Impact of Prior Distributions and Central Tendency Measures on Bayesian Intravoxel Incoherent Motion Model Fitting

Oscar Gustafsson[1,2], Mikael Montelius[1], Göran Starck[1,2], Maria Ljungberg[1,2]

[1]Department of Radiation Physics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.

[2]Department of Medical Physics and Biomedical Engineering, Sahlgrenska University Hospital, Gothenburg, Sweden

Word count: 4574

Running title: Impact of Prior and Central Tendency Measure on Bayesian IVIM Model Fitting

Key words: intravoxel incoherent motion; Bayesian estimation; prior distribution; central tendency measure

Corresponding author:

Oscar Gustafsson

MRI center

Sahlgrenska University Hospital

Bruna stråket 13, 413 45 Göteborg, Sweden

Email: oscar.gustafsson@gu.se

Abstract

Purpose: Bayesian model fitting has been proposed as a robust alternative for IVIM parameter estimation. However, consensus regarding choice of prior distribution and posterior distribution central tendency measure is needed. The aim of this study was to compare the quality of IVIM parameter estimates produced by different prior distributions and central tendency measures, and to gain knowledge about the effect of these choices.

Methods: Three prior distributions (uniform, reciprocal and lognormal) and two measures of central tendency (mean and mode) found in the literature were studied using simulations and in vivo data from a tumor mouse model.

Results: Simulations showed that the uniform and lognormal priors were superior to the reciprocal prior, especially for the parameters *D* and *f* and clinically relevant SNR levels. The choice of central tendency measure had less effect on the results, but had some effects on estimation bias. Results based on simulations and in vivo data agreed well, indicating high validity of the simulations.

Conclusion: Choice of prior distribution and central tendency measure affects the results of Bayesian IVIM parameter estimates. This must be considered when comparing results from different studies. The best overall quality of IVIM parameter estimates was obtained using the lognormal prior.

Key words: intravoxel incoherent motion; Bayesian estimation; prior distribution; central tendency measure

Introduction

Intravoxel incoherent motion (IVIM) imaging has gained increased interest during the last years due to its applicability in abdominal imaging (1-3). It enables both diffusion and perfusion information to be extracted from MR images completely non-invasively. Commonly, IVIM is described by a biexponential model:

$$S_i = S_0\left((1-f)e^{-b_iD} + fe^{-b_iD^*}\right) + \varepsilon_i \quad [1]$$

where S_i is the measured signal using the ith b-value, D is the diffusion coefficient, D^* is the pseudo diffusion coefficient, f is the perfusion fraction, S_0 is the signal intensity without diffusion weighting and ε_i is a random deviation from the model following some appropriate distribution, often assumed to be Gaussian (2). Since the signal attenuation due to perfusion effects is separated from attenuation due to diffusion in the model, the diffusion coefficient obtained from IVIM model fitting is less dependent on the choice of b-values as long as the effect of the dispersion of the diffusion coefficient at high b-values is avoided (4, 5). However, biexponential model fitting is significantly more difficult than monoexponential fitting, as the former cannot be transformed into a linear model and since it may have multiple, local, optima. Furthermore, in the limit as the perfusion fraction approaches zero, the biexponential model transforms into a monoexponential model and D^* cannot be estimated, which may lead to unstable fitting results.

The difficulties concerning the fitting of the IVIM model has resulted in several studies comparing model-fitting strategies (6-15). Most of these studies compare different least-squares methods, but a Bayesian approach was early shown to be a robust alternative (12). This was recently reported also by Barbieri et al. in a comparison between most of the IVIM model fitting strategies found in the literature (6), although the Bayesian approaches used in these studies were slightly different.

The aim of a Bayesian IVIM model fit is to estimate the joint posterior distribution from which the marginal posterior distributions of the IVIM model parameters of interest, i.e. $P(D|S)$, $P(D^*|S)$ and $P(f|S)$ can be derived. The joint posterior distribution is given by Bayes' rule:

$$P(D, D^*, f, S_0|S) \propto P(S|D, D^*, f, S_0) \cdot P(D, D^*, f, S_0) \quad [2]$$

where S is the measured data, $P(S|D,D^*,f,S_0)$ is the likelihood function and $P(D,D^*,f,S_0)$ is the joint prior distribution. Apart from choosing an appropriate likelihood function a prior distribution has to be chosen as well. Most previous studies involving Bayesian IVIM model fitting have used various non-informative or low-informative priors (6, 12, 16, 17), although some more advanced approaches have been proposed (13, 18). To make the result of a Bayesian fit more comprehensible, the

marginalized probability distributions need to be summarized, most importantly in terms of central tendency to describe the center or location of the distribution, but possibly also in terms of width and skewness. In previous studies the mode (6, 12) or mean (13, 16) have been used to describe the central tendency. Estimates of the mode and mean of the marginal posterior distributions are also commonly referred to as the marginal maximum a posteriori (MMAP) estimates and the minimum mean square error (MMSE) estimates respectively.

Numerous studies concerning IVIM model fitting have been performed, and the Bayesian approach has shown great promise. Yet none of the previous studies has, to our knowledge, aimed to assess the impact of the methodology used in the Bayesian model fitting, most importantly including the choices of prior distribution and central tendency measure, which often differ between studies (6, 12, 16, 17). These choices may impact parameter estimation performance, especially when noise limits the information available to the model fitting. Potential estimation bias due to these choices must be studied and taken into account when comparing results of studies with different Bayesian model-fitting approaches.

The aim of this study was to evaluate the implications of the choice of prior distribution and posterior distribution central tendency measure on the IVIM parameters using simulations and in vivo data.

Methods

An IVIM experiment setup with 12 b-values (1.4, 5, 10, 20, 35, 50, 75, 100, 201, 401, 602 and 802 s/mm^2) was used and data were acquired both through simulations and in vivo measurements for subsequent analysis. The simulations enable separate analysis of bias and variability of the estimated model parameters. However, the simulations are only of interest if they are a good representation of the reality. If results based on simulations and in vivo measurements agree well, this warrants further conclusions to be drawn from the results based on the simulations. Therefore, the in vivo data was analyzed and compared with results from the simulations.

Simulations

Artificial data were generated by Monte Carlo simulations using the IVIM model (Eq. 1) with certain sets of tissue parameters and noise levels. 10,000 data series were generated for each noise level and white Gaussian noise was added independently to the real and imaginary part, followed by calculation of the absolute value. The standard deviation of the noise in the real and imaginary parts was equal and chosen such that the signal-to-noise ratio (SNR) levels before diffusion weighting were 10, 20 and 40. The SNR was defined as the simulated S_0 divided by the standard deviation of the

Gaussian noise. The tissue parameters used in the simulations were randomly generated from bounded uniform distributions except for S_0, which was set to 100. Bounds for the remaining tissue parameters were *D*: [0.5 1.5] μm^2/ms, *D**: [10 100] μm^2/ms and *f*: [0 0.3].

In vivo animal study

In vivo data were acquired from a tumor mouse model. Five female BALB/c nude mice (Charles River, Japan and Germany), with subcutaneous xenografts of the human midgut carcinoid cell line GOT1 transplanted into the neck region, were subjected to MR imaging when the tumor diameters were approximately 15 mm. The study was approved by the ethical committee on animal research at the University of Gothenburg, Gothenburg, Sweden.

