG. Generalized linear mixed models. In WR Gilks, S Richardson, DJ Spiegelhalter, eds., Markov Chain Monte Carlo in Practice. Chapman & Hall, London, 1996; 275–301.
- [27] Johnson NL, Kotz S, Balakrishnan N. Continuous Univariate Distributions, volume 1. New York: John Wiley & Sons, 2nd edition, 1994.
- [28] Parker G, Suckling J, Tanner S, Padhani A, Husband J, Leach M. MRIW: parametric analysis software for contrast-enhanced dynamic MR imaging in cancer. Radiographics 1998;18:497–506.
- [29] Johnson NL, Kotz S, Balakrishnan N. Continuous Univariate Distributions, volume 2. New York: John Wiley & Sons, 1995.
- [30] Ah-See MLW, Makris A, Taylor NJ, Burcombe RJ, Harrison M, Stirling JJ, Richman PI, Leach MO, Padhani AR. Does vascular imaging with MRI predict response to neoadjuvant chemotherapy in primary breast cancer? Journal of Clinical Oncology (Meeting Abstracts) 2004;22(14S):582.
- [31] Parker G, Suckling J, Tanner S, Padhani A, Revell P, Husband J, Leach M. Probing tumor microvascularity by measurement, analysis and display of contrast agent uptake kinetics. Journal of Magnetic Resonance Imaging 1997;7:564–574.
- [32] d’Arcy J, Collins D, Padhani A, Walker-Samuel S, Suckling J, Leach M. Informatics in radiology (infoRAD): Magnetic resonance imaging workbench: analysis and visualization of dynamic contrast-enhanced MR imaging data. Radiographics 2006;26(2):621–632.
- [33] Leach MO, Brindle KM, Evelhoch JL, Griffiths JR, Horsman MR, Jackson A, Jayson GC, Judson IR, Knopp MV, Maxwell RJ, McIntyre D, Padhani AR, Price P, Rathbone R, Rustin GJ, Tofts PS, Tozer GM, Vennart W, Waterton JC, Williams SR, Workman P. The assessment of antiangiogenic and antivascular therapies in early-stage clinical trials using magnetic resonance imaging: issues and recommendations. British Journal of Cancer 2005;92:1599–1610.
- [34] Gilks WR, Richardson S, Spiegelhalter DJ, eds. Markov Chain Monte Carlo in Practice. Chapman & Hall, London, 1996.
- [35] Liu J, Wong W, Kong A. Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. Biometrika 1994;81:27–40.
- [36] Schmid VJ, Whitcher B, Yang GZ, Taylor NJ, Padhani AR. Statistical analysis of pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging. In J Duncan, G Gerig, eds., Medical Imaging Computing and Computer-Assisted Intervention – MICCAI 2005, number 3750 in Lecture Notes in Computer Science. Berlin: Springer, 2005; 886–893.
- [37] Silverman BW. Density Estimation. London: Chapman & Hall, 1986.
- [38] Beresford MJ, Padhani AR, Taylor NJ, Ah-See ML, Stirling JJ, Makris A, d’Arcy JA, Collins DJ. Inter- and intraobserver variability in the evaluation of dynamic breast cancer MRI. Journal of Magnetic Resonance Imaging 2006;24:1316–1325.
- [39] StLawrence KS, Lee TY. An adiabatic approximation to the tissue homogeneity model for water

exchange in the brain: I. theoretical derivation. *Journal of Cerebral Blood Flow & Metabolism* 1998;18:1365–1377.

[40] Port RE, Knopp MV, Hoffmann U, Milker-Zabel S, Brix G. Multicompartment analysis of gadolinium chelate kinetics: Blood-tissue exchange in mammary tumors as monitored by dynamic MR imaging. *Journal of Magnetic Resonance Imaging* 1999;10:233–241.

[41] Rue H. Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society Series B* 2001;63:325–338.