A horizontal-bore 7T system (Bruker BioSpin 70/20AS MRI GmbH, Ettlingen, Germany; software: ParaVision 5.1), equipped with a maximum 400 mT/m gradient system, a 72 mm volume transmit coil, and an actively decoupled 4-channel array rat brain receiver coil (RAPID Biomedical GmbH, Rimpar, Germany) was used for image acquisition. Diffusion-weighted images were acquired with a spin echo-echo planar imaging (SE-EPI) method based on the Stejskal-Tanner pulse sequence (Δ = 9 ms, δ = 4 ms), with three orthogonal gradient directions and the b-values listed above. Other imaging parameters were: TR = 1500 ms, TE = 22 ms, number of signal averages = 3, number of segments = 1, pixel size = 320^2 μm^2, slice thickness = 1000 μm, slice gap = 500 μm, partial Fourier acceleration = 1.5, EPI echo spacing = 0.3 ms. The field of view (FOV) included the tumor and only very small amounts of other tissues. The animal was anaesthetized and fixed in supine position on a custom made plastic cradle during image acquisition, and the tumor was immobilized in a cut out hole in the cradle to avoid motion artefacts. Total scan time was approximately 3 minutes. A 2x2 in-plane median filter was applied to reduce the effects of residual motion.

The noise level in the in vivo images was estimate by calculating the square root of the mean squared error from a voxelwise monoexponential non-linear least squares fit of images with b-values higher than or equal to 200 s/mm^2. The SNR in the images with the smallest b-value was 15-25 in the tumor tissue.

Model fitting

By assuming white Gaussian noise the likelihood function for the IVIM model is given by:

$$P(S|D,D^*,f,S_0,\sigma) = (2\pi\sigma^2)^{-n/2}\exp\left\{\frac{1}{2\sigma^2}\sum_{i=1}^{n}(S_i - S_0\left((1-f)e^{-b_iD} + fe^{-b_iD^*}\right)^2\right\} \quad [3]$$

where σ is the standard deviation of the noise and n is the number of b-values. Using a reciprocal prior on the noise parameter σ, an analytical marginalization was performed, as described by Bretthorst et al. (19), to yield the likelihood function:

$$P(S|D,D^*,f,S_0) \propto \left[\frac{1}{2}\sum_{i=1}^{n}(S_i - S_0\left((1-f)e^{-b_iD} + fe^{-b_iD^*}\right)^2\right]^{-n/2} \quad [4]$$

While the Gaussian likelihood function is the one most commonly used, a Rician distribution would be more appropriate when magnitude data is used and the SNR is low (20). The likelihood function based on Rician noise is given by:

$$P(S|D,D^*,f,S_0,\sigma) = \frac{S_i}{\sigma^2}\exp\left\{-\frac{S_i^2+S(b_i)^2}{2\sigma^2}\right\} I_0\left(\frac{S_iS(b_i)}{\sigma^2}\right) \quad [5]$$

where I_0 is the modified zeroth order Bessel function of the first kind (20). When using the same reciprocal prior on the noise parameter σ as for the Gaussian likelihood (eq. 4), an analytical marginalization is not available. Instead, σ was marginalized numerically as described below for the IVIM model parameters.

In order to be able to estimate the marginal posterior distributions of the model parameters (*D, D*, f* and S_0) a joint prior distribution P(*D, D*, f*,S_0) must be chosen. Focusing on the simple priors used in most published studies, three distinguished kinds of priors where identified in the literature and included in this study. The joint prior was given by the product of the individual marginal priors with the additional constraint D < D*. This constraint was implemented by setting the joint prior distribution to zero where D* < D. The marginal priors were bounded uniform distributions for *f* and S_0, while the distributions over *D* and *D** were either bounded uniform (denoted 1 in figures) (6, 17), reciprocal (1/x) (12) or lognormal (logN) (16). The reciprocal prior is described by:

$$P(\theta) \propto 1/\theta \quad [5]$$

whereas the lognormal distribution is described by:

$$P(\theta|\mu_\theta,\sigma_\theta) = \frac{1}{\theta\sigma_\theta\sqrt{2\pi}}\exp\left\{-\frac{(\ln\theta-\mu_\theta)^2}{2\sigma_\theta^2}\right\} \quad [6]$$

where μ_θ and σ_θ are the mean and standard deviation of the natural logarithm of the random variable Θ. The lognormal priors for *D* and *D** both had $\sigma_\theta = 1$, while μ_θ was -6 for *D* and -3.5 for *D** with units translating into mm^2/s.

The marginal posterior distributions were estimated using a Markov Chain Monte Carlo (MCMC) setup based on Gibbs sampling and the Metropolis-Hastings algorithm, similar to the voxelwise

method described by Orton et al. (13). The procedure was performed for each data series (simulations) or voxel (in vivo experiment) separately. The start values for the MCMC algorithm were the true model parameter values in the simulations and based on a two-step fit for the in vivo experiments (6). The step length parameters used in the MCMC procedure were updated every 100th iteration during the first 2000 iterations to reach a level where approximately 50 % of the samples were accepted, to improve convergence rate. All step length parameters were initialized as one tenth of the start value of the corresponding model parameter. The initial magnitude of the step length will only affect the convergence speed and is thus not important as long as the number of iterations is large enough. Another 2000 iterations were run before the start of the sampling which lasted 40,000 iterations. By studying the results from different chains based on the same data, this was found to be sufficient to assure convergence. For each kind of prior distribution, parameter estimates were obtained by calculation of the mean and mode of the resulting marginalized parameter distributions. The mode was calculated using the half sample mode method (21). All prior distributions were set to zero outside the limits given by D: [0 5] µm^2/ms, f: [0 1], D^*: [0 1000] µm^2/ms and S_0: [0 2S_{max}], where S_{max} is the overall maximum measured or simulated signal depending on the context. For the simulated data parameter estimation was performed using both the Gaussian likelihood (eq. 4) and the Rician likelihood (eq. 5), whereas for the in vivo data parameter estimation was only based on the Gaussian likelihood. MATLAB R2014b (The MathWorks, Natick, MA, USA) was used for all simulations and model fitting.

Evaluation of estimation methods

Parameter estimation error, defined as estimated value minus simulated value, was used to assess the bias and variability of parameter estimates based on simulated data. For in vivo data, parameter estimates and local standard deviation, calculated in 3×3 neighborhoods around each pixel in the parameter maps, were used to analyze relative differences in estimation bias and variability between methods.

Results

Representative data sets from simulations and in vivo examinations were chosen to illustrate the impact of different prior distribution and central tendency measures. Individual data points and signal intensity curves derived from Eq. 1 using the estimated parameter values based on the Gaussian likelihood (Eq. 4) can be seen in Figure 1. Figure 2 shows the corresponding prior and posterior parameter distributions based on the same data. Parameter maps of a representative tumor chosen from the five tumors included in the study are shown in Figure 3. These examples show that the uniform and lognormal priors resulted in similar estimates of D and f, but produced

different estimates of *D** where the marginal posterior distribution was highly affected by the prior. Estimations based on the reciprocal prior were much more sensitive to noise, as demonstrated for example in the posterior distribution of *D* (Fig. 2), which is dominated by the information in data only for the highest SNR level. In contrast, the uniform and lognormal priors resulted in distinct posterior distributions of *D* even at the lowest SNR level. Parameter maps of all other slices in tumor number one and also the remaining four tumors show similar results as those seen in Figure 3 and can be found in Supporting Figures S1-S34.

The difference between estimation with the Gaussian likelihood function (Eq. 4) and the Rician one (Eq. 5) was in general small. The only clear trend was a slightly larger bias and variability with the Rician likelihood for *D* at large values of *f* and *D*, when combining the uniform prior and the mean as central tendency measure. A detailed comparison of the bias and variability of the two likelihood functions can be found in Supporting Figures S35-S52. All results presented below refer to estimation using the Gaussian likelihood.

Parameter estimation errors are presented as boxplots in Figure 4. It shows that the estimation error of *D* and *f* is similar using the uniform or lognormal prior and substantially smaller compared with the error when using the reciprocal prior. The choice of central tendency measure did in some cases tend to affect either bias or variability. This is for example seen for *D** and the uniform prior where the mode resulted in smaller bias but larger variability compared with the mean.

As a result of the range of simulated values and the fixed limits of the priors, the maximum negative and positive errors depend on the simulated parameter values. Figure 5 shows how the error of *D* depends on the simulated parameter values for the reciprocal prior and with mode as central tendency measure, while other combinations of parameter, prior and central tendency measure are found in Supporting Figures S53-S70. One can see in the leftmost column of Figure 5 that a majority of the estimates of *D* are fixed at the lower limit for SNR = 10 and 20. This effect cannot be visualized in an ordinary boxplot (Fig. 4 and right column in Fig. 5). To quantify the dependence of the estimation error on simulated values Spearman correlation coefficients were calculated (Fig. 6 for SNR = 20 and Supporting Fig. S71 and S72 for SNR 10 and 40 respectively). Strong negative correlations are seen, for example, between the error in *D** and simulated *f* for the uniform prior, as a result of the decreasing estimation bias of *D** as *f* increases (Supporting Figs. S65 and S66), and between the error in D* and simulated *D** for the reciprocal prior, as a result of estimation at the lower limit similar to that seen for *D* in Figure 5 (Supporting Figs. S67 and S68). Negative correlation was also found between error in *f* and simulated *f* for all prior distributions using the mean as central

tendency measure. This correlation was substantially smaller for the uniform and lognormal priors when the mode was used as central tendency measure (Fig. 6 and Supporting Figs. S53-S58).

The tumor median parameter estimates for all five tumors are shown in Figure 7. A high degree of similarity between the median error derived from simulated data and median estimated value from in vivo data was seen (compare Figures 4 and 7). When comparing in vivo estimation variability (Fig. 8) with the estimation variability from simulations (interquartile range seen in Figure 4) a good agreement was seen for most methods.

Discussion

In this study we have explored the characteristics of Bayesian IVIM model fitting for three prior distributions (uniform, reciprocal and lognormal) and two distribution central tendency measures (mean and mode) found in the literature. The choice of prior distribution was found to have a major impact on the resulting parameter estimates, whereas the central tendency measure appeared to have a less pronounced role. Results from simulations were in agreement with the results from in vivo data, indicating high validity of the simulations.

The median estimation error derived from simulations and median estimated parameter value from in vivo data showed similar dependencies on choice of prior and central tendency measure (Figs. 4 and 7). For an appropriate simulation this is to be anticipated since the estimated value is the true value, which is the same for all estimation strategies, plus the estimation error. Also the variability of the estimation error derived from simulations and the local variability of estimated parameter values from in vivo data showed similarities. Although the use of standard deviation in small neighborhoods is a less straight forward approach it gives a rough estimate of the estimation variability under the assumption that the variability of the true tissue parameter value from pixel to pixel is small. Due to the good correspondence between results based on simulations and in vivo data conclusions based on simulations are likely to apply also to analysis of in vivo data.

The choice of prior distribution is an important choice since the prior distribution combined with the likelihood function gives the posterior distribution that is used in Bayesian model fitting (Eq. 2). The choice of prior distribution is even more critical at low SNR, which is common in diffusion-weighted imaging. Comparisons of estimates based on the different priors showed that the uniform and lognormal priors resulted in similar estimates of D and f, while the reciprocal prior dominated the posterior distribution, yielding estimates of low quality. A closer inspection of Figure 4 reveals that the bias imposed by the priors is different for D and f when comparing the uniform and lognormal priors, however, the choice is mostly a question of taste. The uniform prior is less informative and

may therefore be considered more attractive, whereas the bias appears to be slightly smaller using the lognormal prior. On the other hand, the estimation of *D** is of substantially higher quality using the lognormal prior compared with the other ones. It is especially seen for lower values of *D** and when the mean is used to summarize the marginal posterior distribution. This is reasonable since the lognormal prior on *D** is fairly informative and has most of its density in the region 3-40 $\mu m^2/ms$ (>50 % of the peak value of the distribution in this range), while the mean is able to include information in the right tail of the distribution thereby possibly reducing the bias for high values of *D**. Furthermore, when visually inspected, the *D** maps obtained using the lognormal prior show an apparent good quality with pattern and characteristics not clearly seen in the other parameter maps. Due to the informative prior the estimates may be strongly biased, but may still prove interesting in for example comparisons between groups or longitudinal analyses. The parameter estimation using reciprocal priors was in general poor, but for high SNR (40 in this study or higher) the performance seems to move towards that of the other priors. The good performance reported by Neil et al. (12) is thus limited to examinations with SNR substantially higher than what is commonly seen in diffusion-weighted MRI.

The choice of central tendency measure has no impact on the parameter estimate as long as the marginal posterior distribution is symmetric. It is clear, however, from Figure 2 that there are cases with strongly skewed distributions. Choosing the most suitable central tendency measure could have a substantial influence on the quality of the results in these cases. Almost all of the marginal posteriors of *f* seen in Figure 2 are rightly skewed, resulting in smaller parameter estimates using the mode compared with using the mean as central tendency measure. This difference is seen in Figure 4 which shows that using the mean imposes a positive bias on the estimate of *f*. The effect of skewed distributions are also seen for *D** where the estimates based on a uniform prior have smaller bias using the mode compared with using the mean. The opposite was seen when using a lognormal prior. The overall tendency was that using the mode instead of the mean resulted in reduced bias but slightly higher variability of the parameter estimates. For *D* and *f* it was also seen that the mode resulted in reduced dependencies between the error and the true parameter values (Fig. 6).