## Figure Captions

1. Schematic overview of the Bayesian hierarchical model for the observed contrast agent concentration time curves.
2. Prior probability density functions for patient-specific effects  $\exp(\gamma)$  and  $\exp(\delta)$  (top) and voxel effects  $\exp(\epsilon)$  (bottom). The prior for the patient-specific effects allows a deviation from the baseline kinetic parameters in a range of biologically meaningful values. The prior distribution for the voxel effects only permits small deviations from the baseline. R1.7
3. Contrast concentration time curves (CTCs) for pre- and post-treatment scans in three different subjects. Solid lines indicate the voxel-specific fit to the CTC, dashed lines the combined patient- and treatment-specific CTCs, and dotted lines the global pre- and post-treatment median CTCs for the entire study.
4. Samples from the marginal posterior distributions of  $K^{\text{trans}}$  at the study level. At pre-treatment  $K^{\text{trans}}$  is given by  $\exp(\alpha_1)$  and at post-treatment  $K^{\text{trans}}$  is given by  $\exp(\alpha_1 + \beta_1)$ .
5. Samples from the marginal posterior distributions of  $K^{\text{trans}}$  at the patient level. At pre-treatment  $K^{\text{trans}}$  is given by  $\exp(\alpha_1 + \gamma_{j1})$  for patient  $j$  and at post-treatment  $K^{\text{trans}}$  is given by  $\exp(\alpha_1 + \beta_1 + \gamma_{j1} + \delta_{j1})$  for patient  $j$ .
6. Smoothed histograms summarizing the values of the posterior median  $K^{\text{trans}}$  at the voxel level. At pre-treatment  $K^{\text{trans}}$  is given by  $\exp(\alpha_1 + \gamma_{j1} + \epsilon_{1jk1})$  for scan 1, patient  $j$  and voxel  $k$ . At post-treatment  $K^{\text{trans}}$  is given by  $\exp(\alpha_1 + \beta_1 + \gamma_{1j} + \delta_{1j} + \epsilon_{12jk})$  for scan 2, patient  $j$  and voxel  $k$ . The  $x$ -axis has been restricted to  $[0, 1]$  for visualization.
7. Smoothed histograms summarizing the values of  $K^{\text{trans}}$  from voxel-wise non-linear regression analysis. The  $x$ -axis has been restricted to  $[0, 1]$  for visualization.
8. Prior (bold line) and posterior (thin lines) probability density functions for patient-specific effects  $\gamma$  (top) and voxel effects  $\epsilon$  (bottom). For the voxel-specific effects, 25 random voxels from the pre-treatment scan of patient 2 have been displayed. R1.9

Table 1: Median  $K^{\text{trans}}$  values from the standard analysis (R = responder, NR = non-responder).

Patient ID	1	2	3	4	5	6	7	8	9	10	11	12
pathological	R	R	R	NR	NR	R	NR	NR	R	NR	NR	R
pre	0.208	0.355	0.255	0.230	0.199	0.154	0.264	0.198	0.305	0.267	0.432	0.174
post	0.161	0.120	0.031	0.245	0.208	0.173	0.327	0.223	0.122	0.221	0.111	0.113