The clinical goal is the ability to differentiate between different tissues, for example tumor and normal tissue, or to be able to assure if parameters change in time e.g. after treatment. To be able to accomplish this using model parameter estimates will depend on the variability of parameter values within the groups and the difference between groups. Since the difference between groups is reduced by a negative correlation between the parameter estimation error and the true parameter value, the correlation of the bias must be taken into account when choosing the model fitting approach for a study. Such a negative a correlation was found for *f* using the mean as central

tendency measure, but not when the mode was used in combination with a uniform or lognormal prior (Fig. 6). The mode may thus be preferable in that case. However, the optimal choice will depend on sample size and statistical method. The Gaussian and Rician likelihood functions were found to give similar results. For high SNR this is expected since the Rician distribution then is well approximated by a Gaussian (20). However, at low SNR (< 5) the signal is substantially positively biased, especially at SNR < 2. Such SNR levels are only found for the highest b-value in this study (800 s/mm^2) and the most attenuating IVIM model parameter combinations (in our simulations D = 1.5 μm^2/ms and f = 0.3 respectively). The gain in estimation performance when using a Rician likelihood function would thus most likely be largest for large values of D and f. However, no such trend was seen in the simulations (Supporting Figures S35-S52). Instead an increased bias and variability was seen when using the Rician likelihood for *D* at large values of *f* and *D* for the combination of uniform prior and mean as central tendency measure. A possible explanation of this result may be that the combination of a Rician likelihood and a uniform prior is too flexible and thereby more sensitive to noise. Based on these results, estimation using the Gaussian likelihood is preferable since it is less complicated and appears to yield estimates with similar or smaller bias and variability. However, it should be noted that the effect of Rician noise may be more pronounced with fewer or higher b-values where the influence of single, high b-values would be larger.

The results in this study indicate that an informative prior is needed for stable estimation of *D**. Using the voxelwise priors that have been studied here, this can only be achieved by choosing priors with a narrow shape. However, recently more advanced priors, which exploits the information of neighboring voxels or voxels in a region of interest, have been proposed (13, 18). The benefit with these approaches is that the narrowness and location of the prior is deduced from the data itself instead of a priori knowledge. However, while the results are promising, assumptions regarding the joint prior distribution of model parameters across voxels need to be made and a recent study showed that these approaches may give unwanted results such as disappearing structures (22). Further studies are thus needed to improve and validate these methods.

The results of this study show that different prior distributions and distribution central tendency measures profoundly affect the quality of Bayesian IVIM model fitting, and that these differences must be considered when comparing the results of studies using different model fitting strategies. Our results show that the same data may give very different results even though the quality of the resulting parameter maps appears satisfactory with seemingly acceptable noise levels (e.g. *D* when comparing uniform and lognormal priors in Fig 7). Therefore, comparative studies, such as this and others (6-15), are important for better comparability between past and future studies, but also in

order to reach consensus regarding model-fitting strategies, which will increase comparability between future studies.

The need of a thorough analysis of estimation errors in simulation studies as this is apparent in Figure 5 and Supporting Figures S53-S70. This type of data is often visualized as boxplots; however, all important information cannot be visualized in a boxplot if a range of parameter values are simulated. This is clearly demonstrated if one compares columns 1-3 with column 4, where the graphs are equivalent to boxplots, in these figures. Depending on the simulated range this could lead to a misleading boxplot. Although more complicated, the graphs in the leftmost column in Figure 5, showing the error of *D* as a function of simulated *D,* gives substantially more information and should preferably be supplied as a complement to boxplots.

There are limitations in this study. First, only one set of b-values was used. It is, however, similar to what is used in many clinical IVIM studies (23). Second, only one in vivo setup was used, resulting in only one SNR level and one type of tissue being examined. Third, the studied central tendency measures and prior distributions were limited to those used in previous IVIM studies. Among other possible central tendency measures, the median is probably the most appealing. Compared with the mean, it is less sensitive to skewed distributions, while it is substantially easier to obtain from a sampled distribution compared with the mode. Furthermore, the lognormal and uniform priors may be altered by choosing their parameters differently. The parameters used for the lognormal prior in this study were based on the ones used by Dyvorne et al. (16) whereas the limits for the uniform priors are similar to those used by Barbieri et al. (6). These parameters give reasonable results, but might be optimized and also further studied.

Conclusions

At practically achievable SNR levels in IVIM imaging, the choice of prior distribution and distribution central tendency measure plays an important role in the performance of Bayesian IVIM model fitting. In particular, the choice of prior distribution was found to have a significant impact. Among the priors assessed in this study, the uniform and lognormal priors provided more stable estimates in both simulations and in vivo experiments. However, it should be noted that the bias imposed by the two priors were of opposite sign. The choice of central tendency measure was of minor importance for the estimation of *D* and *f*, but our results indicate that a proper choice could give an increased number of voxels where the estimation error of the *D** parameter was of acceptable magnitude, especially for the uniform prior. The lognormal prior had the best overall performance, although the uniform prior could be considered if only *D* and *f* are of interest and high objectivity is prioritized. The

actual choices of prior distribution and central tendency measure need to be taken into account when comparing results from different studies.

Acknowledgements

We are greatly thankful to Prof Eva Forssell-Aronsson and Prof Ola Nilsson for letting us use the GOT1 mouse model for the in vivo validation of the simulations. We also thank Lilian Karlsson and Ann Wikström for their skilled technical work with the animals as well as PhD Hadrien Dyvorne for supplying addition information about prior distributions used in their study.

This study was supported by research grants from the Swedish Cancer Society, the Swedish Research Council, the King Gustaf V Jubilee Clinic Cancer Research Foundation, the Sahlgrenska University Hospitals Research Foundations, the Assar Gabrielsson Foundation, the Wilhelm and Martina Lundgren science foundation, the Adlerbert student foundations and the Royal Society of Arts and Sciences in Gothenburg (KVVS).

References

1. Le Bihan D. Intravoxel Incoherent Motion Perfusion MR Imaging: A Wake-Up Call. Radiology 2008;249:748-752.
2. Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology 1988;168:497-505.
3. Luciani A, Vignaud A, Cavet M, Van Nhieu J, Mallat A, Ruel L, Laurrent A, Deux J, Brugieres P, Rahmoini A. Liver Cirrhosis: Intravoxel incoherent Motion MR Imaging - Pilot Study. Radiology 2008;249:891-899.
4. Iima M, Yano K, Kataoka M, Umehana M, Murata K, Kanao S, Togashi K, Le Bihan D. Quantitative Non-Gaussian Diffusion and Intravoxel Incoherent Motion Magnetic Resonance Imaging: Differentiation of Malignant and Benign Breast Lesions. Investigative radiology 2015;50:205-211.
5. Jensen JH, Helpern JA. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. NMR in biomedicine 2010;23:698-710.
6. Barbieri S, Donati OF, Froehlich JM, Thoeny HC. Impact of the calculation algorithm on biexponential fitting of diffusion-weighted MRI in upper abdominal organs. Magnetic resonance in medicine 2015;75:2175-2184.
7. Cho GY, Moy L, Zhang JL, Baete S, Lattanzi R, Moccaldi M, Babb JS, Kim S, Sodickson DK, Sigmund EE. Comparison of fitting methods and b-value sampling strategies for intravoxel incoherent motion in breast cancer. Magnetic resonance in medicine 2015;74:1077-1085.
8. Fusco R, Sansone M, Petrillo A. A comparison of fitting algorithms for diffusion-weighted MRI data analysis using an intravoxel incoherent motion model. Magma (New York, N.Y.) 2016;
9. Heusch P, Wittsack HJ, Pentang G, Buchbender C, Miese F, Schek J, Kropil P, Antoch G, Lanzman RS. Biexponential analysis of diffusion-weighted imaging: comparison of three different calculation methods in transplanted kidneys. Acta radiologica (Stockholm, Sweden : 1987) 2013;54:1210-1217.
10. Meeus EM, Novak J, Withey SB, Zarinabad N, Dehghani H, Peet AC. Evaluation of intravoxel incoherent motion fitting methods in low-perfused tissue. Journal of magnetic resonance imaging : JMRI 2016;
11. Merisaari H, Movahedi P, Perez IM, et al. Fitting methods for intravoxel incoherent motion imaging of prostate cancer on region of interest level: Repeatability and gleason score prediction. Magnetic resonance in medicine 2016;
12. Neil JJ, Bretthorst GL. On the use of Bayesian probability theory for analysis of exponential decay data: an example taken from intravoxel incoherent motion experiments. Magnetic resonance in medicine 1993;29:642-647.
13. Orton MR, Collins DJ, Koh DM, Leach MO. Improved intravoxel incoherent motion analysis of diffusion weighted imaging by data driven Bayesian modeling. Magnetic resonance in medicine 2014;71:411-420.
14. Suo S, Lin N, Wang H, Zhang L, Wang R, Zhang S, Hua J, Xu J. Intravoxel incoherent motion diffusion-weighted MR imaging of breast cancer at 3.0 tesla: Comparison of different curve-fitting methods. Journal of magnetic resonance imaging : JMRI 2015;42:362-370.
15. Park HJ, Sung YS, Lee SS, Lee Y, Cheong H, Kim YJ, Lee MG. Intravoxel incoherent motion diffusion-weighted MRI of the abdomen: The effect of fitting algorithms on the accuracy and reliability of the parameters. Journal of magnetic resonance imaging : JMRI 2016;
16. Dyvorne HA, Galea N, Nevers T, et al. Diffusion-weighted imaging of the liver with multiple b values: effect of diffusion gradient polarity and breathing acquisition on image quality and intravoxel incoherent motion parameters--a pilot study. Radiology 2013;266:920-929.

17. Rydhög AS, Van Osch MJ, Lindgren E, Nilsson M, Lätt J, Ståhlberg F, Wirestam R, Knutsson L. Intravoxel incoherent motion (IVIM) imaging at different magnetic field strengths: what is feasible? Magnetic resonance imaging 2014;32:1247-1258.
18. Freiman M, Perez-Rossello JM, Callahan MJ, Voss SD, Ecklund K, Mulkern RV, Warfield SK. Reliable estimation of incoherent motion parametric maps from diffusion-weighted MRI using fusion bootstrap moves. Medical image analysis 2013;17:325-336.
19. Bretthorst GL, Hutton WC, Garbow JR, Ackerman JJH. Exponential Parameter Estimation (in NMR) Using Bayesian Probability Theory. Concepts in Magnetic Resonance Part A 2005;27A:55-63.
20. Gudbjartsson H, Patz S. The rician distribution of noisy mri data. Magnetic resonance in medicine 1995;34:910-914.
21. Bickel DR, Frühwirth R. On a fast, robust estimator of the mode: Comparisons to other robust estimators with applications. Computational Statistics & Data Analysis 2006;50:3500-3530.
22. While PT. A comparative simulation study of bayesian fitting approaches to intravoxel incoherent motion modeling in diffusion-weighted MRI. Magnetic resonance in medicine 2017;
23. Lemke A, Stieltjes B, Schad LR, Laun FB. Toward an optimal distribution of b values for intravoxel incoherent motion imaging. Magnetic resonance imaging 2011;29:766-776.

Figures

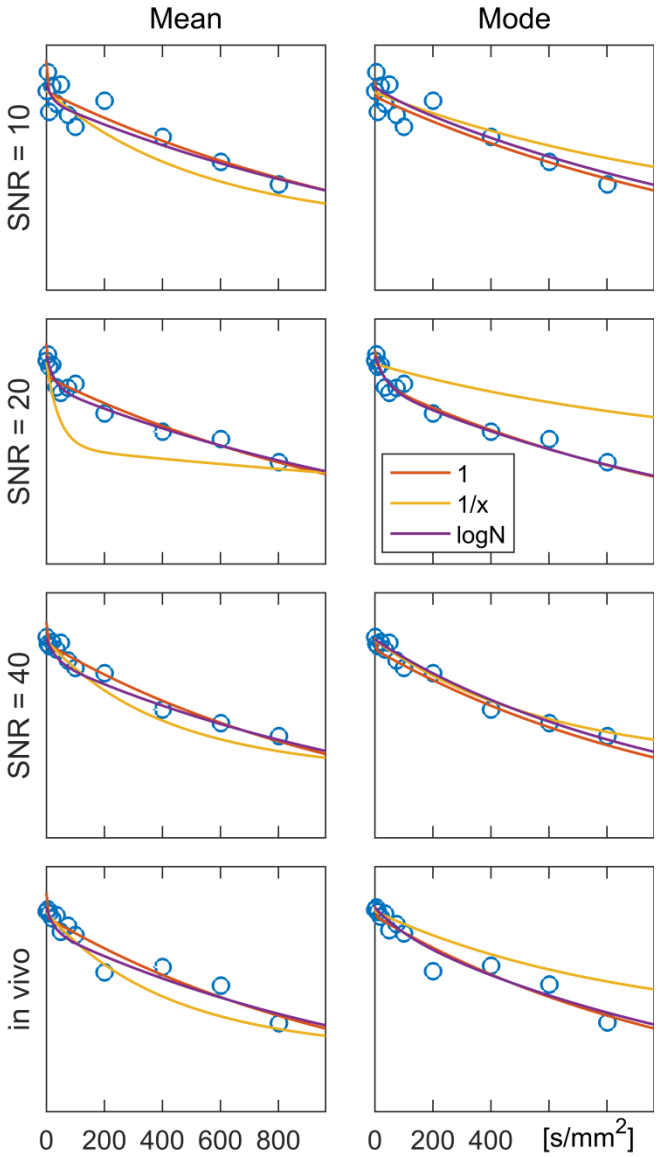

Figure 1. Signal vs. b-value for both simulated (a single noise realization) and in vivo data (single voxel from tumor 1 seen in Figure 3, 6 and 7). Simulated values were D = 0.7 µm^2/ms, D^* = 20 µm^2/ms and f = 0.1. Using uniform or lognormal priors, the estimates of D and f based on in vivo data were similar to the simulated ones. Signal curves were calculated from parameter estimates using the Gaussian likelihood (Eq. 4) and all combinations of prior and central tendency measure through insertion into Equation 1

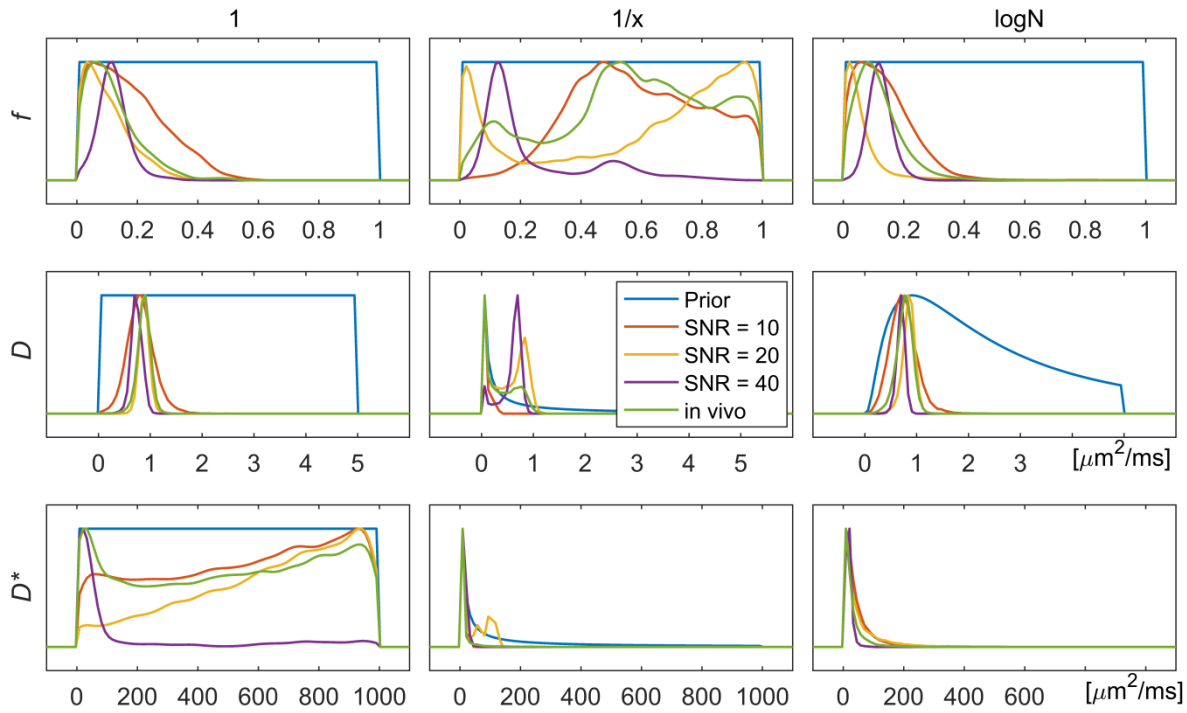

Figure 2. Prior and posterior distributions, using the Gaussian likelihood (Eq. 4), for all IVIM model parameters and priors based on the same simulated and in vivo data as in Figure 1. Simulated values were D = 0.7 µm^2/ms, D^* = 20 µm^2/ms and f = 0.1. Using uniform or lognormal priors, the estimates of D and f based on in vivo data were similar to the simulated ones. The maximum height of the distributions was normalized in the figure for better visibility. Kernel density estimation with a Gaussian kernel was used to illustrate the posterior distributions

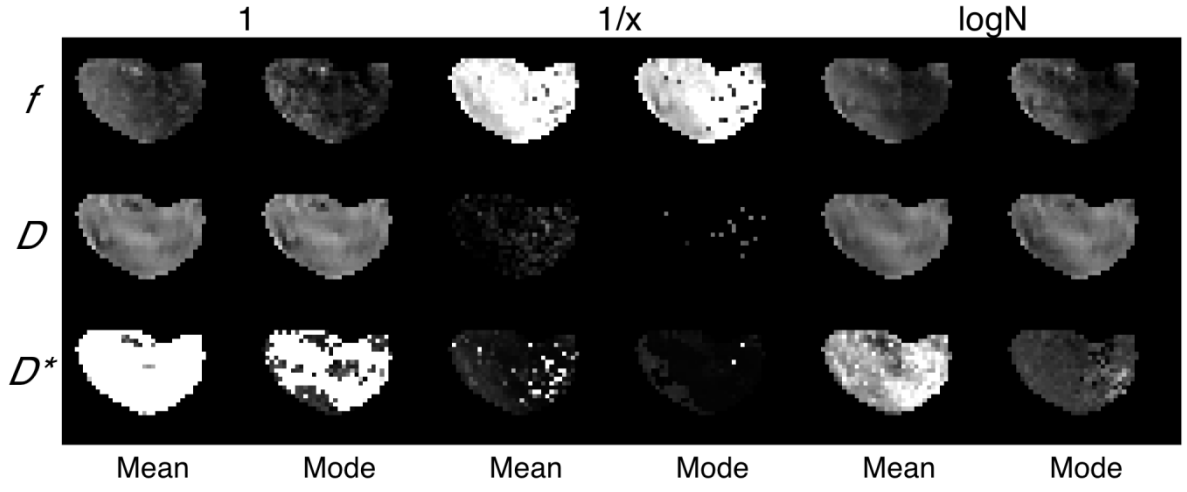

Figure 3. IVIM parameter maps, estimated with a Gaussian likelihood function (Eq. 4) and all combinations of prior and central tendency measure, of the central slice in a representative tumor (tumor 1 in Figures 7 and 8). The display ranges are D: [0 1.5] µm^2/ms, D^*: [0 50] µm^2/ms and f: [0

0.5]. A mask based on manual delineation of the tumor has been applied so that only tumor tissue is visible in the maps

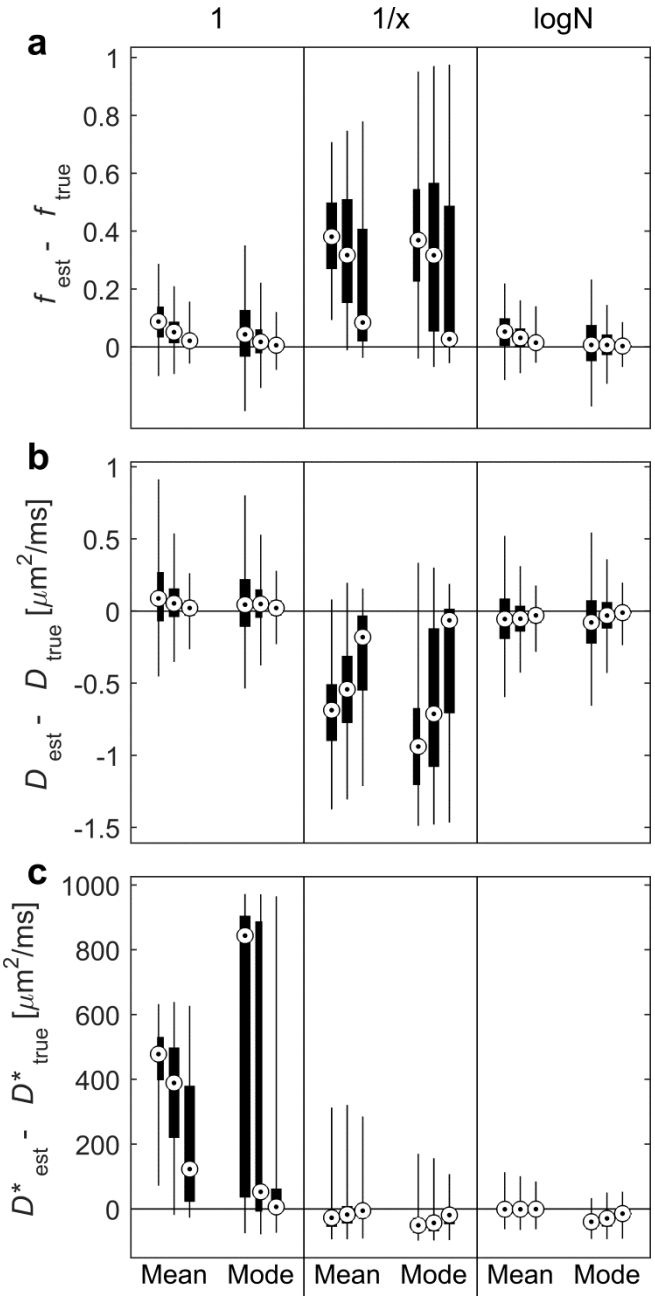

Figure 4. Parameter estimation error for all combinations of prior, central tendency measure and SNR based on simulated data, for a) *f*, b) *D* and c) *D** respectively. The three tightly clustered boxplots represent, from left to right, SNR = 10, 20 and 40. Each boxplot shows the median (dot in circle), the 25th and 75th percentile (lower and upper limit of the box), and the 1st and 99th percentile (endpoints of the whiskers). The horizontal black lines indicate an error equal to zero

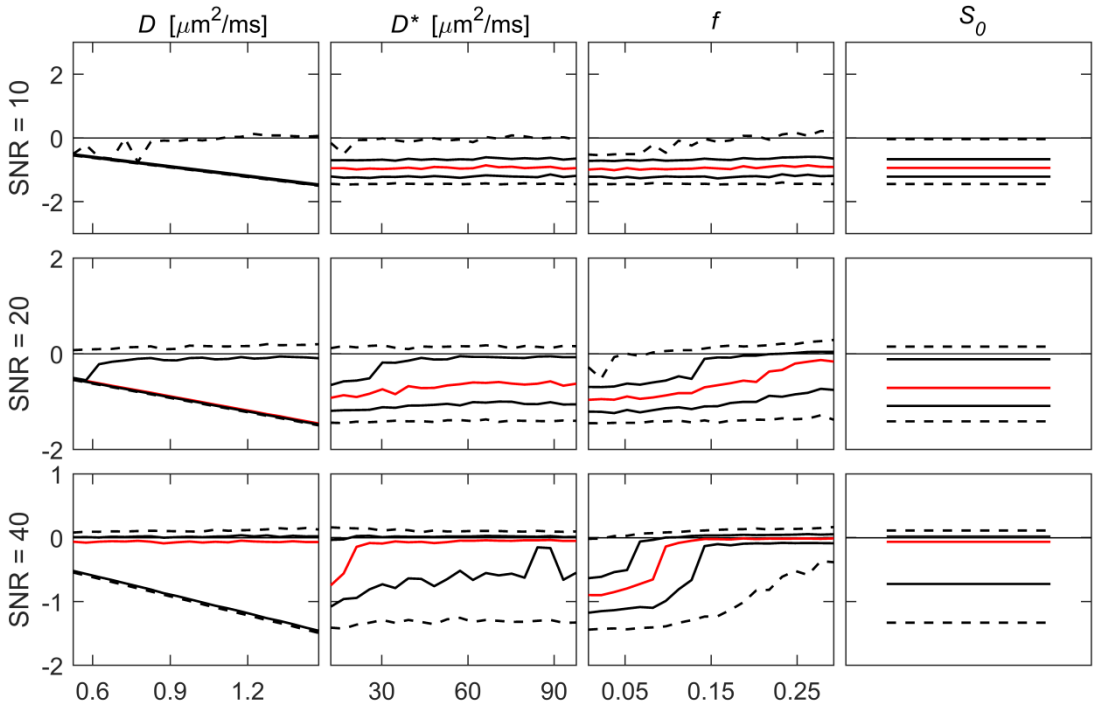

Figure 5. Parameter estimation error for *D,* using a reciprocal prior and mode as central tendency measure, plotted as a function of simulated parameter values for all SNR levels. The error was defined as estimated value minus simulated value. Errors are summarized by the 1st, 25th, 50th, 75th and 99th percentiles (bottom dashed black line to top dashed black line respectively). The horizontal thin black lines indicate an error equal to zero. Since only one value was simulated for S_0 the rightmost plot contains similar information as the boxplot in Figure 4. The almost straight lines with negative slope seen in the leftmost column and especially at low SNR is due to that a large proportion of the parameter estimates are at the lower limit. Since the distance from the lower limit increases as the simulated value increases, the magnitude of the maximum negative error increases. The variability in the errors seen in the columns 2-4 is thus mainly due to the varied simulated values. Note that this lack of information in the estimates is not easily seen in the other plots in the same row. This phenomenon can appear in plots of error of a certain parameter vs. the simulated value of the same parameter. See Supporting Figures S53-S70 for all other combinations of IVIM parameter, prior and central tendency measure

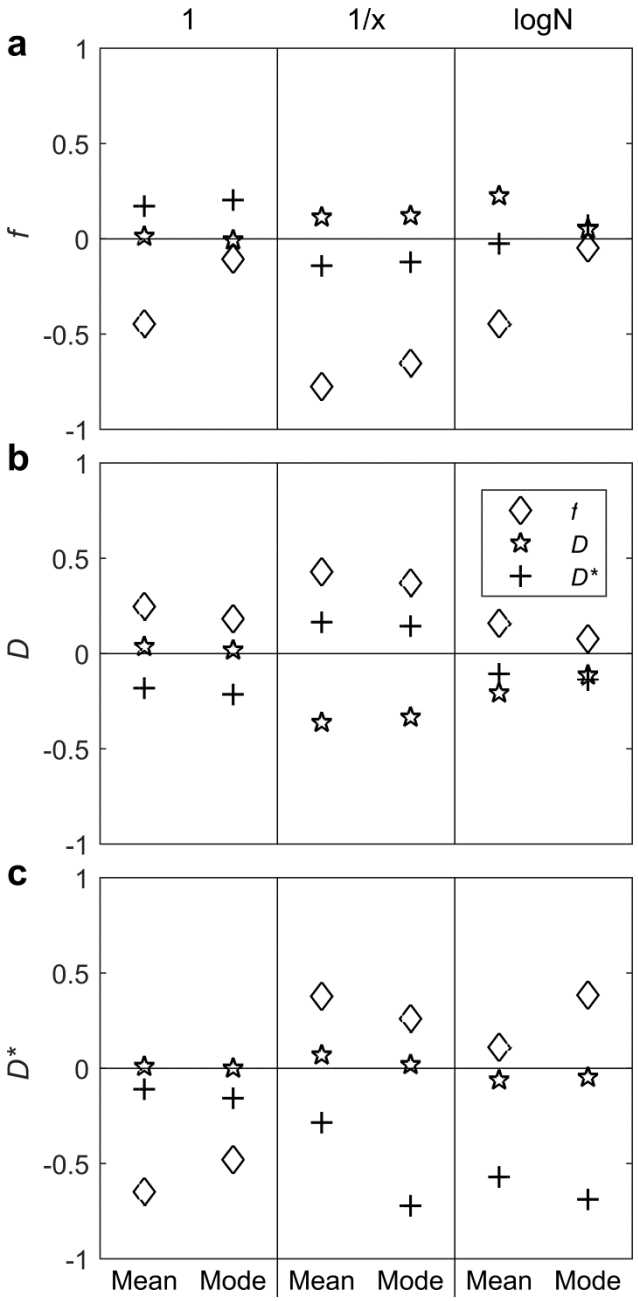

Figure 6. Spearman correlation between parameter estimation error (estimated value minus simulated value) and simulated parameter value for all combinations of prior and central tendency measure at SNR = 20, for a) *f*, b) *D* and c) *D** respectively. Strong correlations may for example appear in cases of consequent underestimation or if the estimation bias increases/decreases as a function of a parameter, which is seen in many cases when the dependence on *f* is studied (Fig. 5 and Supporting Figs. S53-S70)

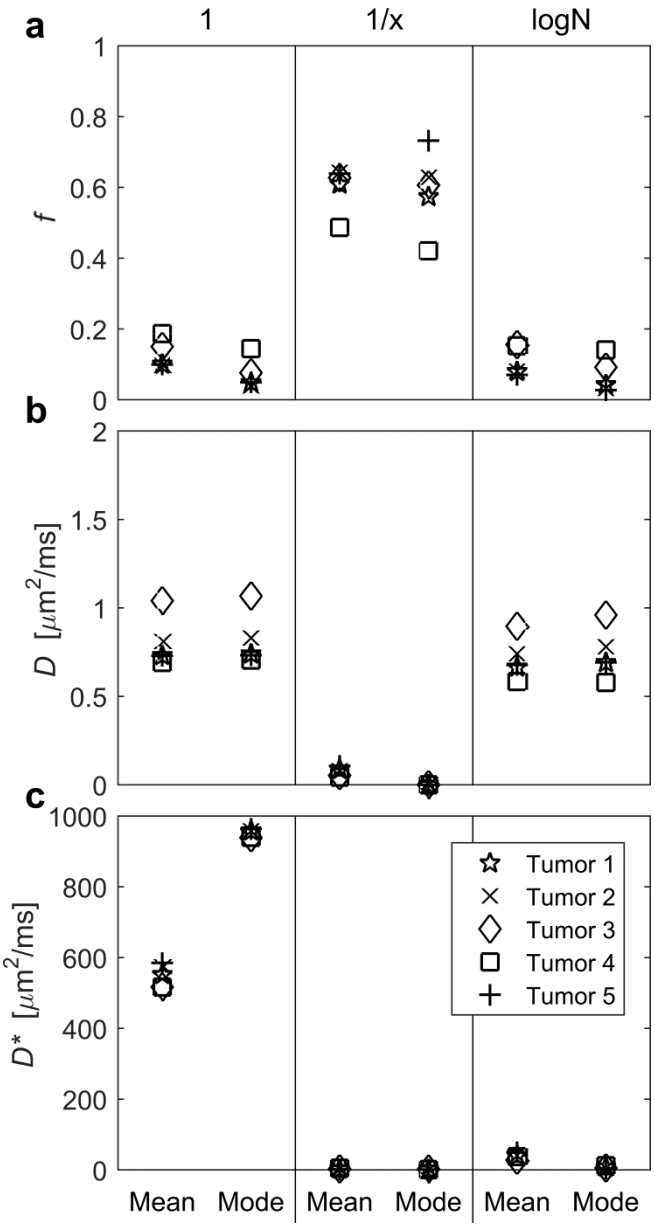

Figure 7. Tumor median parameter estimates for all combinations of prior and central tendency measure, for a) *f*, b) *D* and c) *D** respectively. IVIM parameter estimates from all slices included in the region defined as tumor were included in the calculation of the median

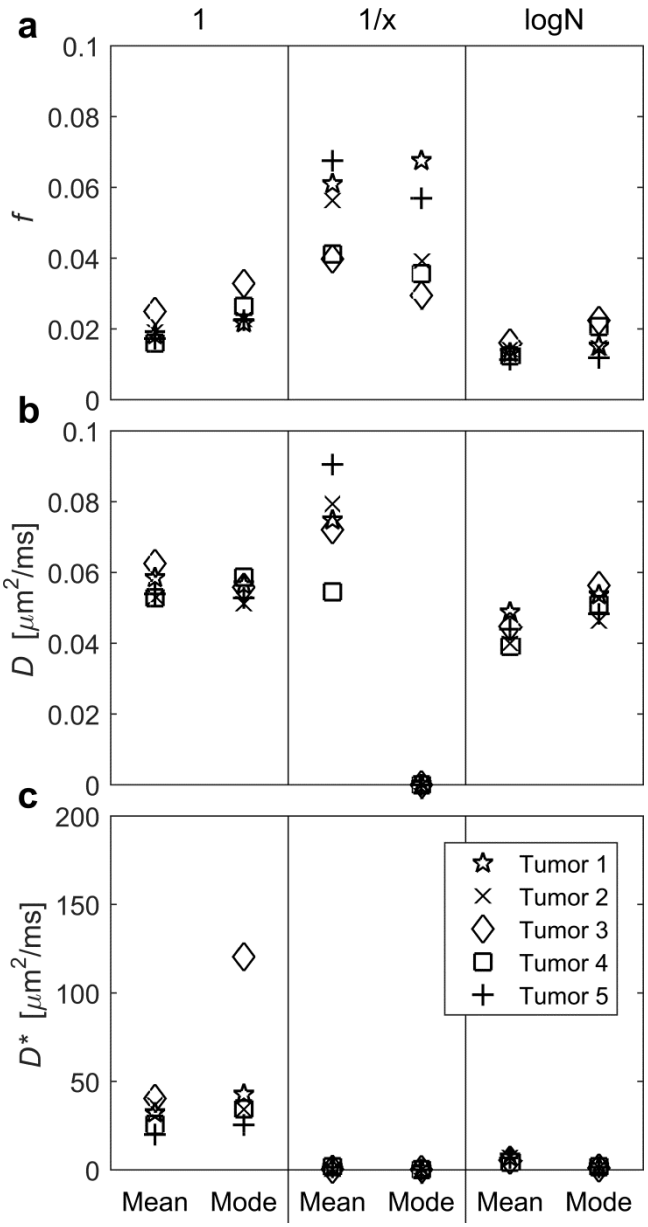

Figure 8. in vivo estimation variability quantified as median local standard deviation derived from 3x3 neighborhoods for each pixel, for all combinations of prior and central tendency measure, for a) *f*, b) *D* and c) *D** respectively. Local standard deviation measures were included from the region defined as tumor in all slices when calculating the median, i.e. the same regions as used in Figure 7

Supporting Figure captions

Supporting Figures S1-S34: Parametric maps as seen in Figure 3 displaying all slices in all tumors included in the study except the one seen in Figure 3

Supporting Figures S35-S52: Bias (median estimation error) and variability (IQR of estimation error) vs. simulated parameter value based on a Gaussian (Eq. 4) and a Rician (Eq. 5) likelihood function for all combinations of prior distribution and central tendency measure

Supporting Figures S53-S70: Estimation error vs. simulated parameter value as seen Figure 5 for all other combinations of prior distribution and central tendency measure

Supporting Figures S71-S72: Correlation between estimation error and simulated parameter value as seen in Figure 6 for SNR = 10 and SNR = 40