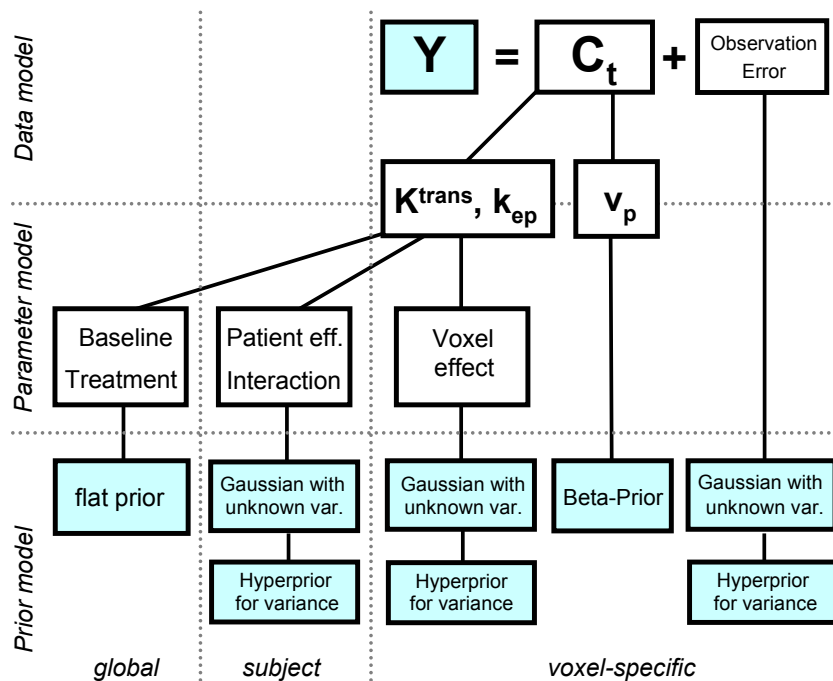


Figure 1:

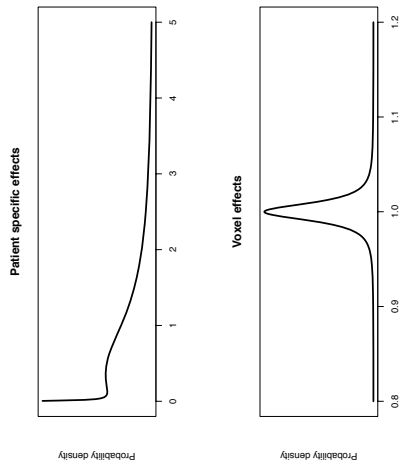


Figure 2:

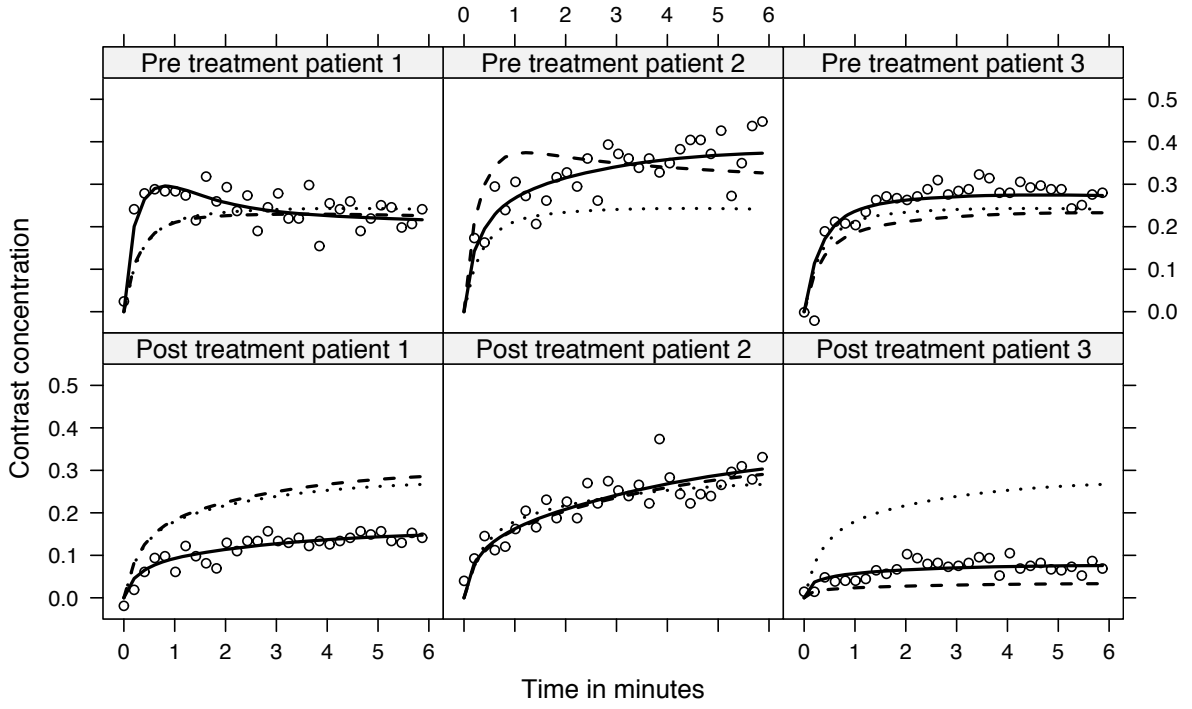


Figure 3:

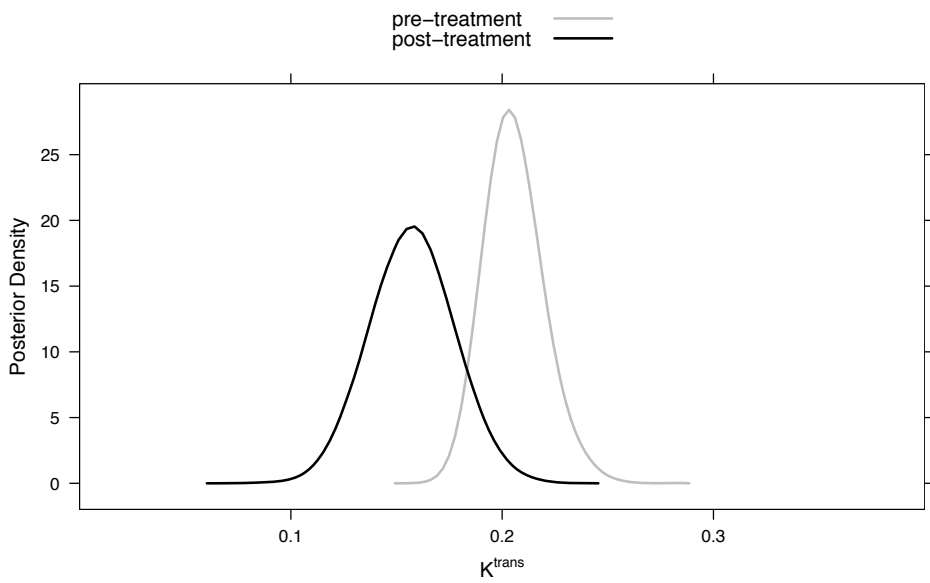


Figure 4:



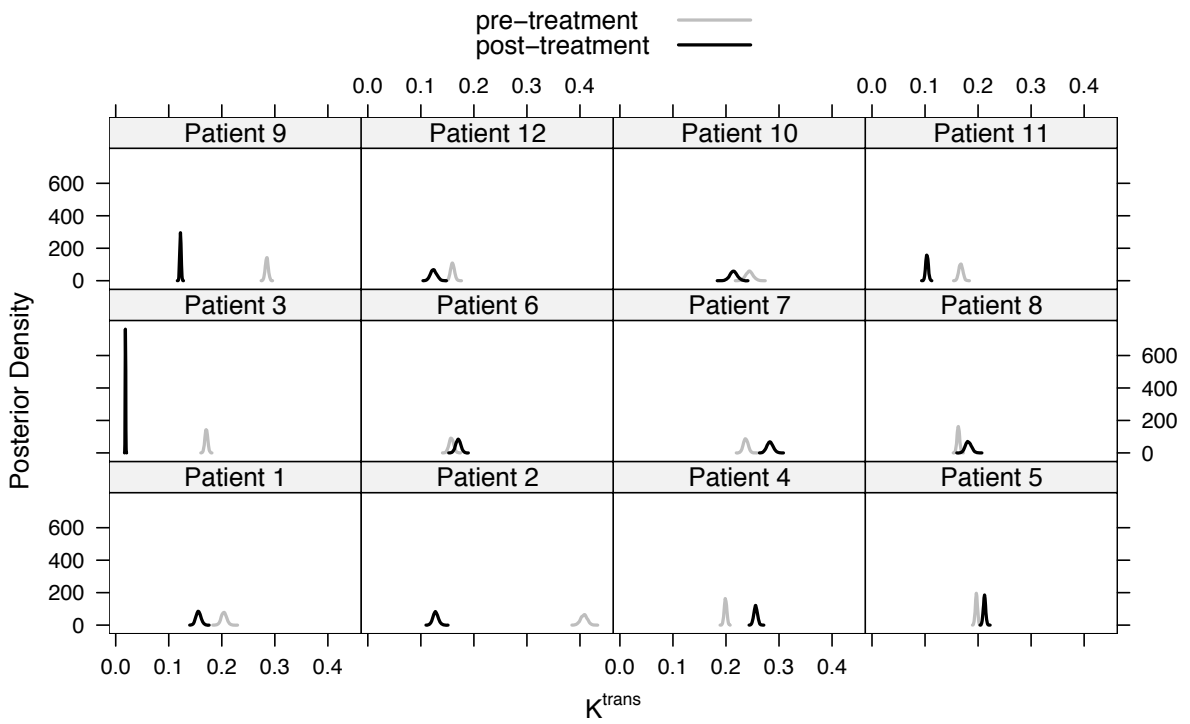


Figure 5:

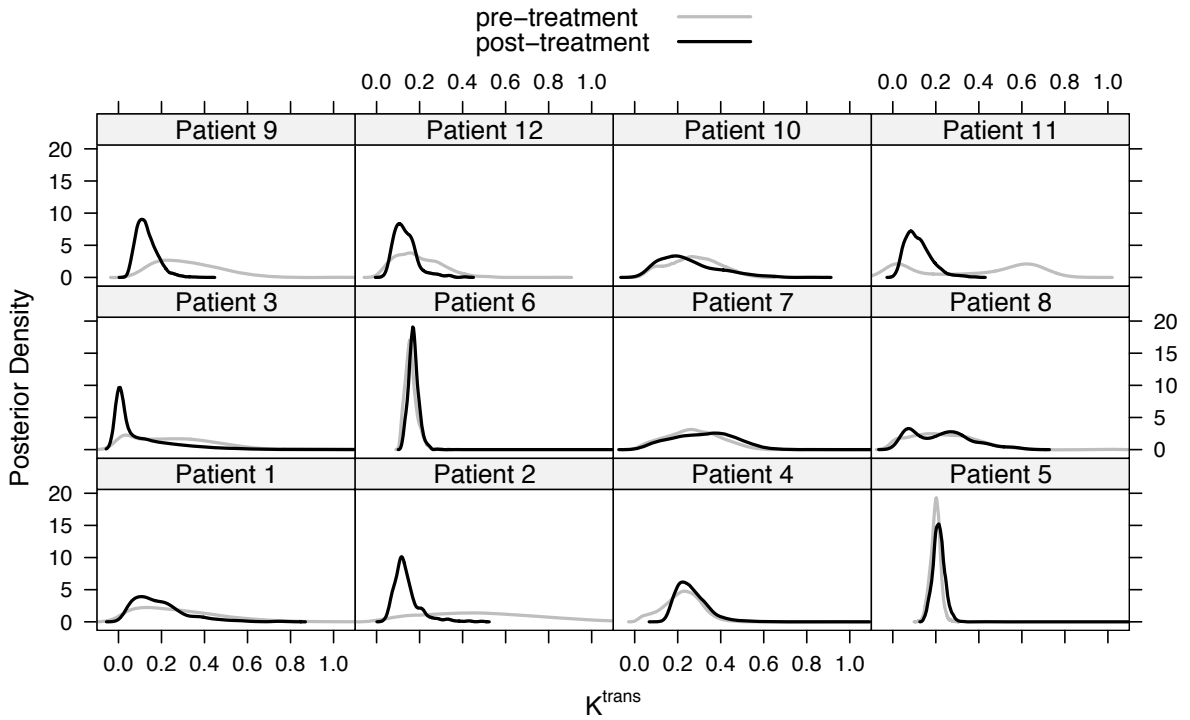


Figure 6:

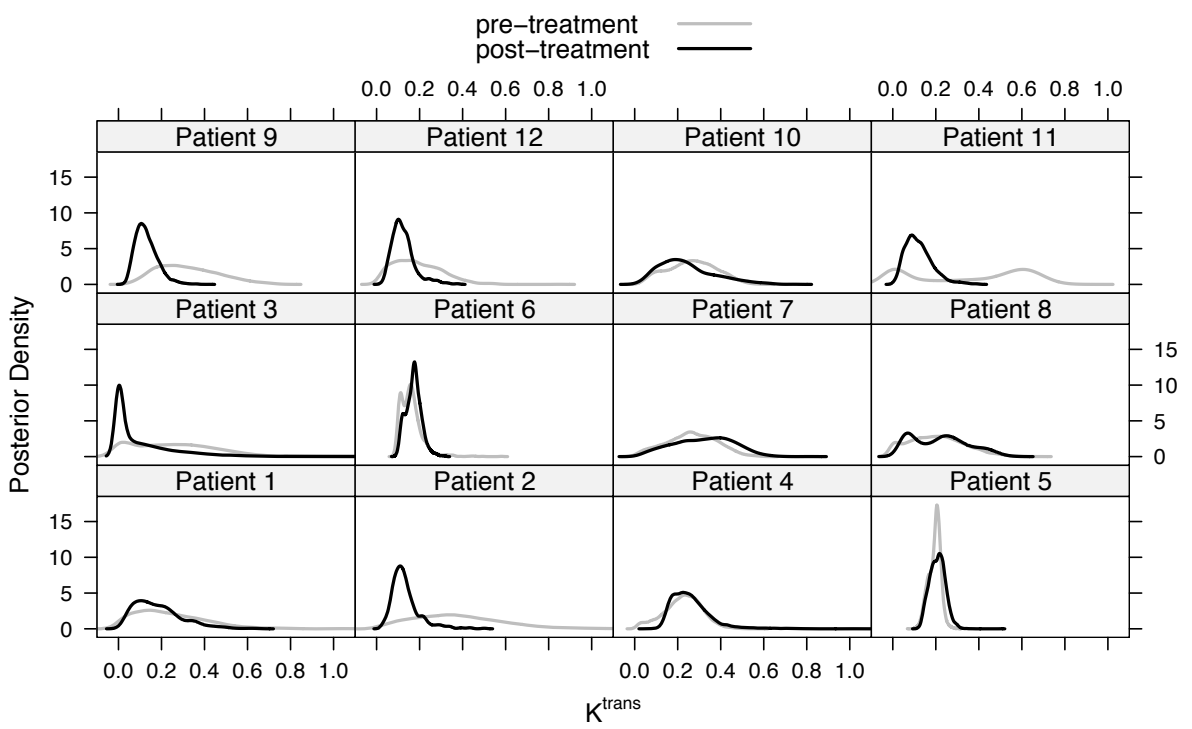


Figure 7:

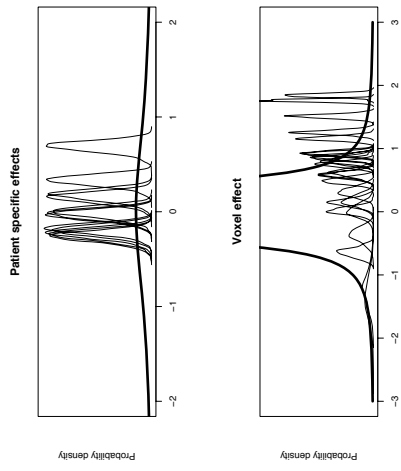


Figure 8: