Quantifying Regional and Global Liver Function Via Gadoxetic Acid Uptake

by

Josiah Simeth

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biomedical Engineering)
in the University of Michigan
2020

Doctoral Committee:

Professor Yue Cao, Chair
Professor Craig Galban
Professor Luis Hernandez-Garcia
Professor Douglas Noll

Josiah J. Simeth

jjsimeth@umich.edu

ORCID iD: 0000-0001-6847-565X

Acknowledgements

My gratitude to Yue Cao, for her guidance and support through both my masters and doctorate, and for the encouragement to pursue doctoral studies to begin with.

Thanks to the other members of my committee, Craig Galban, Luis Hernandez-Garcia, and Douglas Noll, for their insight and advice.

Additional thanks to Natalie Baughan, Janell Dow, Randal Ten Haken, Martha Matuszak, Mary Feng, Michelle Mierzwa, Dawn Owen, Kyle Cuneo, and Theodore S. Lawrence, for their assistance, advice, and collaboration in enabling and forwarding the research presented here. Particular thanks to Madhava Aryal and Adam Johansson both for their contributions to my research, and their camaraderie as lab mates.

This work is supported in part by NIH grants of R01 CA132834 and P01 CA059827. The author thanks Siemens Healthineer for providing the Radial VIBE pulse sequence.

Table of Contents

Acknowledgements ii

List of Tables v

List of Figures vi

List of Appendices ix

ABSTRACT x

CHAPTER 1 - Introduction 1

CHAPTER 2 - A Linearized Two-Compartment Model of Gadoxetic-Acid Kinetics in the Liver 7

2.1 Introduction 7

2.2 Methods and Materials 8

2.2.1 Dual-Input Two-Compartment Model 8

2.2.2 Linear single-input two-compartment model 9

2.2.3 Evaluation 13

2.3 Results 17

2.3.1 k_1 estimations by the LSITC model vs. DITC model 17

2.3.2 k_1 estimation by simulation of missing the arterial peak 19

2.3.3 k_1 estimation from the clinical MRI and comparison with ICG retention 22

2.3.4 Comparison to Literature values 24

2.4 Discussion 24

2.5 Conclusion 27

CHAPTER 3 - Validation of Gadoxetic Acid Uptake Rate as a Measure of Liver Function 29

3.1 Introduction 29

3.2 Methods 30

3.2.1 Data acquisition 30

3.2.2 Image processing and analysis 31

3.3 Results 33

3.3.1 Similarity between perfusion and uptake rate maps 33

3.3.2 Imaging derived measures of total liver function compared to ICG and ALBI 35

3.4 Discussion 40

3.5 Conclusion 43

CHAPTER 4 - Impact of Mismatch between Liver Function and Hepatic Perfusion on Functional Avoidance Treatment Planning 44

4.1 Introduction 44

4.2 Methods 44

4.3 Results 45

4.4 Discussion 47
4.5 Conclusion 48
CHAPTER 5 - Neural Networks for Robust Quantification of Gadoxetic-Acid Uptake Rate From Suboptimal Acquired MRI 49
5.1 Introduction 49
5.2 Methods 50
5.2.1 Data acquisition 50
5.2.2 Least squares fitting of LSITC model 51
5.2.3 Neural network – rationale and implementation 51
5.2.4 GAN 52
5.2.5 1D CNN 55
5.2.6 Error metric for evaluation of analysis methods and acquisition paradigms 55
5.3 Results 57
5.3.1 Fitting of LSITC model 57
5.3.2 NN model 59
5.3.3 GAN augmented NN model 59
5.3.4 Optimization of time points for the LSITC model fitting 61
5.3.5 1D CNN 63
5.4 Discussion 65
5.5 Conclusions 68
CHAPTER 6 - Conclusions, Summary, and Extensions 69
APPENDICES 72
Appendix A: formulation of dual-input two-compartment equations 72
Appendix B: solution to three parameter linearization 74
REFERENCES 75

List of Tables

Table 1.1 Table of abbreviations appearing in text with corresponding definitions 5

Table 2.1 The demographic, pathological, and clinical Child Pugh scores for the patient groups used. 14

Table 2.2. The ranges of the parameter values for the simulation, where $k_{pvp} = k_{pv}(1\text{-Hct})$ and $k_{ap} = k_a(1\text{-Hct})$. 20

Table 3.1. The patient demographics for all 40 patients 31

Table 3.2. Linear Regression Models of Imaging Derived Total Liver Function Measures to ICGR15 and ALBI. 37

Table 5.1. The values used for the generation of training data using the dual-input, two-compartment model. Note that U(a,b) is the uniform distribution from a to b, and $N(\mu,\sigma^2)$ is the normal distribution about μ with standard deviation σ. In this case the normal distribution was truncated to remove results outside the range [0,1]. 54

Table 5.2. The abbreviations used for each method and data pairing evaluated along with a description of the relevant method and data. 56

Table 5.3. Error rates (NRMSE) for each method as function of data length. Statistically significant improvements in NRMSE over LSITC HTR are indicated by an asterisk (*). Statistically significant increases in error are indicated by a negated asterisk. Significance was estimated based on a two sample t-test with a significance level of 0.05, except for the Max row, where a single sample t-test was used. 63

List of Figures

Figure 2.1. A dual-input two-compartment pharmacokinetic model of gadoxetic acid in the liver. 8

Figure 2.2. An example of the relationship between the vectors y and x in equation (2.10) from a volume of interest in the liver. Note that the linear assumptions of the model only bear out after the transition point x_p. 12

Figure 2.3. Example slices of the k_1 maps estimated from the HTR DCE scans of four patients by the linearized single-input two-compartment model (left) and dual-input two-compartment model (right). Note that the units are mL/100mL/min. 18

Figure 2.4. Scatter plot of the k_1 values estimated by the two models for the 8 patients. The k_1 values in the horizontal axis were estimated from the dual-input 2-compartment model and the ones in the vertical axis from the linearized single-input two-compartment model. 19

Figure 2.5. Early section of the AIFs to demonstrate subsampling used to investigate delays in acquisition. a) shows an AIF with the full dataset, b) shows an AIF with missing data points prior to the arterial peak, and c) shows an AIF with missing data points up to 20 seconds after the arterial peak. 20

Figure 2.6. Plots of the input k_1 values against the k_1 values obtained by LSITC model. The correlation coefficient was R = 0.999, 0,998 and 0.999 for the full dataset (a), the dataset missing time points prior to the arterial peak (b), and the dataset missing time points up to 20 seconds after the arterial peak (c) respectively. 21

Figure 2.7. Boxplots for the WAPE in the LSITC model results relative to the simulated values at 3 noise levels for the three simulated datasets: a) the full dataset, b) the dataset missing data points prior to the arterial peak, and c) the dataset missing data points up to 20 seconds after the arterial peak. The median values are indicated by the horizontal line, the mean is indicated by the diamond, and error bars are 1.5 times the interquartile range. 22

Figure 2.8. Example maps of k_1 and v_{dis} in the assessed livers. Notice that high values of v_{dis} can be seen to correspond to vasculature. 23

Figure 2.9. Plot of the log of ICG retention at 15 minutes against the sum of K_1 values in the contoured total liver volume. R = -0.72 (p = 0.0004, n = 20). 23

Figure 2.10. This plot shows the mean values obtained in several studies for background liver uptake rate of gadoxetic acid in cancerous livers, and a group of healthy livers for reference. Error bars indicate the standard deviation. The HTR values are based on a removed outlier..... 24

Figure 3.1. Example slices of k_1 (top row) and k_{pv} maps (bottom row) presented in 100 quantiles of 4 patients with different ICGR15 and ALBI values. Note that large vessels generally show as dark in uptake rate maps (lacking hepatocytes), while in perfusion maps they can be either dark (arterial vessels) or bright (portal venous vasculature), depending on the type of vessels..... 33

Figure 3.2. histogram of the distribution of ranked correlations between k_1 and k_{pv} values (mean: 0.095, median: 0.12, n=53). 34

Figure 3.3a-3g $\log_{10}$(ICGR15) plotted against: (a) mean k_1 by Total Volume, (b) summed K_1 = summed k_1V_{dis}, (c) total liver Volume outside GTV, (d) mean $k_{pv,}$ (e) weight normalized mean k_1 by Total Volume, (f)) weight normalized summed K_1 = summed k_1V_{dis}, (g)) weight normalized total liver Volume outside GTV. The plotted linear regression fit in (a), (b), (e), and (f) each ignore one outlier in terms of k_1, and the fit for (g) ignores 4 outliers in terms of weight normalized liver volume. Correlation coefficients do not exclude outliers..... 38

Figure 3.4a-4g. The raw ALBI plotted against: (a) mean k_1 by Total Volume, (b) summed K_1 = summed k_1V_{dis}, (c) total liver Volume outside GTV, (d) mean $k_{pv,}$ (e) weight normalized mean k_1 by Total Volume, (f)) weight normalized summed K_1 = summed k_1V_{dis}, (g)) weight normalized total liver Volume outside GTV. The plotted linear regression fit in (a), (b), (e), and (f) each ignore one outlier in terms of k_1, and the fit for (g) ignores 5 outliers in terms of weight normalized liver volume. Correlation coefficients do not exclude outliers..... 39

Figure 4.1. A demonstration of sparing region binning and comparison between the sparing and non-sparing regions as defined by the function and perfusion maps (k_1 and k_2). Note profound differences in top and bottom images. 46

Figure 4.2. The mean dose to the function sparing region for the liver volume sparing plan, perfusion based plan, and function based plan. Ordered by the descending ranked correlations between maps of k_1 and k_2, with the correlations given above each set of columns. 47

Figure 4.3. The dose distribution for the liver volume, function, and perfusion sparing plans for an example patient. Note that in this case the function and perfusion based plans are quite similar compared to the liver sparing plan. (Dose distribution not decipherable in greyscale)..... 47

Figure 5.1. Illustration of characteristics of densely sampled high temporal resolution (HTR – left) and more sparsely sampled low temporal resolution (LTR - right) datasets. HTR data is regularly sampled at 5-10 s intervals for the duration of 16-20 min. LTR data involves the acquisition of three post contrast samples uniformly spaced at intervals of 15 to 35 seconds, followed by two points, one at roughly 10 min and another at roughly 20 min post contrast. LTR data is the clinical norm. 51

Figure 5.2. Architecture of the simple fully connected neural network used to predict voxel-wise k_1..... 52

Figure 5.3. The design of the GAN used for generation of C_a and C_{pv} curves. Parenthetical values represent the dropout rate for dropout layers, the gradient of the leaky Relu, and the number of size for all other layers. 53

Figure 5.4. The design of the 1D CNN applied to HTR data. Convolutional layer 1 contains 8 filters with width 9, convolutional layer 2 contains 16 filters with width 5, and the fully connected contains 10 neurons). 55

Figure 5.5. Errors of estimated k_1 values with varied acquisition lengths for the tested methods. 58

Figure 5.6. The k_1 maps created using the HTR and LTR data truncated at 10 min both from directly fitting the LSITC model (second and third columns) and from the NN and GAN augmented NN models (fourth and fifth columns respectfully). The first column displays the reference k_1 images by fitting the LSITC model to full length HTR data acquired over approximately 20 min. 58

Figure 5.7. Examples of generated (top row) and nearest neighbors from the measured (bottom row) C_a and C_{pv} curve pairs. Nearest neighbors were calculated based on the sum of squared differences in C_a alone. 60

Figure 5.8. For each of the 5 GANs used, the probability distributions for L_2 norm of the distance between randomly selected C_a and C_{pv} curves for GAN generated data is shown in red. The probabilities for the measured data are shown in blue as reference. 61

Figure 5.9. The errors in simulated and real data as a function of the number of optimum sampling points using a procedure derived from the genetic algorithm. Note that error in the data leveled off after 8 points. 62

Figure 5.10. Example maps of uptake rate based on (top to bottom in each block) LSITC and DSITC models for full length HTR data, and the CNN for only 10 minutes of data. The colormap depicts the value from 0 to 25 mL/100mL/min. 64

List of Appendices

Appendix A. The formulation of the dual-input two-compartment equations 72

Appendix B. The solution to the three parameter formulation of the LSITC model. 74

ABSTRACT

Liver function is a dominant factor in the survival of patients with hepatocellular carcinoma (HCC). Measures of regional and global liver function are critical in guiding treatments for intrahepatic cancers. Regional and global liver function assessments important for defining the magnitude and spatial distribution of radiation dose to preserve functional liver parenchyma and reduce incidence of hepatotoxicity from radiation therapy (RT) for intrahepatic cancer treatment. This individualized liver function-guided RT strategy is critical for patients with heterogeneous and poor liver function, often observed in cirrhotic patients treated for HCC. Dynamic gadoxetic-acid enhanced (DGAE) magnetic resonance imaging (MRI) allows investigation of liver function through observation of the uptake of contrast agent into the hepatocytes.

This work seeks to determine if gadoxetic uptake rate can be used as a reliable measure of liver function, and to develop robust methods for uptake estimation with an interest in the therapeutic application of this knowledge in the case of intrahepatic cancers. Since voxel-by voxel fitting of the preexisting nonlinear dual-input two-compartment model is highly susceptible to over fitting, and highly dependent on data that is both temporally very well characterized and low in noise, this work proposes and validates a new model for quantifying the voxel-wise uptake rate of gadoxetic acid as a measure of regional liver function. A linearized single-input two-compartment (LSITC) model is a linearization of the pre-existing dual-input model but is designed to perform uptake quantification in a more robust, computationally simpler, and much faster manner. The method is validated against the preexisting dual-input model for both real and simulated data. Simulations are used to investigate the effects of noise as well as issues related to the sampling of the arterial peak in the characteristic input functions of DGAE MRI.

Further validation explores the relationship between gadoxetic acid uptake rate and two well established global measures of liver function, namely: Indocyanine Green retention (ICGR) and Albumin-Bilirubin (ALBI) score. This work also establishes the relationships between these

scores and imaging derived measures of whole liver function using uptake rate. Additionally, the same comparisons are performed for portal venous perfusion, a pharmacokinetic parameter that has been observed to correlate with function in patients with relatively good liver function, and has been used as a guide for individualized liver function-guided RT. For the patients assessed, gadoxetic acid uptake rate performs significantly better as a predictor of whole liver function than portal venous perfusion.

This work also investigates the possible gains that could be introduced through use of gadoxetic uptake rate maps in the creation of function-guided RT plans. To this end, plans were created using both perfusion and uptake, and both were compared to plans that did not use functional guidance. While the plans were generally broadly similar, significant differences were observed in patients with severely compromised uptake that did not correspond with compromised perfusion.

This dissertation also deals with the problem of quantifying uptake rate in suboptimal very temporally sparse or short DGAE MRI acquisitions. In addition to testing the limits of the LSITC model for these limited datasets (both realistic and extreme), a neural network-based approach to quantification of uptake rate is developed, allowing for increased robustness over current models.

CHAPTER 1
Introduction

In the treatment of liver cancer, the competing goals of controlling tumor growth and maintaining liver function must be balanced against one another. The fundamental goal is to maintain sufficient liver function to prevent short term liver failure, while sufficiently controlling tumor growth to prevent further growth of the cancer. This is of particular concern in radiation therapy, where the dose targeted for the cancer will inevitably cause some, albeit reduced, damage to the tissues near the lesion. It is thus essential to ensure, prior to treatment, that the retained liver function after treatment or retained "functional volume" is sufficient[1].

Metastatic cancers in the liver are most likely to leave overall liver function outside the tumor volumes relatively unaffected. In these cases, it is logical to assume that overall functional loss will be roughly proportional to the non-tumor area damaged by therapy. The radiation induced damage on normal tissue can be modeled based on the observed response in a sample population[1–3]. This damage can be assessed using a measure of whole liver function, e.g. indocyanine green (ICG) retention[1,4] or Albumin-bilirubin (ALBI) score[5], to ensure that the radiation dose does not exceed the threshold to cause liver failure.

However, some pathologies, such as cirrhosis, compromise function throughout the liver[4,6]. Additionally, this compromise is often inhomogeneous, and not limited to identifiable lesions. Cirrhosis is often a comorbidity with Hepatocellular Carcinoma (HCC), meaning treatment of HCC tumors must consider this inhomogeneous functional distribution. It is in cases like this that measures of regional liver function are particularly necessary[7,8]. Without regional measures the uncertainty in retained function is greatly raised, making it more difficult to effectively control tumor growth without increasing the likelihood of liver failure to an unacceptable degree. Thus, while global measures do allow for clinicians to monitor treatment and pause or modify treatment plans to avoid radiation induced liver disease (RILD) and liver failure, regional

measures allow a more granular understanding of functional distribution, enabling radiation targeting such that the likelihood of both RILD and of continued tumor growth are simultaneously lowered.

Various methods currently exist for determining regional liver function. Positron emission tomography (PET) and single positron emission computed tomography (SPECT) with radioactive hepatobiliary tracers, have been developed for direct measurement of regional liver function[7,9,10]. MRI-based methods benefit from superior spatial and temporal resolutions and soft tissue contrast. Similar advantages have motivated MRI guided radiation therapy[11,12]. Dynamic contrast enhanced (DCE) MRI with the correct contrast allows for physiological investigation of liver function. Vascular contrast agents, such as Gadobenic acid have enabled interrogation of portal venous perfusion through dynamic Gadobenic acid enhanced MRI scans of the liver[13–15]. Analysis of the resulting imaging allows for determination of the perfusion parameters throughout the liver. Portal vein perfusion in particular has been implemented for this purpose, operating on the assumption that poor liver health is correlated with compromised perfusion in that region. Portal venous perfusion has been found to be significantly correlated with overall function evaluated via indocyanine green (ICG) clearance or retention[13,16,17]. However, since this is an indirect surrogate for function there is reason to doubt this is an accurate measure when function is more heterogeneous as in HCC.

Though its kinetics are more complex, a hepatobiliary contrast, such as gadoxetic acid more directly applicable to the assessment of function. Marketed in the United States as Eovist, and as Primovist in Europe, gadoxetic acid is a hepatobiliary contrast that allows direct assessment of liver function through contrast uptake in liver parenchyma[18]. Hepatobiliary contrasts are distinguished from vascular agents in that they are taken up into the liver cells, allowing more direct interrogation of liver function. Over the course of a gadoxetic acid enhanced (DGAE) MRI, the gadoxetic acid uptake rate in a given region is a meaningful quantifier of functioning hepatocytes in that region. Since hepatocytes are the drivers of liver function, this is very direct method of quantifying liver function[14,19–21].

While the hepatobiliary property of gadoxetic acid is of great interest for the assessment of liver function, currently gadoxetic acid is only indicated for use in assessing intrahepatic lesions clinically[18]. Correspondingly, most current gadoxetic acid scans are conducted for diagnostic purposes, allowing the localization, identification, and characterization of tumors within the liver. The vascular and hepatobiliary properties of the contrast allow for discrimination of tumor properties from a few MRI volumes acquired at a few key times. Recommended acquisitions generally include 6 image volumes over approximately 20 min. The first four phases are meant to capture the pre-enhancement phase, arterial phase, portal venous phase, and "mixing phase" respectively. The remaining volumes attempt to characterize the hepatobiliary phase, where uptake is a predominant feature. While only indicated for lesion characterization, the same properties that motivate this use case also open the door for interrogation of voxelwise liver function, rather than mere tumor identification[18].

The kinetics of gadoxetic acid can be described using a dual-input two-compartment (DITC) model of liver function, where the dual inputs are due to the unique physiology of the liver's vasculature[20]. In addition to an arterial blood supply, the liver has a portal venous blood supply, responsible for carrying nutrients from the digestive system to the liver. Blood from both the arteries and the portal vein leave the liver via the hepatic vein. The two compartments represent the relevant intra and extracellular spaces. With the introduction of gadoxetic acid, the relative signal strength at a given location can be used to estimate the concentration of contrast at that location. Using the concentration over time in the arterial and portal venous input, along with the concentration at an ROI, the contrast kinetics can be fit and quantified using the DITC model. The DITC model of gadoxetic acid kinetics is non-linear, and depends on simultaneous fitting of 6 parameters, a problem with significant computational complexity and challenges for robust fitting[20]. Such a method completely rules out using the relatively common low temporal resolution (LTR) scans, as the data points typically required leave the problem very underdetermined. An accurate fit using the DITC model requires well temporally characterized data, with imaging volumes acquired every 5 to 10 seconds. This fitting is also sensitive to noise, an issue that is compounded by the short acquisition time allowed for each volume.[22]

In addition to the direct benefits of an accurate measure of liver function in tailoring radiation therapy plans, a sufficiently accurate measure of regional liver function would also allow for replacement of tests used to quantify whole liver function, which are important for evaluating individual patient risk of RILD. This would then allow for a single acquisition to characterize the tumors, map liver regional liver function, and evaluate the whole liver function.

In any imaging examination, the time taken for the patient and the clinicians should be minimized for overall convenience and economic viability, as well as increased likelihood of patient capacity to complete the scan with minimal discomfort or motion. Machine learning approaches such as neural networks have introduced powerful means for extracting useful information, regression, classification, and decision making from restricted data sets, using learned features that do not have to be identified and crafted by human agents. In the best cases neural network based approaches can perform as well as the best crafted models with less information. If neural network based methods could give comparable results to 20 minutes of temporally well characterized data assessed via preexisting model based methods, using smaller, more restricted datasets, it would be extremely valuable. Additionally, if neural network based approaches could derive reliable uptake rate maps using only LTR data, the generalizability of gadoxetic acid uptake rate would be substantially improved with minimal workflow changes required.

This work seeks to determine if gadoxetic uptake rate can be used as a reliable measure of liver function, and to develop robust methods for uptake estimation with an interest in the therapeutic application of this knowledge in the case of intrahepatic cancers.

Chapter 2 introduces a model for quantifying the voxel wise uptake rate of gadoxetic acid as a measure of regional liver function. This model is derived from the pre-existing dual-input two-compartment model but is designed to perform uptake quantification in a computationally simpler, more robust, and much faster manner. The method is compared against the dual-input two-compartment model for real and simulated data. Simulations are used to investigate the effects of noise as well as issues related to the sampling of the arterial peak in the characteristic input functions of dynamic gadoxetic acid enhanced MRI. This chapter also Goes into more

depth on the specifics of the DITC model, and the relative advantages of the gadoxetic acid uptake rate used here, as compared to other measures that depend on gadoxetic uptake.

Chapter 3 goes into greater depth in validating this measure of liver function against well regarded measures of global liver function (ICG and ALBI), and compares uptake to a promising pre-existing surrogate for liver function: portal venous perfusion. The comparisons with global functional measures also provide calibrations for calculating equivalent ICG and ALBI scores from DGAE imaging. This enables adherence to guidelines defined using ICG or ALBI but informed only by the imaging measure. This chapter also contains comparisons to contemporary work related to the quantification of regional uptake rate of Gadoxetic acid.

Chapter 4 evaluates the relative gains in using uptake rate instead of perfusion in functional avoidance RT planning. This is performed by comparing RT plans created using each method for a small set of patients. Evaluations determine the differences between the dose distributions in each plan, with a focus on reduced dose to highly functioning regions.

Chapter 5 tests the limits of the model for suboptimal data, in terms of both the temporal length of the acquisition and the extreme under sampling of LTR data. Furthermore, this chapter examines machine learning techniques to improve performance in these extreme cases. These approaches focus on LTR style undersampling, using a simple neural network to estimate actual uptake from this restricted dataset. Training is tested using both real and simulated data. Additionally, data augmentation is investigated via the generation of random realistic arterial input functions using a generative adversarial neural network.

Table 1.1 *Table of abbreviations appearing in text with corresponding definitions*

Abbreviation	**Definition**
AIF	Arterial Input Function
ALBI	Albumin-Bilirubin
CNN	Convolutional Neural Network
CP	Child-Pugh
DCE	Dynamic Contrast Enhanced
DGAE	Dynamic Gadoxetic Acid Enhanced
DITC	Dual-Input, Two-Compartment

EMD	Earth Mover's Distance
GAN	Generative Adversarial Network
GTV	Gross Tumor Volume
HCC	Hepatocellular MRI
Hct	Hematocrit
HEF	Hepatic Extraction Fraction
HTR	High Temporal Resolution
ICG	Indocyanine Green
ICGR	Indocyanine Green Retention
ICGR15	ICG retention at 15 minutes
LLS	Linear Least Squares
LSITC	Linearized Single-Input, Two-Compartment
LTR	Low Temporal Resolution
LV/SLV	Liver to spleen ratio
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
NLLS	Non-Linear Least Squares
NN	Neural Network
NRMSE	Normalized Root Mean Squared Error
NTCP	Normal Tissue Complication Probability
OAR	Organ At Risk
PK	Pharmacokinetic
PVIF	Portal Venous Input Function
RE	Relative Enhancment
RILD	Radiation Induced Liver Disease
ROI	Region of Interest
RT	Radiation Therapy
SBRT	Stereotactic Body Radiation Therapy
SD	Standard Deviation
SNR	Signal to Noise Ratio
WAPE	Weighted Absolute Percent Error
WMAPE	Weighted Mean Absolute Percent Error

CHAPTER 2

A Linearized Two-Compartment Model of Gadoxetic-Acid Kinetics in the Liver

This chapter is substantially equivalent to work the author has already published in *NMR in Biomedicine*[22].

2.1 Introduction

The hepatobiliary properteris of gadoxetic acid make it an attractive contrast for assessing regional liver function. Using this agent, hepatic extraction fraction (HEF) can be estimated to assess liver function. While HEF is directly related to uptake rate, it cannot isolate uptake rate from the effects of plasma flow[14,23]. Semi-quantitative measures such as relative enhancement (RE) and enhancement relative to spleen similarly cannot differentiate between uptake rate and plasma flow, while additionally either ignoring fluid enhancement or assuming its uniform conformity to fluid enhancement in the spleen[14]. A dual-input two-compartment model of liver function can be used to directly estimate the uptake rate, but the model requires high temporal resolution images to adequately characterize the concentration curves used as inputs, can have as many as 6 unknown parameters, and may be susceptible to over fitting variance in the data [20].

This chapter develops and applies a linear model based on the dual-input two compartment model of liver function. Ideally, this model can be used to estimate both flow dependent and independent measures of liver function with decreased computational complexity and susceptibility to variance as compared to the dual-input two-compartment model. To this end we compare the results of the developed model to the dual-input two-compartment model as applied to high-temporal resolution gadoxetic-acid enhanced MRI data, and to ICG retention in livers with corresponding clinical temporally-sparse MRI data. Through simulations we also assess the impact of failure to capture the peak of the arterial input function on the estimate of uptake. This allows us to confirm the correspondence of our approach to the two-compartment model for liver function, and to an independent and reliable measure of whole liver function found in the

extraction by the liver of ICG[24], while confirming the applicability to low temporal resolution clinical datasets.

2.2 Methods and Materials

2.2.1 Dual-Input Two-Compartment Model

2.2.1.1 Model Description. The dual-input two-compartment (DITC) pharmacokinetic (PK) model of gadoxetic acid in the liver[20] is illustrated in figure 2.1. This model describes the hemodynamics of gadoxetic acid (hepatic arterial and portal venous perfusion), and contrast uptake of hepatocytes. In this model, after injection of a bolus of gadoxetic acid, the contrast circulates in the blood by flowing in from both hepatic artery and portal vein into the sinusoids, distributing in the space of Disse and flowing out through the central and hepatic veins. This assumes fast exchange between the sinusoids and the space of Disse. Meanwhile, hepatocytes take up the contrast through the sinusoid membrane. The contrast uptake of hepatocytes is assumed to be unilateral, by omitting the minor efflux of the contrast back to sinusoids in the initial retention period[25]. This model also omits the slow and delayed excretion process.

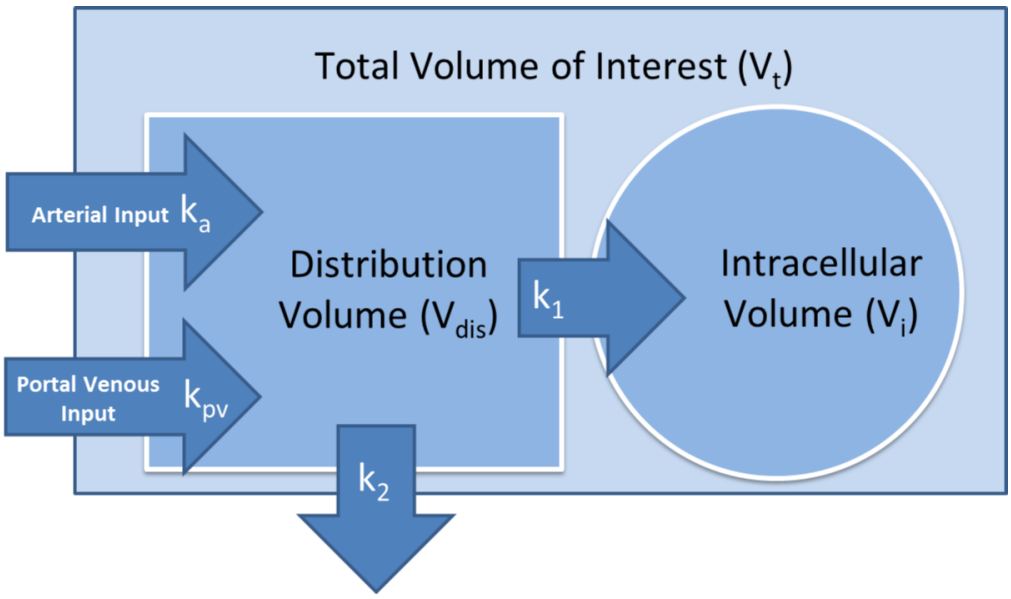

Figure 2.1. *A dual-input two-compartment pharmacokinetic model of gadoxetic acid in the liver.*

If we consider a voxel or a volume of interest with a total volume of V_t, the total amount of contrast in the voxel is a sum of the amounts of contrast in the extracellular and intracellular spaces, and can be described by the following equations:

$$\overbrace{V_t C_t(t)}^{\text{Contrast in Tissue}} = \overbrace{V_{dis} C_{dis}(t)}^{\text{Extracellular Contrast}} + \overbrace{k_1 \int_0^t V_{dis} C_{dis}(\tau) d\tau}^{\text{Intracellular Contrast}} \quad (2.1)$$

$$V_{dis} C_{dis}(t) = V_{dis} \int_0^t \left(k_a C_a(\tau - \tau_a) + k_{pv} C_{pv}(\tau - \tau_{pv}) \right) e^{-(t-\tau)(k_2 + k_1)} d\tau \quad (2.2)$$

where V_{dis} is the distribution volume of blood; C_t, C_{dis} , C_a, and C_{pv} are contrast concentrations as a function of time in the respective total, distribution, arterial, and portal vein volumes; τ_a and τ_{pv} describe respective arrival time delays of the arterial and portal vein input functions at each voxel; k_a and k_{pv} describe the normalized arterial and portal venous flow rates, and k_2 is the normalized flow rate leaving the volume of interest through the central vein. k_1 is the normalized rate of uptake of contrast to the intracellular space. The distribution volume includes the Space of Disse and sinusoids. We also define a fractional distribution volume of $v_{dis} = \frac{V_{dis}}{V_t}$. The derivation and a detailed description of equations are given in appendix A.

C_t, C_a, and C_{pv} are measurable from the intensity of dynamic contrast enhanced MRI at regions or voxels of hepatic tissue, artery, and portal vein respectively. As a result, equations (2.1) and (2.2) have 6 unknown variables (k_a, k_{pv}, τ_a, τ_{pv}, k_1 *and* v_{dis}) to be determined.

2.2.1.2 Optimization. To determine the 6 unknown variables in equations (2.1) and (2.2), the cost function

$$\sum_{i=0}^{N_t-1} \left[C_t(iT) - \hat{C}_t(iT) \right]^2 \quad (2.3)$$

is optimized. Here $\hat{C}_t$ is the estimate of C_t given by the model in equation (2.1) with guessed values of k_a, k_{pv}, τ_a, τ_{pv}, k_1 and v_{dis} during the optimization process. T is the temporal interval between time points and N_t is the total number of time points in the DCE curves. This work used the Nelder Mead Simplex algorithm to perform the optimization.

2.2.2 Linear single-input two-compartment model

2.2.2.1 Rationale. Since fitting the DITC PK model requires the optimization of 6 parameters, it is susceptible to overfitting of variations due to noise, and also is time consuming when fitting a long dynamic series of data in the whole liver. Estimating k_1 (the contrast uptake rate of hepatocytes) requires a long time period of observation of the contrast accumulation in

hepatocytes. The hemodynamic changes after the initial transient time following the contrast bolus injection become slow. This offers an opportunity to solve the problem in a different manner, producing a computationally simpler problem and, ideally, reducing susceptibility to variation. Assumptions used in the derivation and formula are described in the following subsections

2.2.2.2 Assumptions & Formulation. The change in the total amount of contrast in the distribution volume in a voxel is:

$$V_{dis}\frac{dC_{dis}(t)}{dt} = V_{dis}[k_a C_a(t-\tau_a)+k_{pv}C_{pv}(t-\tau_{pv})]-V_{dis}(k_2+k_1)C_{dis}(t) \quad (2.4)$$

which is equation (2.2) in the derivative form. Given the long acquisition period for observation of gadoxetic-acid uptake in hepatocytes, after a few circulations of the contrast bolus in the blood ($t > t_p$), the contrast concentration in the portal vein blood is eventually equal to that in the arterial blood, $C_a(t) = C_{pv}(t)$. Under this condition, equation (2.4) can be re-written as:

$$V_{dis}\frac{dC_{dis}(t)}{dt} = V_{dis}\,k_t C_a(t) - (k_1 + k_2)V_{dis}C_{dis}(t) \quad when\ t>t_p \quad (2.5)$$

where $k_t = (k_a+k_{pv}) = (1-Hct)k_2$, the normalized total blood flow rate in V_t, and *Hct* is hematocrit. Equation (2.5) can be re-arranged as:

$$C_{dis}(t) = \frac{k_t}{k_1+k_2}C_a(t) - \frac{1}{k_1+k_2}\frac{dC_{dis}(t)}{dt} \quad when\ t>t_p \quad (2.6)$$

Substituting C_{dis} in equation (2.6) into the second term in equation (2.6), equation (2.6) can be further re-written as:

$$C_{dis}(t) = \frac{k_t}{k_1+k_2}C_a(t) - \frac{k_t}{(k_1+k_2)^2}\frac{dC_a(t)}{dt} + O(\frac{d^2C_a(t)}{dt^2}) + \ldots \quad when\ t>t_p \quad (2.7)$$

where the first term depends upon C_a, the second term depends upon the first derivative of C_a, the third term depends upon the second derivative of C_a, and so on. If the second derivative of C_a is small enough to be neglected, substituting equation (2.7) into equation (2.1) and re-arranging the terms, we have:

$$(1 - Hct)C_t(t) = v_{dis}\frac{k_2}{k_1+k_2}\left\{\left(1 - \frac{k_1}{k_1+k_2}\right)C_a(t) + k_1\int_0^t C_a(\tau)\,d\tau - \frac{1}{k_1+k_2}\frac{dC_a(t)}{dt}\right\} \quad (2.8)$$

Equation (2.8) can be considered as the linear problem $y = ax_1+bx_2+cx_3$, where $y = (1 - Hct)C_t(t)$, and $\vec{x} = (x_1, x_2, x_3) = (C_a(t), \int_0^t C_a(\tau)\,d\tau, \frac{dC_a(t)}{dt})$. A linear least squares (LLS) fit can estimate coefficients of *a*, *b* and *c*. k_1, k_2 and v_{dis} can be solved from the coefficients (see appendix B).

If the second term (related to the first derivative of C_a) in equation (2.7) can be neglected, we have:

$$(1 - Hct)C_t(t) = v_{dis}\frac{k_2}{k_1+k_2}\left\{C_a(t) + k_1\int_0^t C_a(\tau)\,d\tau\right\} \quad (2.9)$$

Again, equation (2.9) is a linear problem, $y = ax_1+bx_2$, where $y = (1 - Hct)C_t(t)$, and $\vec{x} = (x_1, x_2) = (C_a(t), \int_0^t C_a(\tau)\,d\tau)$, which can be solved by LLS fitting. In this case, $k_1 = \frac{b}{a}$. Note that there is no assumption made relating to k_1 and k_2; and but k_2 and v_{dis} cannot be solved. Also, equation (2.9) can be re-arranged to be:

$$\overbrace{\frac{(1-Hct)C_t(t)}{C_a(t)}}^{y} = \overbrace{v_{dis}k_1\frac{k_2}{k_1+k_2}}^{slope}\overbrace{\frac{\int_0^t C_a(\tau)d\tau}{C_a(t)}}^{x} + \overbrace{v_{dis}\frac{k_2}{k_1+k_2}}^{intercept} \quad (2.10)$$

which is a form of the Patlak analysis[26]. Note that k_1 can be calculated by slope/intercept, and is not affected by the relationship between k_1 and k_2, which is different from the Patlak analysis. The intercept in the Patlak analysis is called v_{eff} and is usually greater than the true blood distribution volume. The intercept in our case, $v_{dis}\frac{k_2}{k_1+k_2}$, is smaller than v_{dis}. However, if we assumed that $k_2 >> k_1$, v_{dis} can be estimated by the intercept. Note that we only used the assumptions: $C_a(t) = C_{pv}(t)$ and that there is slow contrast change in the blood after $t > t_p$, to derive these equations. We will call it the linearized single-input two-compartment (LSITC) model hereafter.

2.2.2.3 Optimization. Optimization of equation (2.10) involves first computing the vector *x* and the set of vectors *y* (one for each voxel). Additionally, t_p (or x_p corresponding to t_p) needs to be determined. Based on the assumptions of the model, *x* and *y* will be linearly related after x_p,

suggesting a linearity test is needed. If it is assumed that x_p is relatively consistent throughout the liver a single test can be performed, reducing noise effects and saving computation time.

To obtain x_p, the vectors y are averaged over all voxels within the liver to form a single vector. The two singular values of the centered data matrix [x y] after the tested x_p are acquired by the singular value decomposition. The tested x_p is varied within a time interval between the arterial peak and 2 min before the last data point. The ratio of the first singular value to the second is calculated to determine the linearity of the relationship. x_p is then chosen to maximize this ratio. Figure 2.2 illustrates the behavior of the y vector before and after x_p in a region of interest. However, in cases with sparse temporal sampling the process can be simplified by setting t_p based on the DCE data with high temporal sampling.

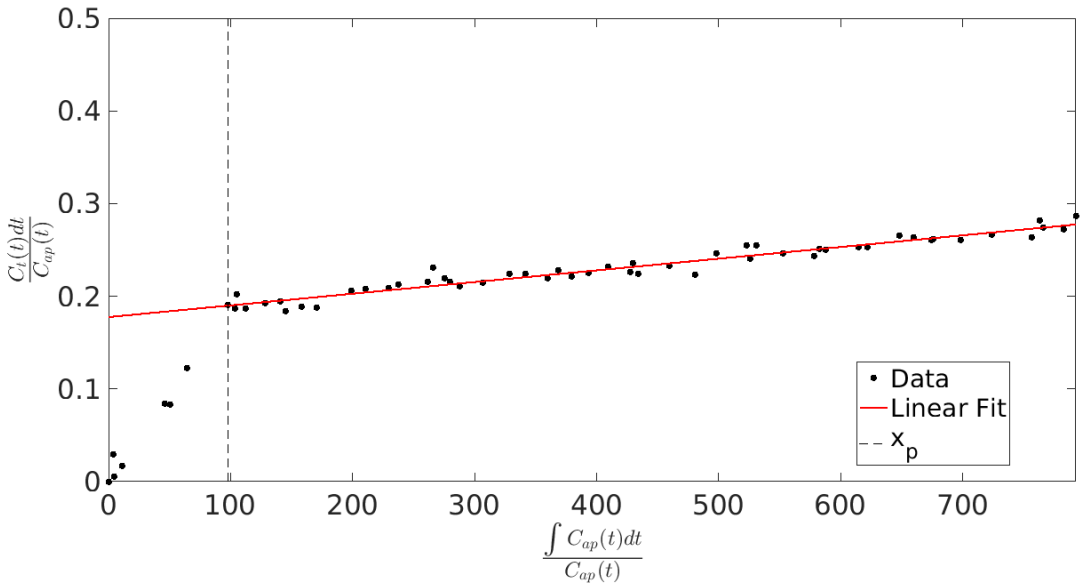

Figure 2.2. *An example of the relationship between the vectors y and x in equation (2.10) from a volume of interest in the liver. Note that the linear assumptions of the model only bear out after the transition point* x_p.

After selection of x_p total least squares regression is performed for each voxel using the data after x_p to minimize the impact of errors in both x and y. The slope of the resulting fit is divided by the intercept to determine the value of k_1 in the voxel. In cases where the intercept is less than 0.02, k_1 is set to zero to prevent values from blowing up. This is also justified in that a low intercept, corresponding to a sufficiently low v_{dis} will effectively preclude meaningful levels of uptake in the voxel.

2.2.3 Evaluation

This chapter evaluates the performance of the LSITC model, with reference to the results of 6 parameter optimization of the established DITC model, simulated data based on the DITC model, global liver function assessment through ICG retention in the plasma, and a comparison to literature values.

2.2.3.1 Data Acquisition. In order to compare the results of the proposed LSITC model with the DITC model, 3D volumetric DCE MRI of the liver were acquired during the intravenous injection of a single standard dose of gadoxetic acid using a Golden-Angle Radial sampling VIBE sequence on a 3T scanner (Skyra, Siemens Healthineer) in a prospective protocol approved by University of Michigan Institutional Review Board. A total of 14 exams from 13 patients with intrahepatic cancers prior to radiation therapy were acquired during free breathing. Demographic, pathological, and clinical Child Pugh scores of this group of patients (called Group 1) are provided in Table 2.1. Note that 54% of the patients had Child Pugh scores of 5, indicating good liver function. In this group of the patients, 3D DCE images were acquired with temporal resolutions of 3.5 to 10 seconds and total acquisition times of 4 to 24 minutes, and covered the whole liver with 64 to 72 slices with slice thicknesses of 2.6 to 4.5 mm and in-plane resolution of 2.1×2.1 mm to 2.4×2.4 mm (192×192 pixels). These scans will be referred to as high temporal resolution (HTR) scans. The high temporal resolution scans could be fitted to both the DITC model and the LSITC model, allowing comparison of the results obtained from the two methods.

Table 2.1 *The demographic, pathological, and clinical Child Pugh scores for the patient groups used.*

Characteristic	HTR n=13	LTR n=19	Total n=32
Median Age (range)	60.7 (56.5 - 72.0)	61.1 (52.7 - 78.9)	61.0 (52.7 - 78.9)
Gender			
Male	10 (77%)	14	24
Female	3 (23%)	5	8
Cirrhosis			
Positive	8 (62%)	16*	24
Negative	5 (38%)	2*	7
Histology			
Hepatocellular carcinoma	9 (69%)	18*	27
Adenocarcinoma	2 (15%)	0	2
Intrahepatic cholangiocarcinoma	1 (8%)	0	1
Solitary fibrous tissue	1 (8%)	0	1
Baseline Child Pugh			
5	7 (54%)	5 (28%)	12 (40%)
6	4 (31%)	7 (39%)	11 (37%)
7	0 (0%)	5 (28%)	5 (17%)
8	1 (8%)	1 (6%)	1 (3%)
9	1 (8%)	0 (0%)	1 (3%)

In addition, the proposed LSITC model was applied to clinical multi-phase MRI with gadoxetic acid in 19 patients, which was approved by a retrospective protocol of University of Michigan Cancer Center. Demographic, pathological, and clinical Child Pugh scores of this second group of patients are provided in Table 2.1. Pre- and post-RT scans were acquired from 19 patients, for a total of 40 scans. Each patient had been diagnosed with HCC. The scans were comprised of 3D volumetric multi-phase MRI of the liver during the intravenous injection of a single standard dose of gadoxetic acid on either a 1.5 T GE or a 1.5T Philips scanner. Each exam consisted of a pre-contrast volume, three-phase (arterial and portal vein phases) volumes that were each spaced approximately 20 seconds apart, and hepatobiliary phase volumes at approximately 10 and 20 minutes post contrast, for a total of 6 time points. Each acquisition was obtained during a breath hold. 3D MRI had 88 to 124 slices per volume with 256×256 pixels to 512×512 pixels in the

plane. The pixel size varied from 0.7×0.7 mm to 1.4×1.4 mm within each slice, with the slice thickness consistently 2 mm. These scans will be referred to as low temporal resolution (LTR) scans. However, due to changes in flip angle between phases in the scan (particularly in the late phases) and image quality issues, the set of usable scans was only 27 of the original 40. ICG retention exams were carried out near the time of the scan for 20 of the 27 scans, without RT or any other treatment having taken place in the meantime. The ICG retention score as a quantitative overall liver function assessment was measured as the percentage of the original ICG dose remaining 15 minutes after injection, as described previously[24], with higher plasma retention signifying poorer liver function. The patients in this second group had a median baseline ICG retention of 37.2%, with minimum and maximum retention scores of 9.8% and 50.2%.

2.2.3.2 Image pre-processing. For the clinical LTR multi-phase images, interpolation was carried out on the image volumes that had different spatial resolutions between the volumes. All 3D multi-phase LTR MRI volumes in an exam were co-registered using a robust, over-determined image registration method[27]. For all volumes the aorta was contoured from the aortic split to the liver up 3 cm. The 100 voxels within this region that had the highest contrast just before the arterial concentration peak were averaged to form the arterial input function (AIF). For the HTR DCE data, the portal vein was also contoured and selected by the same process to obtain the portal vein input function (PVIF). In both cases relative enhancement was used to create the input functions:

$$C(iT) \propto \frac{SI_i}{SI_{precontrast}} - 1 \tag{2.11}$$

where $C(iT)$ describes the relevant concentration at time point i, given a sampling interval of T, and SI_i and $SI_{precontrast}$ are the average signal intensities in the given region of interest at time point i, and prior to contrast enhancement respectively. The same calculation was performed for each voxel in the liver.

2.2.3.3 Evaluation metrics. k_1 maps were obtained from the HTR DCE series using both the established DITC model and the proposed LSITC model. The k_1 maps obtained from the DITC approach were used as a reference standard in the evaluation of the LSITC approach.

The first evaluation was to assess the similarity and deviation between the two resulting k_1 maps within the liver. The similarity was tested by the linear correlation coefficient between the two k_1 maps. The deviation was evaluated by the weighted mean absolute percent error (WMAPE), where we define the voxel-wise weighted absolute percent error (WAPE) as

$$\frac{|reference(i)-estimate(i)|}{\frac{1}{N}\sum_{n=1}^{n=N} reference(n)}. \tag{2.12}$$

Where i and n are voxel indices, and N is the total number of voxels considered. It should be noted that this metric places higher weight on accuracy for larger measurements. In this case the DITC model uptake rates are the reference values.

The second evaluation was to assess the validity of the LSITC model as applied to the clinical multi-phase LTR MRI data. Due to the low temporal resolution of approximately 20 seconds, the arterial input function peak could be missed or averaged over 20-second sampling. The sampling of the arterial peak affects the integral of the arterial input function in equation (2.10). Considering that the integral is over a long time period of 10 to 20 minutes, the effect of the arterial peak on the k_1 estimation could be small. To evaluate it, a tissue concentration curve with a temporal resolution of 1/s was simulated by direct application of the DITC model, subject to the input of reasonable parameter values and blood concentration curves. To mimic the LTR multi-phase MRI data, a subsampled curve was created by removing all points after the 1st pre contrast point and prior to the peak of the arterial input function. This curve was used to assess the error inherent in neglecting to sample the upswing in the arterial function, even when the image was perfectly timed to correspond to the peak. In order to evaluate the additional bias incurred by mistiming the peak, an additional concentration curve was created by removing all post-contrast data until 20 seconds after the arterial peak. This mimics a 20-second delay in the ideal time to image the arterial peak. For each of these three cases, the primary metrics were the correlation coefficients and WMAPE. The error in the k_1 estimation could represent an upper bound on the error incurred by missing the arterial peak, since acquisition delays longer than this would be easily visually recognizable.

The third assessment involved the application of the LSITC approach to the clinical multi-phase MRI data and comparison to global liver function as measured by the ICG retention rates. These clinical scans were much sparser temporally than those seen in the HTR DCE scans and so fitting

with the full DITC model was impractical. The metric for evaluation was the correlation coefficient between the sum of the estimated K_1 values over the contoured liver volume by the LSITC approach and the log of the ICG retention score, where $K_1 = k_1 V_{dis}$. In the sum of K_1 over the whole contoured liver volume, outliers having K_1 values above the 95^{th} or below the 5^{th} percentiles were rejected to remove edge effects. Additionally, large vessels were excluded by rejecting voxels where v_{dis} was greater than 0.4. The inclusion of the volume term ensures both uptake rate and plasma flow are accounted for. The summation of a regional measure of liver function to allow comparison to a global function such as ICG is not new.[28]

The final assessment involved comparison of the k_1 values obtained in the HTR DCE, and LTR multi-phase MRI data sets to reported values of k_1 in the liver from prior studies.

2.3 Results

2.3.1 k_1 estimations by the LSITC model vs. DITC model

Maps of k_1 values estimated from the liver HTR DCE scans using both the LSITC and DITC models are shown in Figure 2.3.

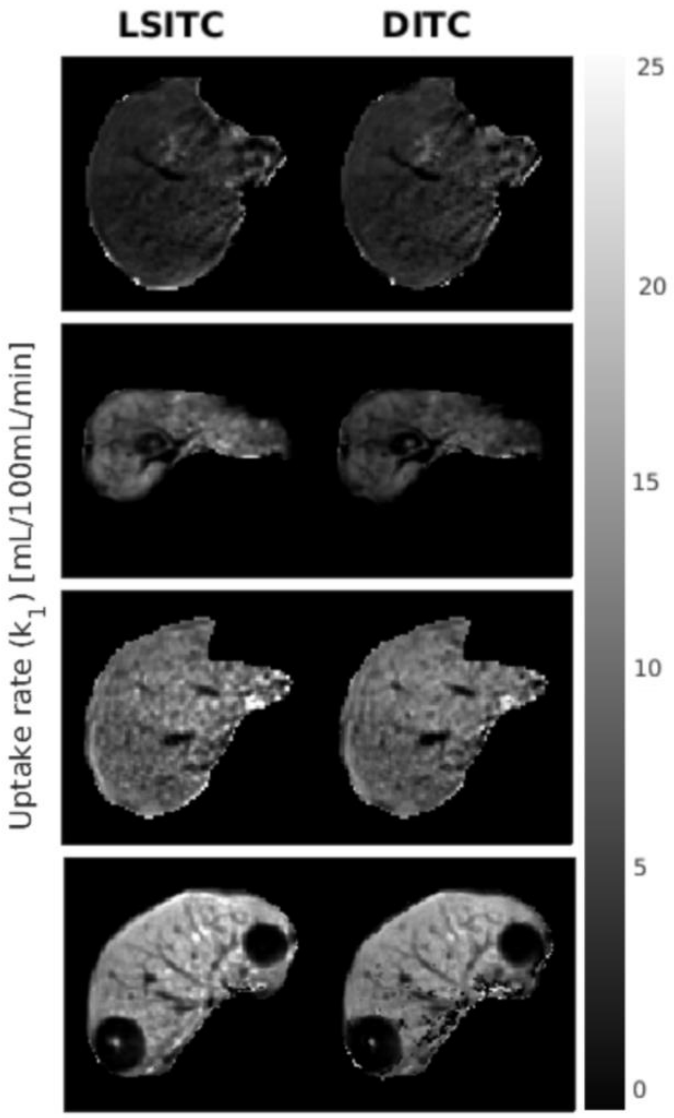

Figure 2.3*. Example slices of the* k_1 *maps estimated from the HTR DCE scans of four patients by the linearized single-input two-compartment model (left) and dual-input two-compartment model (right). Note that the units are mL/100mL/min.*

The linear correlations between the k_1 maps estimated by the two models are shown in Figure 2.4. The correlation was calculated in a randomly selected 5000 voxels within each liver, restricted to the voxels where v_{dis} was above the 25th and below the 75th percentiles, and the k_1 values were greater than 0.01 mL/100mL/min. (In one patient who was a candidate for the liver transplant and had very poor liver function, there was a small volume in the liver that had the non-zero k_1 values, leaving little volume for analysis. Thus, this patient was excluded from the analysis described here). The correlation coefficients ranged from 0.98 to 0.76 with a median of

0.91. The WMAPEs ranged from 9.0% to 39.4% with a median of 17.2%. Note that either estimate can be considered as the ground truth.

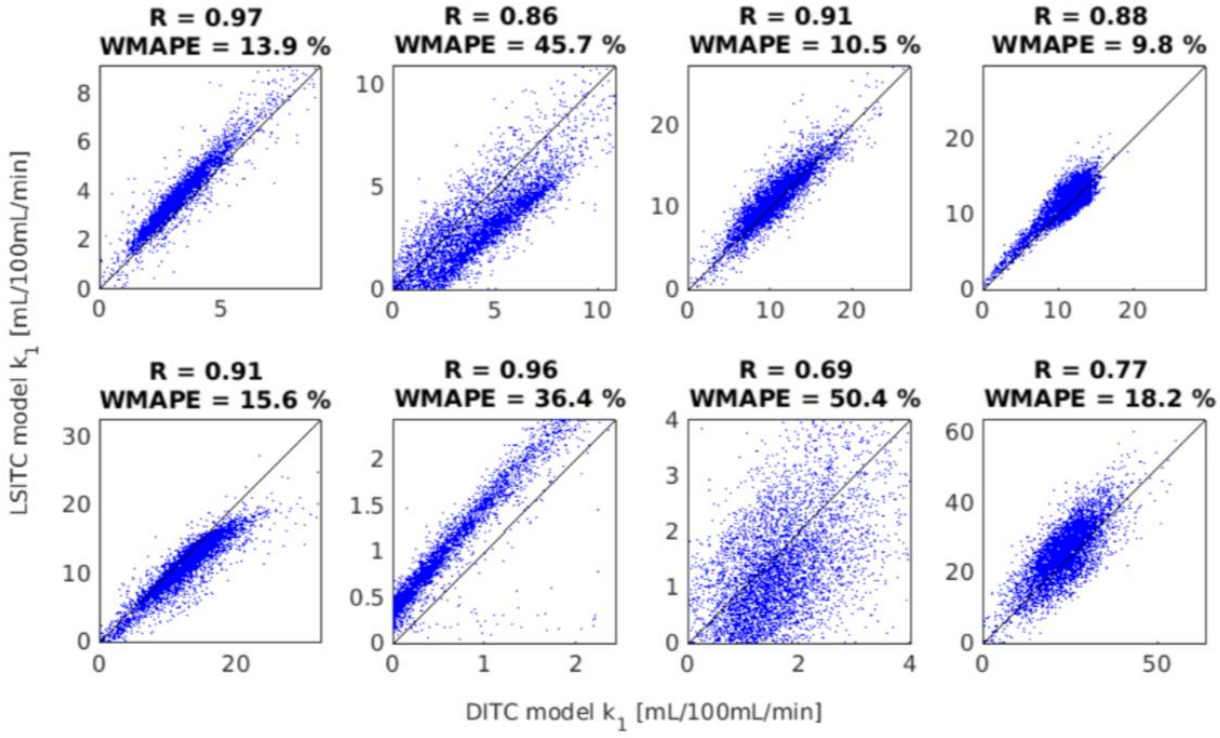

Figure 2.4. *Scatter plot of the* k_1 *values estimated by the two models for the 8 patients. The* k_1 *values in the horizontal axis were estimated from the dual-input 2-compartment model and the ones in the vertical axis from the linearized single-input two-compartment model.*

The estimated transition times (t_p) when the varied from 10 to 313 seconds as measured from the peak of the arterial input function, with a median value of 58 seconds. The computation speed of the LSITC approach was approximately 1000 times faster than the dual-input two-compartment model, taking a few seconds per exam while the DITC model fitting took several hours per exam.

2.3.2 k_1 estimation by simulation of missing the arterial peak

To simulate DCE data, composite input curves were first created by averaging the respective AIF and PVIF after matching the arterial peaks and re-sampling the time curves from all patients. The composite functions converged approximately 75 seconds after the arterial peak. To reduce noise the input data at least 5 min subsequent to the arterial peak was replaced by a double exponential fit.

Simulated parameters were randomly chosen but uniformly distributed over the ranges shown in Table 2.2, while τ_a and τ_{pv} were assumed to be 0 seconds. Using these values and the composite input concentration curves (AIF and PVIF), the resulting tissue concentration curves were calculated using the dual-input two-compartment model.

***Table 2.2**. The ranges of the parameter values for the simulation, where $k_{pvp} = k_{pv}(1\text{-}Hct)$ and $k_{ap} = k_a(1\text{-}Hct)$.*

Parameter	Minimum Value	Maximum value	
v_{dis}	10	20	%
$k_{pvp} + k_{ap}$	*50*	*300*	mL/100mL/min
k_{pvp}	*0.5($k_{pvp} + k_{ap}$)*	*($k_{pvp} + k_{ap}$)*	mL/100mL/min
k_1	0	0.1*($k_{pvp} + k_{ap}$)*	mL/100mL/min

To assess the potential error and variance incurred by missing the arterial peak during the clinical multi-phase MRI scan, the resulting tissue and arterial functions (Figure 2.5a) were then subsampled to remove all points prior to the arterial peak, apart from one pre contrast reference scan (Figure 2.5b). Furthermore, the data points acquired 20 seconds after the arterial peak (including the peak) were removed from the simulated curve and AIF to mimic a possible further delayed acquisition in the clinical data (Figure 2.5c).

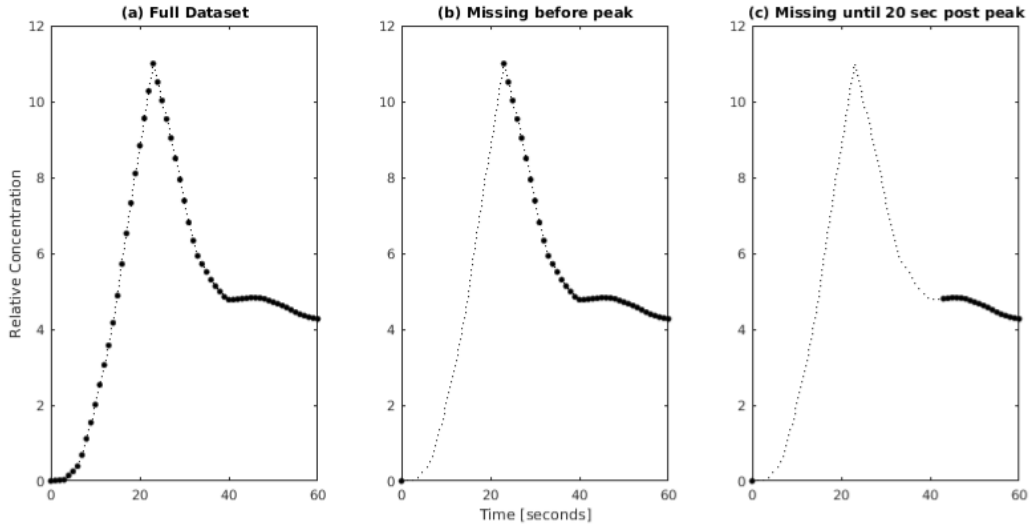

***Figure 2.5.** Early section of the AIFs to demonstrate subsampling used to investigate delays in acquisition. a) shows an AIF with the full dataset, b) shows an AIF with missing data points prior to the arterial peak, and c) shows an AIF with missing data points up to 20 seconds after the arterial peak.*

For each scenario t_p was assumed to be 60 seconds post peak. The simulation was run 10,000 times per case. A strong correlation was found between the LSITC results and input simulation

values. Without noise, correlation coefficients remained above 0.99 for all three cases including the case with an acquisition delay of 20 seconds after the arterial peak.

Results showed similar distributions for estimates of k_1, regardless of the acquisition delay (see Figure 2.6 and Figure 2.7). The WAPE in the noise free simulations was 4.7 ± 3.2% (mean ± standard deviation) for the full dataset, and then shifted to 6.1 ± 3.6% for the data missing time points before the arterial peak and to 6.7 ± 2.4% for the data missing time points up to 20 s after the arterial peak. The WAPE changed little when introducing white Gaussian noise to the generated C_t functions. Note that missing the early time points in the dynamic curves caused a maximum change in the WAPE of 2% (from 4.7% to 6.7% without noise), suggesting other effects predominate in the k_1 estimation errors. Based on these results we would expect similar levels of systematic error resulting from delayed or averaged capture of the arterial peak in the LTR multi-phase data.

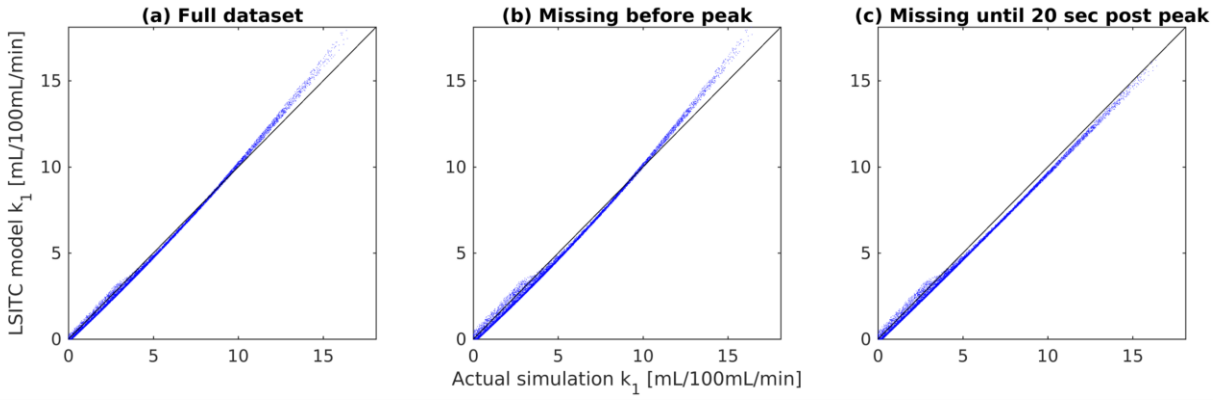

Figure 2.6. *Plots of the input* k_1 *values against the* k_1 *values obtained by LSITC model. The correlation coefficient was R = 0.999, 0,998 and 0.999 for the full dataset (a), the dataset missing time points prior to the arterial peak (b), and the dataset missing time points up to 20 seconds after the arterial peak (c) respectively.*

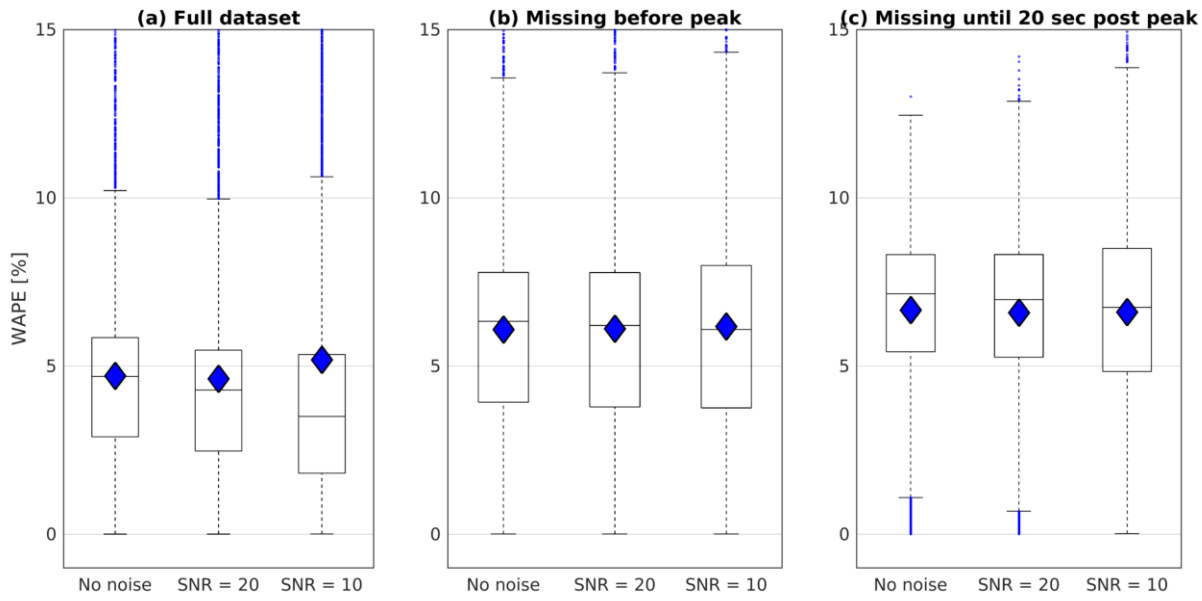

Figure 2.7. *Boxplots for the WAPE in the LSITC model results relative to the simulated values at 3 noise levels for the three simulated datasets: a) the full dataset, b) the dataset missing data points prior to the arterial peak, and c) the dataset missing data points up to 20 seconds after the arterial peak. The median values are indicated by the horizontal line, the mean is indicated by the diamond, and error bars are 1.5 times the interquartile range.*

2.3.3 k_1 estimation from the clinical MRI and comparison with ICG retention

Maps of k_1 and v_{dis} were estimated from the clinical LTR multi-phase MRI of 27 exams using the LSITC model. Example k_1 and v_{dis} maps are shown in Figure 2.8. For the 20 clinical scans with accompanying ICG retention scores, the scores were compared to the K_1 values ($K_1 = k_1 V_{dis}$) summed over the contoured liver volume (Figure 2.9). Voxels with v_{dis} greater than 0.4 were rejected to omit vasculature, and K_1 values less than the 5th or greater than the 95th percentile were rejected to avoid outliers and edge effects. Since the log of ICG retention is inversely proportional to the rate of clearance in the liver, a linear relationship is expected. The sum of K_1 values were significantly correlated to the log of ICG retention values with a correlation coefficient of R = -0.72 (p=0.0004, n = 20).

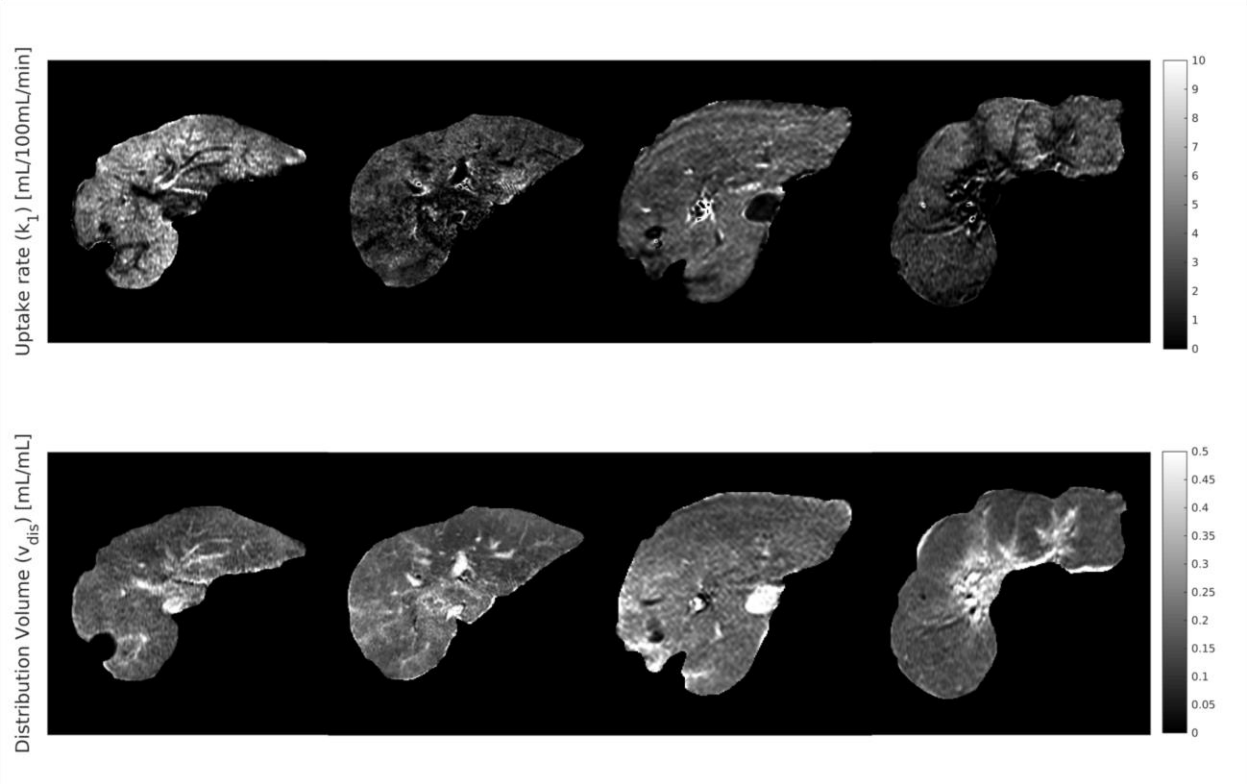

Figure 2.8. *Example maps of k_1 and v_{dis} in the assessed livers. Notice that high values of v_{dis} can be seen to correspond to vasculature.*

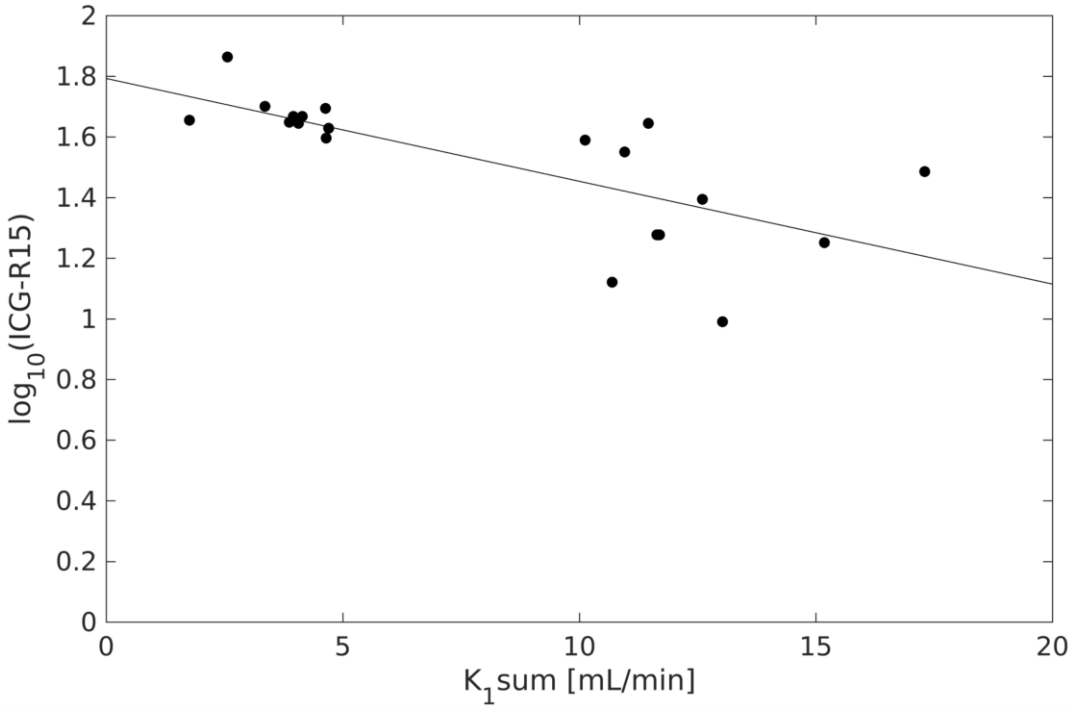

Figure 2.9. *Plot of the log of ICG retention at 15 minutes against the sum of K_1 values in the contoured total liver volume. R = -0.72 (p = 0.0004, n = 20).*

2.3.4 Comparison to Literature values

Maps of k_1 values were generated for the 27 multi-phase LTR liver scans by the LSITC model. In each of the 27 k_1 maps the mean value was calculated over the volume with v_{dis} less than 0.4 and with k_1 values in the central 90th percentile of k_1 value for the liver. The mean value across all multi-phase LTR scans was 3.93 ±1.79 mL/100mL/min. Using the same process for the HTR DCE based maps the mean uptake was measured as 9.17 ± 8.23 mL/100mL/min, across all HTR DCE scans, and 7.44 ± 4.93 mL/100mL/min after removal of an outlier more than twice as large as any other mean uptake. It should be noted that the outlier was also the scan with the shortest total duration.

This is fairly consistent with the prior literature. Previous studies have found mean uptake rates of 3.4 ± 2.1 in background regions of livers with metastases, 3.03 ± 2.1 in cirrhotic livers, and 6.53 ± 2.4 in healthy livers, as seen in Figure 2.10[29,30].

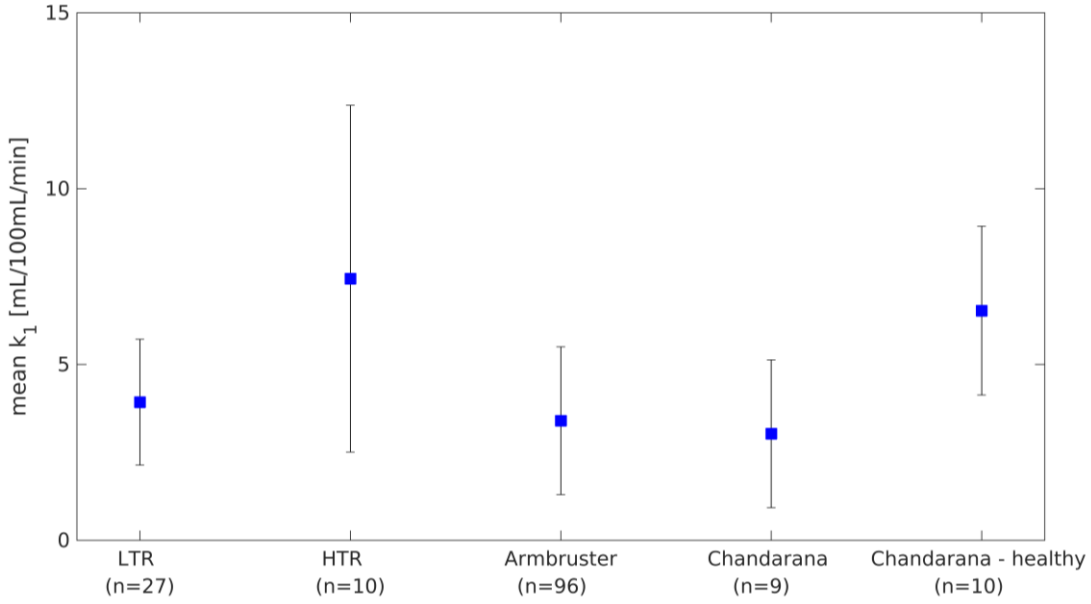

Figure 2.10*. This plot shows the mean values obtained in several studies for background liver uptake rate of gadoxetic acid in cancerous livers, and a group of healthy livers for reference. Error bars indicate the standard deviation. The HTR values are based on a removed outlier.*

2.4 Discussion

This chapter developed a LSITC model to quantify k_1 from dynamic gadoxetic acid enhanced MRI in the liver, and evaluated the results by comparisons to an established uptake model, and measure of whole liver function. The k_1 values estimated by the LSITC and DITC analysis of

the HTR DCE MRI data had a close median correlation (R = 0.91). Application of the LSITC approach to LTR multi-phase MRI data gave similar results to prior studies and correlated relatively well (R = -0.72) with the results of ICG retention exams. Furthermore, concerns about the impact of delayed imaging of the arterial peak were addressed by simulations showing less than 3% related error. The results indicate that the LSITC model is a simple analog to the dual-input two-compartment model, and correlates well with independent scores of liver function. Since this technique can be applied to clinically typical multi-phase data, it presents the possibility of quantitative liver assessment without large changes to existing clinical workflow.

There are several possible sources of error in the determination of k_1 via the LSITC model set forth here. It should first be noted that unlike the Patlak model, the ratio of k_1 to k_2 should not impact the estimate of k_1, since the k_2 dependent term cancels in equation (2.10) when dividing slope by intercept. However, this term will impact the estimate of v_{dis} and K_1, with both underestimated. This also suggests a more complete linear model where equation (2.8) is fit to determine k_1, k_2 and v_{dis}, allowing a more accurate estimate of K_1 (see appendix B).

The model assumes that after t_p, $C_{pv} = C_a$. In most of the measured input curves C_a and C_{pv} were very similar within less than a minute of the arterial peak. In some cases this held true after a constant correction factor, i.e. $C_{pv} \times$ constant = C_a. This could reflect differences in *Hct* between C_a and C_{pv}. In quantification of C_a and C_{pv}, native T_1 of plasma was not considered since images for T_1 quantification are almost never acquired in clinical liver scans. However, native T_1 of plasma should be the same for C_a and C_{pv}. This omission should not contribute to the difference between C_a and C_{pv}. In general, it is anticipated that the peak of C_{pv} is shallower and broader than that of C_a, since the bolus of contrast travels through the vascular systems of intestines and spleen before it slowly returns to the portal vein through the mesenteric vein and splenic vein. In some patients, this contrast return seems to be slower than in others. This could be a factor contributing to difference between C_a and C_{pv} even after the typical time t_p. However, partial volume effects and motion could also affect the measurement of C_{pv}, causing apparent differences between C_{pv} and C_a that do not reflect the actual contrast concentrations. The simplified form of the model also neglects dC_a/dt from equation (2.8). Since dC_a/dt will generally be negative, and decreasing in magnitude after t_p, we expect this to result in a small overestimate of v_{dis} and k_1. The possibility of flow related enhancement causing a

mischaracterization of AIF can be considered as well. Selecting the AIF values from an Aorta contour within 3 cm of the aortic split to the liver ensured the blood in the voxels used were far from the edge of the FOV, and had experienced multiple repetitions within the excitation volume. This minimizes the impact of flow related enhancement on the AIF, making this an unlikely source of error in AIF, or cause for differences between AIF and PVIF. The estimate of x_p could also impact the final solution. Choosing x_p too early would be expected to cause an underestimate of v_{dis}, and an overestimate of k_1. In the absence of noise even a very late estimate of x_p should have little impact on the result. However, with noise we would expect additional uncertainty in the result as we estimate the slope and intercept from fewer data points.

In the LTR clinical data we assumed the last three points after enhancement were after t_p. This was necessary to have an overdetermined solution, but may not have been accurate in some voxels or livers. The time post arterial peak for the first point in the fit varied between 36 and 290 seconds. The median time was 48 seconds. This can be compared to the optimal t_p times chosen in the HTR data, which had a median of 65 seconds. It should be noted that the peak in the LTR clinical data was assumed to coincide with the start of the first post contrast image, so it is possible that there is a hidden delay relative to the physiological arterial peak. Another possible source of error is the sparsity of the clinical data which causes an underestimate of the integral of C_a, and thus an underestimate of the values for x. We would expect this error to increase as the timing of the 1st post contrast measurement was delayed past the peak, but the change in error in the simulation was relatively minor, indicating other effects predominate.

Both the LSITC and DITC models omit several notable features. Firstly, the models omit the excretion of contrast from the hepatocytes into the bile. It was assumed that the rate of excretion is negligible over the timeframe of the exams. Deviations from this assumption would cause error in k_1 and thus impact the total functional estimate. However, this would not impact the comparison between the two models, or effect error in the simulated case. Secondly, the DITC model includes only one extracellular compartment, which may not be valid in tumors or other pathological tissues where movement between the capillary bed and the space of Disse is relatively slow. This would again impact both models. The impact could be assessed by comparison to a dual-input three compartment model.[31]

Contrast concentration for all exams was calculated using relative enhancement. Relative enhancement has been found to correlate linearly with the concentration in a given tissue, though this relationship breaks down at sufficiently high contrast concentrations.[31] However, even if we assume perfect linearity and that all plasma and all liver voxels had uniform respective native T_1 values, the direct use of relative enhancement as relative concentration will introduce a constant bias term in the uptake rate based on the native T_1 in plasma relative to liver tissue. This would not impact the correlation with liver function and could be fixed through a correction constant with knowledge of the ratio of native liver and plasma T_1. If we further consider differences in native T_1 across the liver we would expect additional error even in the relative voxel wise uptake rate. This could be fixed by characterizing the precontrast T_1 with an additional sequence before contrast injection. Ideally this would allow for more accurate quantification of concentration, but does introduce clinical inconvenience and complicates analysis.

In all real data; noise, motion and other random variation contributed to error in the input curves, and the tissue curves. Random variation will be especially harmful in cases where relatively few data points are used in the estimate of k_1. Motion effects are especially apparent at the edges of the liver and liver vasculature. In these locations slight motion can cause apparent jumps in uptake as a motion artefact.

Further work can be done to improve k_1 quantification. For example, the impact of noise and motion could be lessened by the introduction of spatial regularization to the creation of the k_1 map. The full model from equation (2.8) can be used if k_2, or v_{dis} are parameters of interest, though it should be noted that the LSITC model cannot replace the DITC model when arterial or portal venous perfusion are parameters of interest. An evaluation of the impact of omission of native T_1 on k_1 estimation across the population of patients will be conducted. Additional work should also use larger and more varied data sets to further characterize the relationship between liver function and uptake as measured by the LSITC model. Further analysis should also consider tissues in which the DITC model is insufficient and include comparison to a dual-input three compartment model.[31]

2.5 Conclusion

This Chapter proposes and validates the LSITC model for assessing liver function based on the uptake rate of gadoxetic acid. Validation was obtained relative to the[31] predictions of the

accepted dual-input two-compartment model, and independent measurements of whole liver function. The LSITC approach allows the creation of a spatially resolved quantitative image of liver function, using standard clinical acquisitions, and reduces the dependence on impractical, extremely high temporal resolution scans.

CHAPTER 3

Validation of Gadoxetic Acid Uptake Rate as a Measure of Liver Function

Significant portions of the results described in this chapter were previously described by the author in an abstract for the 2020 ASTRO annual meeting.[32,33]

3.1 Introduction

Global liver function has been shown to be an important clinical prognostic factor for RILD as well as overall survival in patients with cirrhosis and hepatic cancers[1,3,34,35]. Liver injury during radiation therapy (RT) of hepatic cancers has been minimized using normal tissue complication probability (NTCP) models by limiting mean liver dose[1–3]. To further minimize risks of liver injury, individualized adaptive RT strategies have been suggested and investigated. Using this strategy, patients cannot be assumed to have the same liver function and the same response to radiation dose. Also, the distributions of hepatic function and dose-response in the liver volume cannot be assumed to be homogeneous[4,6–8,36]. These approaches require measures of global and regional liver function as well as radiation dose response in individual patients, and consider all these factors in treatment strategies and planning.

We hypothesize that global as well as regional liver function can be quantified from a single measure using gadoxetic acid uptake rate (k_1) from dynamic MRI. This chapter aims to validate k_1 as a measure of both global and regional function through comparison with two established global function measures, ICG retention at 15 minutes (ICGR15) and raw albumin–bilirubin (ALBI) score[5]. Additionally, uptake is compared to portal venous perfusion (k_{pv}) in relation to total function and spatial distribution.

3.2 Methods

3.2.1 Data acquisition

3.2.2.1 Patients. Seventy-eight DGAE liver MRI scans were acquired from 40 patients (median age of 64 years) with liver cancers in prospective protocols approved by the University of Michigan Institutional Review Board. All patients signed written consents. The patient demographic information is given in Table 3.1. The liver DGAE MRI scans were obtained pre-RT and one-month post-RT. Within no more than one week of the MRI scans, an ICG retention exam (a total liver function measure) was taken for all but 4 patients. In total, 69 pairs of the DGAE MRI scans and ICG retention scores were available for correlation analysis of global liver function.

3.2.1.2 Image acquisition. 3D volumetric DGAE MRI scans of the whole liver were acquired using a radial sampling VIBE sequence during the intravenous injection of a single standard dose of gadoxetic acid on a 3T scanner (Skyra, Siemens Healthineer). The free-breathing DGAE images of the liver used a 3D golden-angle radial stack-of-stars VIBE sequence that over-samples the center of k-space and is resilient to motion effects[37]. Acquisitions included image volumes from before contrast introduction to 4 to 31 minutes after the initial arterial peak (median 18 minutes). Dynamic image volumes of the whole liver were reconstructed with a temporal sampling rate of 3.4 to 15.2 seconds per volume (median 8.8 seconds), and 64 to 80 slices with thicknesses from 2.6 to 4.8 mm and an in-plane resolution of 2.1×2.1 mm to 2.6×2.6 mm (192 by 192 pixels).

3.2.1.3 ICG retention test. The ICG retention exam used a single dose of ICG with blood sampled at 5, 10, and 15 minutes following injection to measure the fractional retention of ICG. ICG is cleared almost entirely in the liver[38], making measures of ICG clearance a common surrogate for global liver function[6]. ICGR15 was used to test the global liver function calculated from the DGAE MRI scans. See reference [36] for a more in-depth description of ICGR15 procedure.

3.2.1.4 ALBI. The ALBI score is computed using serum bilirubin and albumin levels using the equation given by Johnson et al[5]:

$$\text{ALBI} = (\log_{10}(\text{bilirubin}) \times 0.66) + (\text{albumin} \times -0.085) \tag{3.1}$$

with bilirubin in μmol/L and albumin in g/L. Though this score can be used rate the patient's liver function with a discrete ALBI grade, the raw ALBI score is a continuous measure. This raw ALBI score was used as an additional measure to test the global liver function calculated from the DGAE MRI scans.

Table 3.1. *The patient demographics for all 40 patients*

Characteristic	Value
Age (years): median (range)	64 (48 to 100)
Sex: female/male	9/31
Pre-treatment Cirrhosis (%)	82.5 (33 of 40)
HCC (%)	85.0 (34 of 40)
Liver Volume (L): median (range)	1.80 (0.69 to 3.99)
GTV Volume (mL): median (range)	43 (2.7 to 1251)
Child-Pugh Score: median (range)	6 (5 to 10)
Raw ALBI Score (n=37)	
Pre-RT: median (range)	-2.2 (-3.6 to -1.0)
Post-RT: median (range)	-2.0 (-3.27 to -1.0)
ICGR15 (%) (n=35)	
Pre-RT: median (range)	29.2 (4.0 to 82.3)
Post-RT: median (range)	40.6 (7.5 to 82.3)

3.2.2 Image processing and analysis

3.2.2.1 Image pre-process and model fitting. The time-series DGAE-MRI volumes were co-registered within the liver contour using an over-determined, rigid-body transformation approach[15]. Then, both DITC and LSITC models were applied to the registered time-series DGAE-MRI volumes to estimate k_{pv}, k_1 *and* v_{dis}, using respective NLLS and LLS fitting implemented in Matlab, as described in chapter 2.

3.2.2.2 Similarity between perfusion and uptake rate maps. To assess whether portal venous perfusion (k_{pv}) could be used as a surrogate for liver function, the correspondence between uptake rate (k_1) and portal venous perfusion (k_{pv}) was investigated using the ranked Spearman

correlation. Both k_{pv} and k_1 maps were obtained from NLLS fitting of the DITC model. Additionally, the expected differences of vasculature in the evaluation should be excluded, so all regions with a blood distribution volume (v_{dis}) >0.25 were omitted from comparison. A binned comparison of quantiles of the two maps was performed for visual comparison of the relative distribution of the measures. To this end, all maps were divided into 100 quantiles, and normalized to 0 through 1 for inspection.

3.2.2.3 Global liver function measures derived from imaging. If global and spatially-resolved liver function measures could be derived from a single measurement, e.g., imaging, it would reduce redundant measures. There are several plausible or proposed ways of constructing a global measure of liver function from imaging. A simple measure would be the mean k_{pv} in the liver[13]. A total functional volume would be calculated by incorporating the liver volume with the uptake rate (k_1) as a sum of $k_1 V_t$ across the liver, or equivalently mean k_1 by total volume: $\bar{k}_1 V_L$. A prior study suggests that the blood distribution volume that provides the surface area for the gadoxetic acid uptake by parenchyma might be considered in the functional volume computation, instead of using the absolute liver volume[22]. This measure would be expected to correspond to the total rate of gadoxetic acid uptake as a sum of $K_1=k_1\times V_{dis}$. Finally, a total liver volume, which is a conventional measure used in liver resection[39], could be included for comparison. All four imaging derived candidates of global liver function measures were tested by a linear regression with a well-established global liver function measure, ICGR15, as well as a more recently established measure: ALBI score[5]. Also, these linear regressions serve as a calibration measure to compute global liver function from DGAE MRI. Since the rate of ICG clearance is proportional to the log of ICG retention, a linear relationship would be expected between the log of ICGR15 and a measure of liver function.

The initial three tested global liver function measures: 1) mean $k_1\times$total volume (mL/min), 2) summed K_1, and 3) mean k_{pv} were calculated in the contoured liver volume with uptake rates > 0 mL/100mL/min, and distribution volumes $0.02<v_{dis}<0.25$ to exclude the liver contour variation and blood vessel. The final measure, 4) the volume, included the entire contoured liver volume minus the gross tumor volume (GTV). Since several of these measures is based on a totaled

capacity in the liver (1,2,4), it may be of use to quantify this relative to patient size. To this end, each measure was also compared to ICGR15 and ALBI score after normalization by patient mass

3.3 Results

3.3.1 Similarity between perfusion and uptake rate maps

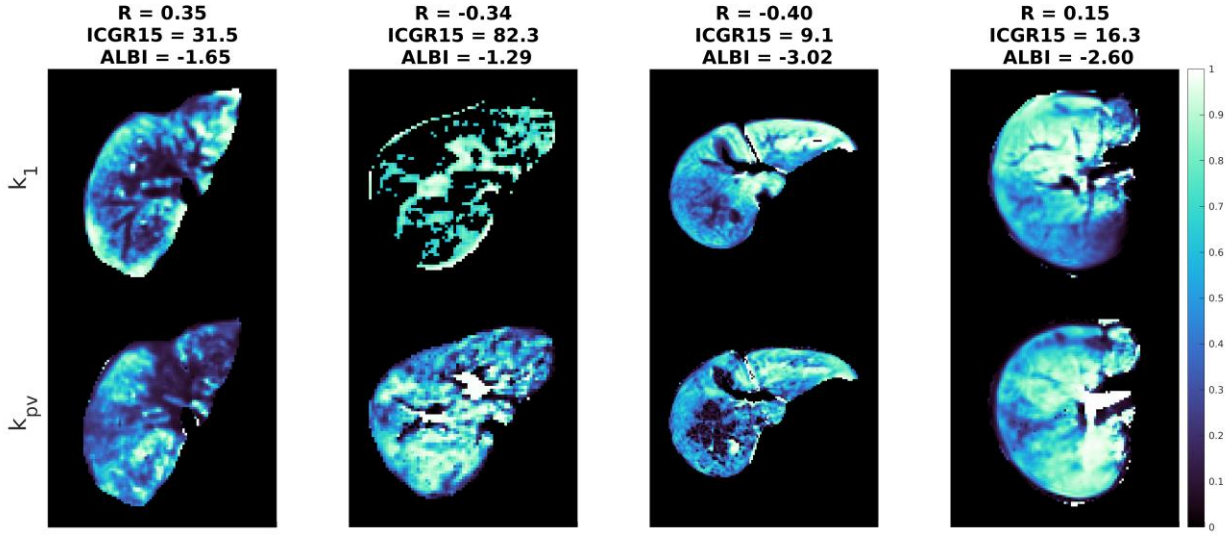

Figure 3.1. *Example slices of* k_1 *(top row) and* k_{pv} *maps (bottom row) presented in 100 quantiles of 4 patients with different ICGR15 and ALBI values. Note that large vessels generally show as dark in uptake rate maps (lacking hepatocytes), while in perfusion maps they can be either dark (arterial vessels) or bright (portal venous vasculature), depending on the type of vessels. Note that* k_1 *maps were derived from the LSTIC model while* k_{pv} *maps were derived from the DITC model.*

Portal venous perfusion and uptake rate varied greatly between patients as well as within patients. Example slices of k_1 and k_{pv} maps of 4 patients with different ICGR15 scores are shown Figure 3.1. The four patients had ICGR15 of 31.5, 82.3, 9.1 and 16.8%, ALBI scores of -1.65, -1.29, -3.02, and -2.60, and correlation coefficients of 0.34, -0.34, 0.27, and 0.15, respectively. Note that the differences were pronounced on a larger scale in some patients, especially for the patients with very poor liver function. The second left column in Figure 3.1 shows that a patient with ICGR15 and ALBI of 82.3% and -1.29 had a negative correlation of -0.34 due to negligible uptake rate of gadoxetic acid throughout the majority of the liver, but relatively healthy perfusion through the liver.

In 26 of the 79 DGAC MRI scans, k_{pv} maps had extremely low values with local variation resembling noise. In these cases we observed a mean $k_{pv} < 0.005$ mL/min/(100mL) , but typically much lower, compared to typical values of 10 to 60 mL/min/(100mL) from other

patients. It is conceivable that this was caused by the existence of cirrhosis and HCC in the patients or, more likely, by unreliable estimation of k_{pv} in the cases with poor hepatic perfusion using one standard dose of gadoxetic-acid (in which Gd counts are one quarter of the counts in one standard dose of Gd-DTPA) and using the DITC model. To make the comparison fair, the 26 instances with negligible perfusion were excluded from the analysis of the correlation between k_{pv} and k_1.

In the remaining 53 exams, the correlation was still poor on average. The mean ranked correlation between uptake rate and perfusion maps was R = 0.095 (median: 0.12 range: -0.45 to 0.56). Figure 3.2 shows the distribution of ranked correlations between k_1 and k_{pv} values. Note that there were only two exams that showed a positive correlation greater than 0.4. Nearly a third to one half of the exams had negative or near zero correlation between k_1 and k_{pv} in the liver, indicating a mismatch or non-match between hepatic perfusion and liver function.

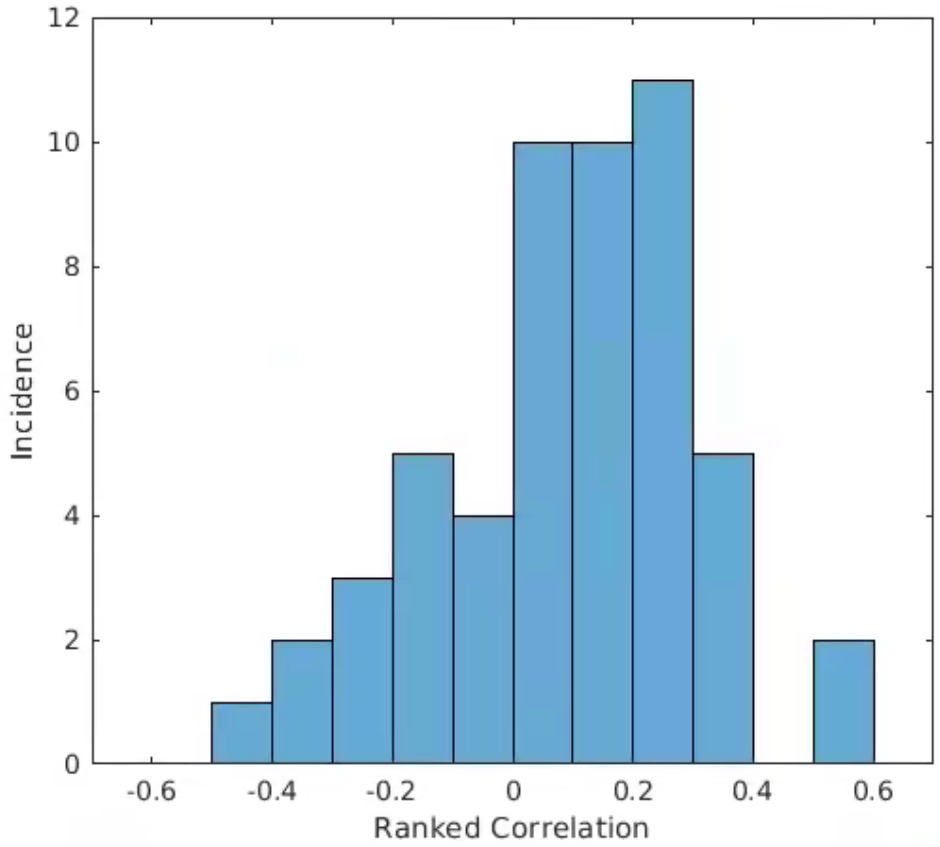

Figure 3.2. *histogram of the distribution of ranked correlations between k_1 and k_{pv} values (mean: 0.095, median: 0.12, n=53).*

3.3.2 Imaging derived measures of total liver function compared to ICG and ALBI

To determine which imaging derived measures were reliable as a measure of total liver function, these measures were compared to an independent measure of total liver function, ICGR15.

3.3.2.1 Uptake derived measures. The total function volume (measure 1 from *3.2.2.3)*, quantified by the product of the mean k_1 in the liver and the total liver volume (determined using the LSITC model) had a strong correlation with both ICGR15 and ALBI (R = -0.67 with 95% confidence interval (95% CI) of [-0.79 -0.53] and R = -0.73 (95% CI) [-0.82 -0.59] respectively), see Figures 3.3a and 3.4a. To calibrate the total function volume against ICGR15, a linear regression model was tested after the removal of an outlier (k_1 >3 median absolute deviations from the median), see Table 3.2. Normalization by patient mass resulted in strengthened correlations (R=-0.76, 95% CI [-0.84 -0.64] and R=-0.75, 95% CI [-0.84 -0.62] for ICGR15 and ALBI respectively), with the corresponding fits given in Table 3.2.

The summed K_1 (measure 2 from *3.2.2.3*, accounting the total blood distribution volume, instead of the total liver volume) correlated well to ICGR15 and ALBI (R=-0.61, 95% CI [-0.73 -0.42], and R=-0.71, 95% CI [-0.81 -0.57]) but weaker than expected for ICGR15 based on preliminary studies[22] or the apparent analogy between total K_1 and ICG uptake rate. There appeared to be approximately 10 exams that had $\log_{10}$(ICGR15) values near 1 but had a systematic deviation from the regression line, see Figure 3.3b. This pattern was not seen in the corresponding ALBI scores. The linear regression models are given in Table 3.2. Normalization by patient mass again resulted in strengthened correlations (R=-0.69, 95%CI [-0.79 -0.54], and R=-0.74, 95% CI [-0.84 -0.63]).

3.3.2.2 Perfusion derived measure. The correlation between mean k_{pv} (measure 3 from *3.2.2.3)* and both $\log_{10}$(ICGR15) and ALBI was poor. 23 exams were removed from the correlation analysis due to a failure in fitting k_{pv} as seen by extremely low k_{pv} values across the liver volume. In these 23 exams, there were no voxels that had k_{pv} values greater than 0.6 mL/(100mL min) and less than 300 mL/(100 mL min) (artifacts). Without these two limits in place the fit was dominated by extreme outliers. As seen in Figures 3.3d and 3.4d and Table 3.2, there was not a significant correlation between mean portal venous perfusion and ICGR15 or ALBI (R=0.03, 95% CI [-0.26 0.32] and R=0.06, 95% CI [-0.24 0.35]).

3.3.2.3 Volume derived measure. Liver volume (measure 4 from *3.2.2.3)* showed a moderate correlation with $\log_{10}$(ICGR15) (R = -0.30, 95% CI [-0.50 -0.06]), which was similar to preexisting studies[28,40,41](see Figure 3.3d and Table 3.2), and comparable correlations for ALBI (R = -0.40, 95% CI [-0.58 -0.18]) (see Figure 3.4d and Table 3.2). Only the ICGR15 correlation was strengthened by normalization by patient mass (R=-0.39, 95% CI: [-0.57 -0.17]). The correlation for mass normalized liver volume and ALBI showed an apparent strengthening of the linearity of the relationship for most points, but created several outliers that canceled out this effect (R=-0.40, 95% CI: [-0.58 -0.17]).

Table 3.2. *Linear Regression Models of Imaging Derived Total Liver Function Measures to ICGR15 and ALBI.*

Measure Used to Predict log10(ICGR15)	R	Slope (95% CI)	Intercept (95% CI)
$\bar{k}_{pv}$ [/s]	0.03	0.0838 [-0.64 0.81]	1.51 [1.26 1.75]
$\bar{k}_1 V_L$ [L/min]	-0.67	-2.97 [-3.63 -2.31]	1.74 [1.66 1.83]
$\bar{k}_1 V_L$/W [L/min/kg]	-0.76	-308.5 [-362.6 -254.6]	1.79 [1.72 1.87]
Summed K_1 [L/min]	-0.61	-14.85 [-19.29 -10.40]	1.67 [1.58 1.76]
Summed K_1/W [L/min/kg]	-0.69	-1749 [-2167 -1330]	1.73 [1.64 1.81]
V_L [L]	-0.30	-0.145 [-0.26 -0.03]	1.71 [1.50 1.92]
V_L/W [L/kg]	-0.39	41.84 [-56.44 -27.24]	2.21 [1.94 2.48]
Measure Used to Predict ALBI	R	Slope (95% CI)	Intercept (95% CI)
$\bar{k}_{pv}$ [/s]	0.06	0.3155 [-1.25 1.88]	-2.06 [-2.60 -1.52]
$\bar{k}_1 V_L$ [L/min]	-0.73	-6.7242 [-8.00 -5.45]	-1.45 [-1.60 -1.29]
$\bar{k}_1 V_L$/W [L/min/kg]	-0.75	-604.7 [-723.3 -486.1]	-1.44 [-1.60 -1.28]
Summed K_1 [L/min]	-0.71	-38.50 [-47.07 -29.94]	-1.56 [-1.72 -1.40]
Summed K_1/W [L/min/kg]	-0.76	-4215 [-5027 -3403]	-1.46 [-1.62 -1.30]
V_L [L]	-0.40	-0.38 [-0.59 -0.16]	-1.42 [-1.83 -1.01]
V_L/W [L/kg]	-0.40	-105.1 [-129.2 -81.11]	-0.23 [-0.67 0.22]

Least squared fitting of the Global measures (log(ICGR15) or ALBI) from the given measure using the equation: Global Measure = Slope×measure + Intercept. The fits ignore the outliers marked in figures 3 and 4. W represents the patient weight.

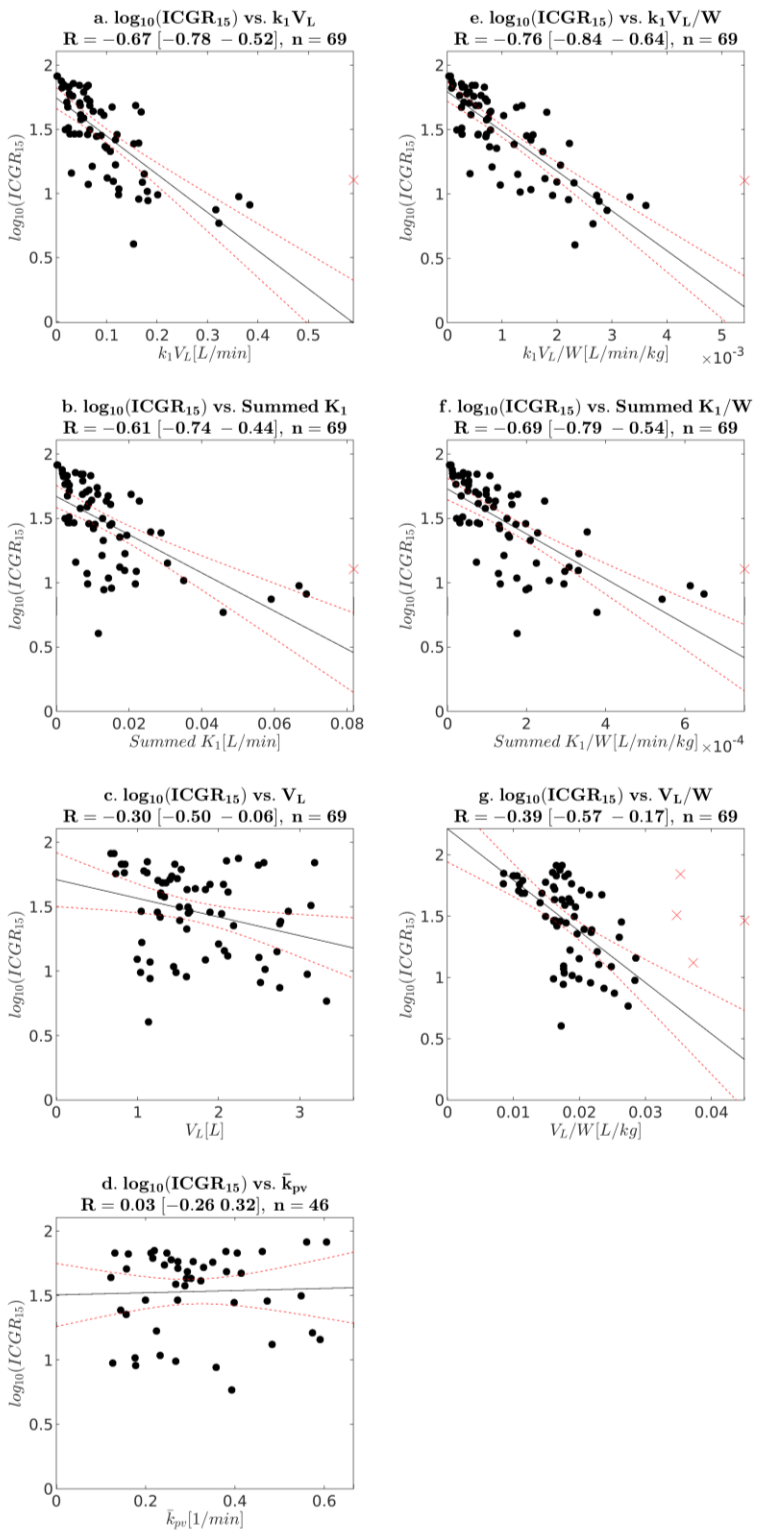

Figure 3.3a-3g *log10(ICGR15) plotted against: (a) mean k_1 by Total Volume, (b) summed K_1 = summed k_1V_{dis}, (c) total liver Volume outside GTV, (d) mean k_{pv}, (e) weight normalized mean k_1 by Total Volume, (f)) weight normalized summed K_1 = summed k_1V_{dis}, (g)) weight normalized total liver Volume outside GTV. The plotted linear regression fit in (a), (b), (e), and (f) each ignore one outlier in terms of k_1, and the fit for (g) ignores 4 outliers in terms of weight normalized liver volume. Correlation coefficients do not exclude outliers*

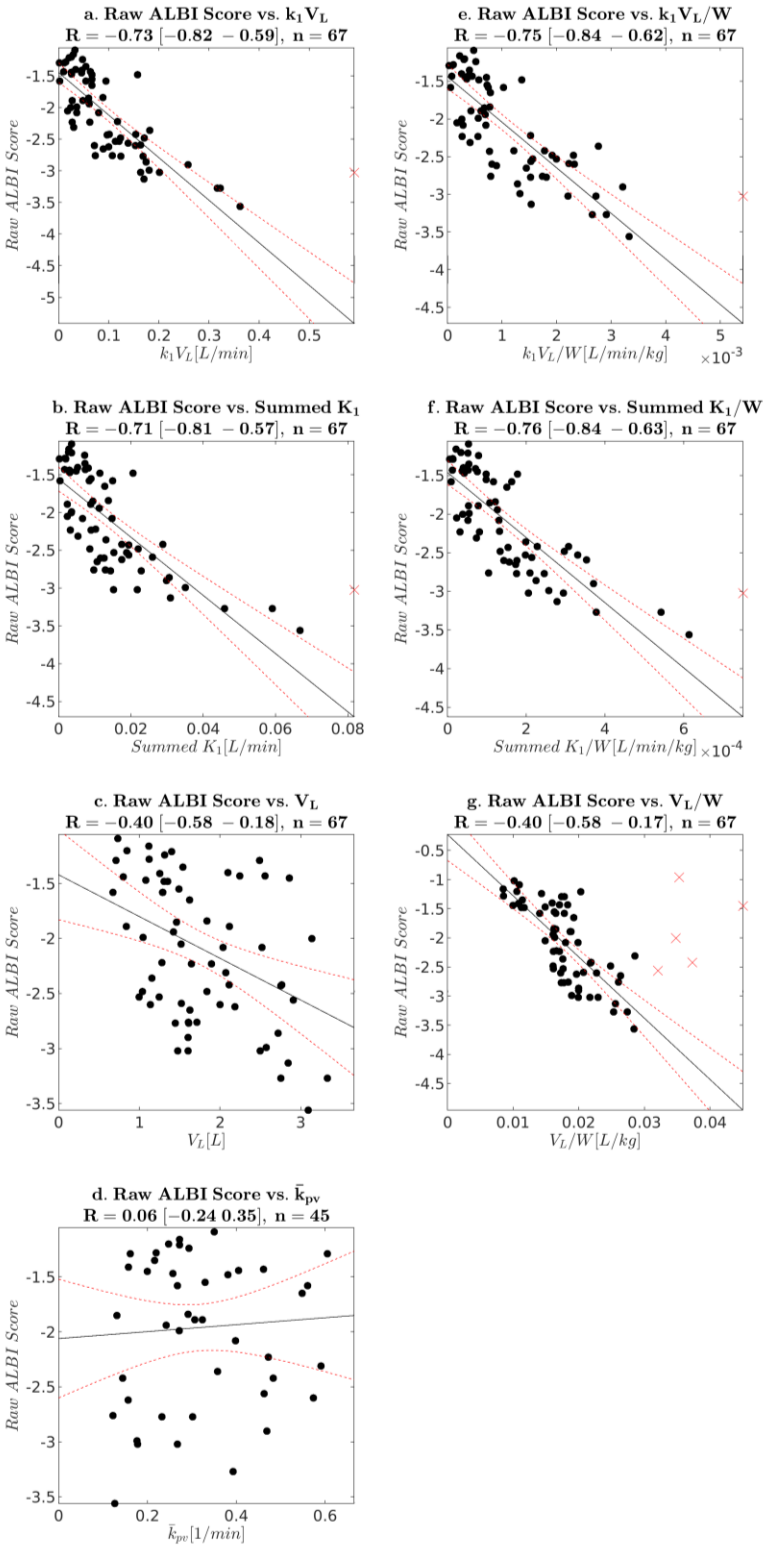

Figure 3.4a-4g. *The raw ALBI plotted against: (a) mean k_1 by Total Volume, (b) summed K_1 = summed k_1V_{dis}, (c) total liver Volume outside GTV, (d) mean k_{pv}, (e) weight normalized mean k_1 by Total Volume, (f)) weight normalized summed K_1 = summed k_1V_{dis}, (g)) weight normalized total liver Volume outside GTV. The plotted linear regression fit in (a), (b), (e), and (f) each ignore one outlier in terms of k_1, and the fit for (g) ignores 5 outliers in terms of weight normalized liver volume. Correlation coefficients do not exclude outliers.*

3.4 Discussion

This chapter evaluated four potential measures of liver function against two well regarded measures of global liver function. The analysis shows that the total function volume quantified from the DGAE-MRI, which accounts for both liver volume and gadoxetic acid uptake rate, is a better measure for global liver function than mean portal venous perfusion or liver volume, especially in the patients with a mis-match between hepatic perfusion and functional parenchyma. This single imaging technique can be used for measures of both global and regional liver function, and aid in liver function preservation during adaptive radiation therapy of hepatic cancers.

The results of this analysis underscore the differences between total functional volume and liver volume. While in most metastatic cases we expect liver function to be uncompromised and relatively homogeneous outside the region of the tumor, this is not a safe assumption for patients with primary liver cancer, and risks overestimation of functional liver volume. As previously mentioned, patients treated for HCC are likely to have preexisting cirrhosis, resulting in significant compromise to liver function outside regions directly affected by the tumor. In this work the correlation between liver volume and total function measured by ICG was similar to prior studies [28,40,41]. The correlation with ALBI was similar. There is obvious and undeniable evidence that non-functional volume cannot contribute to liver function. However, this is not the end of story without accounting for the extent of liver function. While global measures of function such as ICGR15 allow for patient specific prediction of functional reserve, RILD or liver toxicity, thus enabling adaptation in treatment[1,36], regional functional measures open the door to even more intricate tailoring of treatment plans.

Dynamic gadoxetic acid enhanced MRI scans allow for assessments of the extent of hepatocyte function through contrast uptake, rather than assuming uniform function or interrogating the "plumbing" of the liver as in perfusion studies. Quantitative assessment of the extent of function in the regional units (voxels) allows one to compute a sum of the function of units as a measure of global liver function. Various models can be used to get the uptake rate (e.g. DITC, LSITC, DITC with bidirectional exchange or efflux terms). The robustness of these models is of critical importance in ensuring a well-informed treatment plan for every patient, with minimal time

waiting for the patient (whether to receive or finish a procedure). While both intuition and a prior study had indicated that the summed volumetric uptake rate would correlate well with global function, a stronger correlation was found for k_1V_L (measure 1) than for k_1V_{dis} (measure 2), though both had decent (R>0.6) correlations to $\log_{10}$(ICGR15). A similar picture was seen for ALBI, though the distinction between k_1V_L and k_1V_{dis} was less pronounced. It should be noted that while in this chapter both measures (k_1 and $K_1=k_1*v_{dis}$) were computed using LSITC analysis, the same calculations based on the DITC model gave similar results (differing by no more than ±0.05). Both uptake based measures benefited from normalization by patient weight, particularly when compared to ICGR15.

In addition to the measures explicitly compared in this chapter, there are other measures that could be derived from dynamic gadoxetic acid enhanced MRI, such as hepatic extraction fraction (HEF), and liver to spleen ratio (LV/SLV). HEF and LV/SLV have shown promise in prediction of global liver function[23,28,40–42], though these studies have tended to be among patients with healthier ICG scores. A notable feature of uptake rate, unlike measures such as HEF and (LV/SLV), is that it differentiates contrast enhancement due to uptake by parenchyma from contrast in blood plasma and fluid and doesn't assume conformity to flow enhancement in the spleen. While semi-quantitative measures (e.g., LV/SLV) benefit from simplicity, they are machine dependent measures[43]. The primary drawbacks of hepatocyte transport indices like k_1, is the acquisition and analysis complexity[43], which can be mitigated through approaches such as LSITC. A previous study[44] that performed a VOI-based uptake rate calculation using a different method (that was also a simplification of the DITC model designed for improving robustness of the estimation) found a moderate linear correlation ($r = -0.429$, $P = 0.023$)) between ICGR15 and uptake rate in a comparatively healthy patient sample (median ICGR15 was 12% with a range of 0.9 to 52%, compared to a pretreatment median of 29.2% and range of 4% to 82.3% in the present work). Note that the volume based uptake rate calculation has several differences compared to our functional volume calculation. First, equation (2.10) indicates that averaging the concentration time-curve in the liver volume to calculate uptake rate does not produce an averaged k_1 in the volume due to the spatial variation in the distribution volume (v_{dis}). Second, the volume-based k_1 computation does not account for the volume contribution into global

function. Third, the volume-based k_1 does not provide regional liver functional distribution. All these differences affect the global liver function computation using DGAE MRI scans.

Previous work has shown a strong correlation between global function and mean perfusion[13,17]. However, these studies have had relatively small sample population with higher liver function, and lower rates of HCC and cirrhosis as compared to this work. These studies found linear correlations of 0.70 (17 patients), and 0.92 (9 patients) between ICG rates and mean global k_{pv}, but had a lower representation of patients with HCC, and, predictably, lower rates of respective pre-treatment cirrhosis 35%, and 11%, compared to 83% in the present work. A similar reality can be observed in ICGR15 where the two prior studies had respective ICGR15 ranges of 6.72 to 53.18%, and 9.92 to 34.43%, compared to 4.04 to 82.3% in this work. While cirrhosis does result in compromised portal venous perfusion[45], that doesn't necessarily indicate that the level of functional compromise is predictable via the level of compromise in perfusion.

The removal of vascular regions is instructive in this point. The mere presence of perfusion in those regions is not indicative of healthy uptake. Correspondingly, it is entirely possible for losses in functional hepatocytes to be mismatched from the restriction in perfusion as disease progresses (see the second column of figure 3.1). Wang et al[16] used portal venous perfusion to create functional probability maps, with a good correlation to ICG rates. However, their analysis showed increased uncertainty for patients with HCC, which is consistent with the failure in this work to replicate the strong correlations found between ICG retention rates and mean perfusion in healthier, non-cirrhotic patient populations. Taken together with the indications of Wang et al, the results of this work would indicate that for populations with poor liver function and cirrhosis, the mean perfusion is not a reliable indicator of global function, and that the regional perfusion is not a reliable indicator of regional function.

These indications should be understood to have limitations. Firstly, we should note the difficulty in fitting the DITC model. It is possible that reliable functional prediction from portal venous perfusion requires more precise temporal characterization of the portal venous and arterial input functions, or a stronger signal for these input functions than is typically observed in DGAE MRI. It was certainly the case that many exams were not successfully fit via the DITC model, which is unsurprising given its complexity. We should particularly expect difficulty in differentiating k_a

and k_{pv} when C_a and C_{pv} are very similar, with the first few minutes of acquisition containing the salient differences. In cases where perfusion is the desired measure, gadobenic acid would be the preferred contrast, but it is not a hepatobiliary contrast agent, which precludes analysis of uptake rate. Use of gadoxetic acid in this instance allowed for direct comparison in a single scan without additional registration considerations and is certainly justified in any acquisition where regional liver function is the desired measure.

Portal venous perfusion may also have a more complex relationship with function than characterizable through linear a correlation, so that a more complex statistical model such as that performed by Wang et al could derive better predictions of function, but we must recognize that any measure based upon portal venous perfusion is an indirect measure of function, and the results of Wang et al similarly indicate increased uncertainty in cirrhotic patients.

Further work is necessary to determine the degree to which functional avoidance therapy can increase the retained liver function over perfusion based functional avoidance therapy, or mere minimization of mean liver dose. Significant gains would be expected in cases with highly spatially distinct high and low functioning areas outside the tumor region. However, in cases where function is distributed uniformly, or the treatment plan is already determined by organ at risk constraints, a significant reduction in dose to highly functional areas may be infeasible.

3.5 Conclusion

Gadoxetic uptake is promising not only for regional assessment of liver function, but also as a means of estimating global function, and "functional reserve". This is particularly true in cases where liver function is highly compromised and heterogeneous, where the uptake based measures were most reliable.

CHAPTER 4

Impact of Mismatch between Liver Function and Hepatic Perfusion on Functional Avoidance Treatment Planning

The results described in this chapter were previously described by the author in an abstract for the 2018 AAPM annual meeting.[46]

4.1 Introduction

One strategy to preserve overall liver function during radiation therapy is to optimize the treatment plan to avoid highly functional regions and push residual doses to regions with relatively poor function. Although the ultimate clinical benefit of functional avoidance treatment planning requires demonstration through a clinical trial, the actual "functional" map that ought to be used in functional avoidance planning is an important question to address. This chapter evaluates the radiation dose reduction in the liver from functional avoidance treatment planning compared to standard planning, and compares the impact of mismatch between regional hepatic perfusion and hepatocyte uptake of gadoxetic acid in functional avoidance treatment planning.

4.2 Methods

To assess differences between gadoxetic-acid uptake rate and hepatic perfusion on functional avoidance treatment planning, functional avoidance treatment plans were created in 10 patients using the pre-RT scans (3 Female; 6 with HCC and Cirrhosis). Total hepatic perfusion (k_2) was chosen as a more robust measure than k_{pv}, noting that k_{pv} would be expected to be the dominant component of hepatic perfusion (k_{pv} should be 2-3 times k_a)[20]. Ten patients were selected across the range of similarity between k_2 and k_1. Perfusion and uptake rate maps were created as measures of localized liver function, following the procedures described in chapter 2 for the DITC and LSTIC models respectively.

The goal of functional avoidance treatment planning is to spare the best functioning regions in individual patients (a property of relative function). Therefore, the perfusion and uptake maps were thresholded to create function sparing, non-sparing and intermediate regions for each patient. A representative normal functioning region was manually selected in each patient. Voxels with uptake rate (or perfusion) >80%, >36% but <80%, and <36% of the normal value were designated as function (or perfusion) sparing, intermediate, non-sparing regions, respectively. This strategy also mitigates the influence of local uncertainty of perfusion and uptake rate maps on the optimization of treatment planning.
These MRI derived regions were rigidly registered to the treatment planning CT in Eclipse treatment planning system (Varian Medical Systems). Three VMAT plans were created per patient: (i) a liver volume sparing plan minimizing dose to all liver tissue outside the tumor volume, (ii) a function based plan minimizing dose to function sparing regions, and (iii) a perfusion based plan minimizing dose to perfusion sparing regions.

Each plan was designed to deliver 55 Gy, in 5 fractions. Priority one objectives for all plans included SBRT dose limits for OARs and Liver NTCP. The NTCP model used predicts the likelihood of 2-point increase of Child Pugh (CP) Score within 3 months of RT, CP score has been shown to be important in assessing radiation induced toxicity[47]. Plans were compared based on the mean dose to the sparing regions, with particular note given to the reduction in mean dose relative to the liver volume sparing plan.

4.3 Results

The selected 10 patients for function sparing treatment planning had ranked correlation between k_2 and k_1 ranging from 0.54 to -0.58 (Figure 4.2), which represented the cohort of patients well. Examples of quantile uptake rate and perfusion maps for function sparing treatment planning are shown in Figure 4.1. The mean doses in the regions that were considered for function sparing (the regions with k_1 values greater than >36% of the "normal" value in each patient exam) were compared between the liver volume plan, k_1 based sparing plan, and k_2 based sparing plan in each patient, as shown in Figure 4.2. Overall, k_1 and k_2 based sparing plans resulted in the similar mean dose reductions in the function regions (>36% of the individual k_1 "normal" value) as 12.7% (11.8 SD) and 11.5% (13.3 SD), respectively. Compared to the liver volume plan, k_1

based plan yielded a mean dose reduction in the function region greater than 1Gy (or indeed 1.5 Gy) in only 4 patients (#2, #6, #7 and #10 in Figure 4.2). For the most mismatched case, the function based plan provided an additional 10.1% reduction in the mean dose to function sparing regions comparing to using the perfusion based plan. An example case of three plans, liver volume based, function based and perfusion based, are shown Figure 4.3.

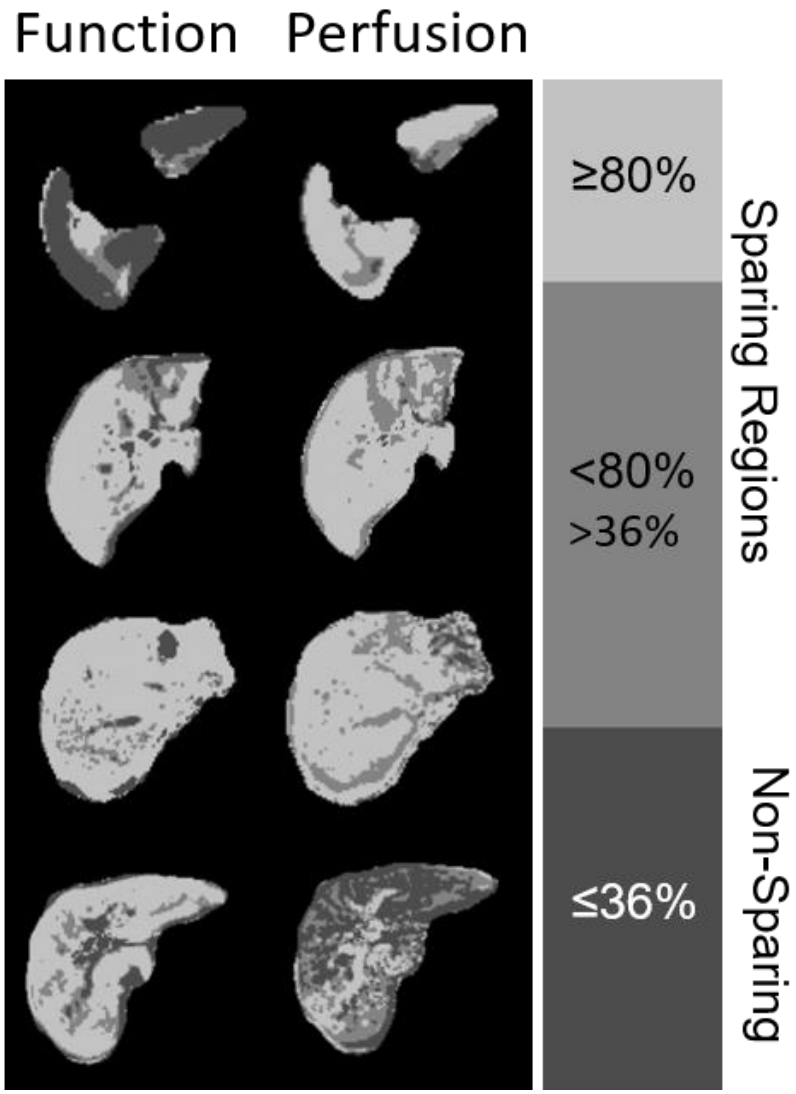

Figure 4.1. *A demonstration of sparing region binning and comparison between the sparing and non-sparing regions as defined by the function and perfusion maps (k_1 and k_2). With k_1 derived from LSITC and k_2 from DITC. Note profound differences in top and bottom images.*

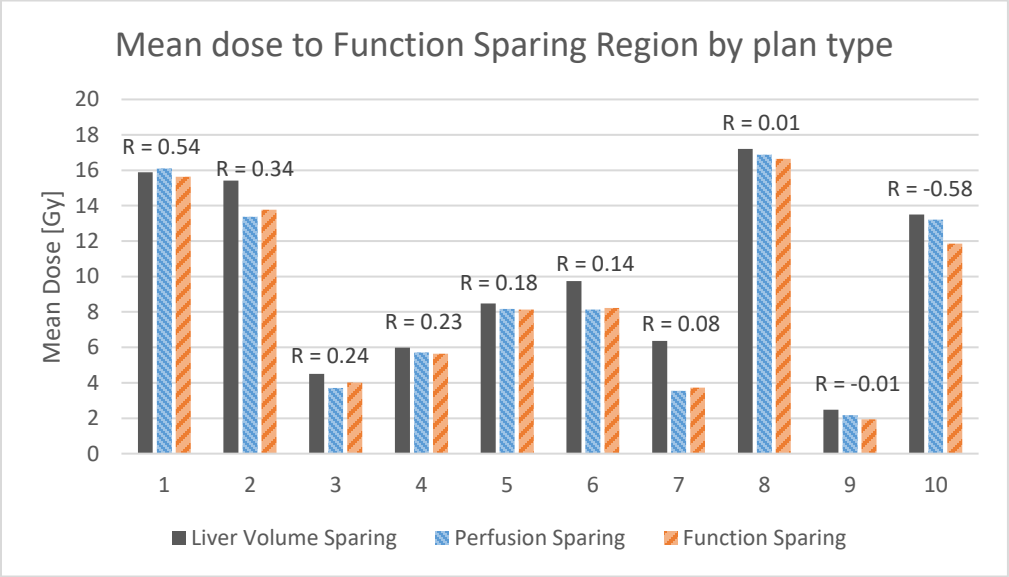

Figure 4.2. *The mean dose to the function sparing region for the liver volume sparing plan, perfusion based plan, and function based plan. Ordered by the descending ranked correlations between maps of k_1 and k_2, with the correlations given above each set of columns.*

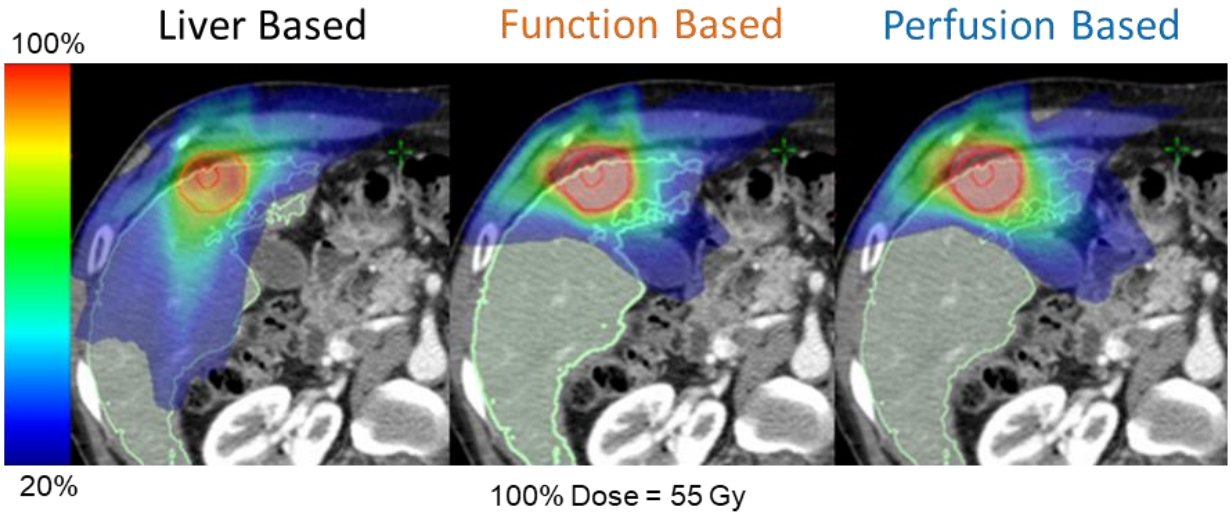

Figure 4.3. *The dose distribution for the liver volume, function, and perfusion sparing plans for an example patient. Note that in this case the function and perfusion based plans are quite similar compared to the liver sparing plan. (Dose distribution not decipherable in greyscale).*

4.4 Discussion

The summary statistics confirm that, as applied, both plans succeed in reducing dose to the specified regions. More interestingly, as applied, overall, either measure will tend to protect function as indicated by the other measure. While that is true overall, in individual cases where the mismatch is most extreme the differences can be quite profound (see figure 4.2).

It is worth considering the reasons for the relevant similarities and differences in sparing regions. Note firstly that there are obvious tissue types that will be evaluated very differently. Vascular regions will have high perfusion, but negligible uptake since they lack hepatocytes (see the bottom comparison of figure 4.1). This principle can be expanded to regions where hepatocytes would normally exist but have been compromised by a pathology. It is entirely possible that low uptake in these regions would correspond with normal perfusion. This is likely seen in the second patient highlighted in figure 4.2, where both gadoxetic uptake rate, and the ICG score showed compromised liver function (ICGR15 = 82.28%), but perfusion did not reveal an abnormality. In cases like this in particular we would expect treatment planning optimization to differ significantly, failing to reap the full benefits of functional avoidance treatment planning if our metric of function is perfusion.

4.5 Conclusion

Functional avoidance treatment planning with perfusion or uptake both result in dose reduction to regions with moderate perfusion or uptake. However, there is potential for large mismatches that limit functional sparing. A clinical trial is required to conclusively determine which method best preserves liver function.

CHAPTER 5

Neural Networks for Robust Quantification of Gadoxetic-Acid Uptake Rate From Suboptimal Acquired MRI

This chapter is substantially equivalent to work the author has already published in *Medical Physics*[48].

5.1 Introduction

As discussed in prior chapters, gadoxetic acid uptake rate can be determined through fitting of the DITC model of gadoxetic acid kinetics, or the DITC derived LSITC model [20,22]. These models are typically applied and validated using HTR DCE scans that collect volumes regularly enough to well characterize the concentration across time in the relevant regions, typically sampling every 5 to 15 seconds. However, the most common clinical gadoxetic acid enhanced MRI exams do not sample this comprehensively. Clinical multiphase scans are obtained for metastases detection and diagnosis. These clinical exams are typically LTR, with as few as 6 volumes irregularly sampling 20 minutes of contrast kinetics. It should also be noted that clinical demands inevitably incentivize shortening exams. If quantification accuracy can be maintained or improved while shortening total acquisition time and eliminating the need for constant acquisition (e.g. LTR style acquisitions), the patient can be given equivalent care with less inconvenience and discomfort, and minimal change to common clinical workflows.

This motivates the development of methods for accurate quantification of regional liver function from short and poorly characterized DCE MRI exams in a robust manner. This chapter develops an artificial neural network (NN) approach to predict k_1 from LTR data. Furthermore, this approach uses data augmentation from a generative adversarial network (GAN) implemented to allow realistic and varied simulation of gadoxetic acid dynamics from the DITC model of gadoxetic acid kinetics in the liver. These approaches are compared to least squares fitting of the LSITC model[22] as applied to both HTR and LTR data. We hypothesize that the new NN

approach allows faster and more convenient MRI acquisition without a sacrifice to the accuracy of functional maps sufficient to compromise treatment guidance.

5.2 Methods

A NN based approach is developed to predict k_1 from LTR data derived from DCE scans. To counter the inherent granularity of the underlying input functions a GAN is used to generate input functions for the augmentation of NN training. The NN based approaches are compared to LSITC analysis for both well characterized HTR data, and the more limited LTR data with varied acquisition duration to assess robustness of the approaches.

5.2.1 Data acquisition

In order to assess error across analysis types and data characteristics, 3D volumetric DCE MRI of the liver were acquired during the intravenous injection of a single standard dose of gadoxetic acid using a Golden-Angle Radial sampling VIBE sequence on a 3T scanner (Skyra, Siemens Healthineer) in a prospective protocol approved by University of Michigan Institutional Review Board. 30 exams were acquired over a set of 22 patients (Age: 50 to 82 years, 6 female) with hepatocellular carcinoma. The 3D free-breathing DCE images of the liver were acquired using a 3D golden-angle radial stack-of-stars VIBE sequence. This sequence over-samples the center of k-space, and allows greater resilience to motion effects than other sequences[37]. The time-series images were co-registered within the liver VOI using an over-determined, rigid-body transformation approach[27]. All acquisitions continued for 16-20 minutes after injection of a single-dose gadoxetic acid contrast and had temporal resolutions of at least 5 samples per minute.

The acquired HTR data was subsampled to produce corresponding LTR data (Figure 5.1). This was done by interpolating (1) a pre-contrast volume, (2) 3 volumes spaced 25 seconds apart designed to capture the arterial and portal venous phases, and (3) two volumes at the end and midpoint of the acquisition (roughly 20 and 10 min, respectively). C_a, C_{pv}, and C_t were obtained as described in a Chapter 2[22].

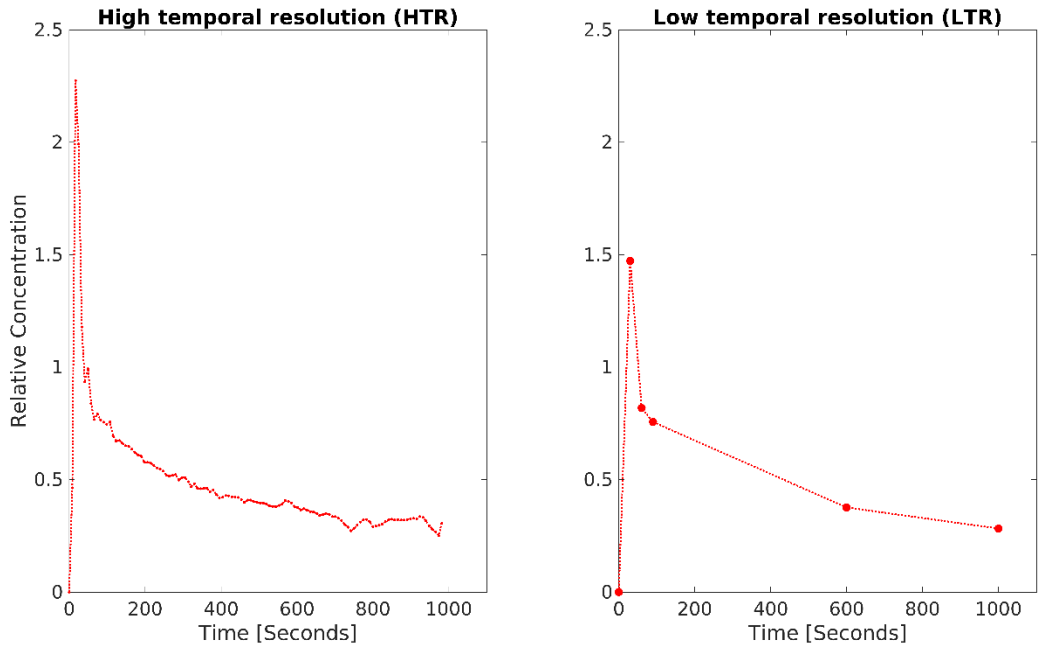

Figure 5.1. *Illustration of characteristics of densely sampled high temporal resolution (HTR – left) and more sparsely sampled low temporal resolution (LTR - right) datasets. HTR data is regularly sampled at 5-10 s intervals for the duration of 16-20 min. LTR data involves the acquisition of three post contrast samples uniformly spaced at intervals of 15 to 35 seconds, followed by two points, one at roughly 10 min and another at roughly 20 min post contrast. LTR data is the clinical norm.*

5.2.2 Least squares fitting of LSITC model

LSITC analysis involved linear regression for the best fit to equation (2.10). For HTR data t_0 was selected to maximize the linearity of the time range being fit, as described in chapter 2. In the analysis of the synthetic LTR data, t_0 was chosen 75 seconds after the initial upswing of the arterial peak. In both cases the resulting estimate of k_1 was the intercept normalized slope of the least squares linear fit from t_0 to the final point. This allowed the linear fit to incorporate 3 points for the LTR data.

5.2.3 Neural network – rationale and implementation

Given a reasonable set of patients with k_1 estimated from HTR data, a machine learning approach is a natural means for creating a prediction from a subset of that data, e.g., multiphase LTR data. To this end, a simple fully connected NN with 4 hidden layers (10,10, 5 and 5 neurons) was trained on voxel-wise LTR data to predict k_1 (Figure 5.2). Both the total acquisition length and the placement of points in the training data were varied considerably to encourage robustness to variation. This was performed by having the arterial and portal venous phase points sampled uniformly 15 to 50 seconds apart, with uniformly distributed perturbation

up to 10% of the sampling period. The endpoint t_{end} was randomly selected from a uniform distribution from 5 minutes after the arterial upswing until the end of the acquisition. The midpoint sample was selected from a uniform distribution from $0.25t_{end}$ to $0.75t_{end}$. Each voxel then consisted of 5 pairs of values representing the x and y vectors calculated from equation (2.10) based on 5 post-contrast time points (as in the right panel of Figure 5.1).

Training was performed by randomly selecting 3 million voxels in the livers from 30 exams, holding $3/5^{ths}$ for training, $1/5^{th}$ for validation, and $1/5^{th}$ for testing. Training and validation data did not have patients that overlapped with the patients in the data held for testing.

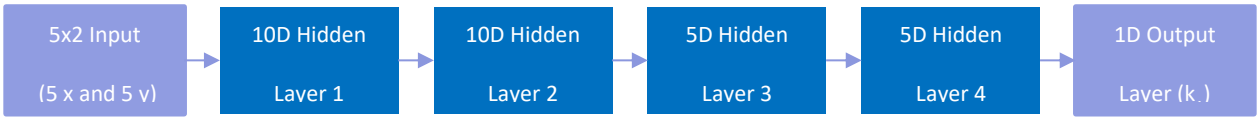

Figure 5.2. *Architecture of the simple fully connected neural network used to predict voxel-wise* k_1.

5.2.4 GAN

5.2.4.1 GAN – rationale. No matter how many voxels are used for training, if we have only a pool of 30 exams, and 22 patients, each voxel will come from one of 30 categories defined by the precise input functions that corresponded to that exam. This inspires data augmentation for the set of input functions to ensure the training data is better spread across the reasonable space of input functions. A GAN is a reasonable choice for this generative task. This approach trains both a generator and a discriminator, who act as adversaries to one another. The generator seeks to generate artificial input functions that are in the space of real input functions. The discriminator attempts to discriminate between the real examples and those generated artificially. Eventually, the generated examples should be essentially indistinguishable from examples drawn from the true dataset. GANs have been applied in a number of circumstances, involving both temporal biological signal[49] and medical image [50,51] generation, including generation for data augmentation [52]. Here we use a generative adversarial neural network to generate arterial and portal venous input functions for gadoxetic acid dynamics in the liver.

5.2.4.2 GAN design and implementation. The GAN consisted of a simple network for conversion of a random vector (length 20) into outputs corresponding to arterial (C_a) and portal

venous (C_{pv}) input functions (two vectors of length 100) along with an indicator of the sampling period T. The network architecture can be seen in Figure 5.3.

The generated input functions are then used as to create tissue concentration curves (C_t) using the DITC model.

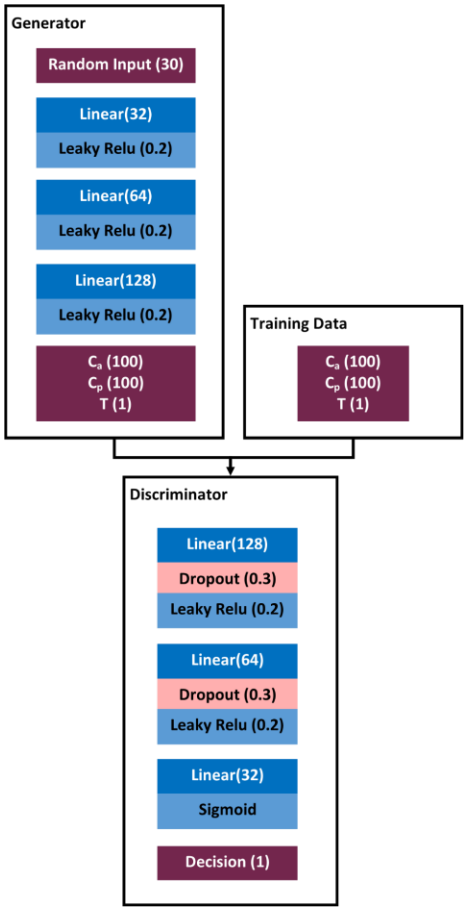

Figure 5.3. *The design of the GAN used for generation of C_a and C_{pv} curves. Parenthetical values represent the dropout rate for dropout layers, the gradient of the leaky Relu, and the number of size for all other layers.*

5.2.4.3 NN augmentation from GAN data. Training using the GAN generated data serves a dual purpose – firstly it acts as a confirmation that the GAN generated data is actually representative of the real C_a and C_{pv} curves, secondly, it could improve prediction accuracy with comparatively minimal chance of overfitting, based on the increased variability in C_a and C_{pv} for the training data. This dataset then has ground truth DITC defined uptake rates with input functions

replicating the variation observed empirically. This data can be used to augment the real data in training neural models to determine uptake from restricted datasets.

In order to train a network to generated C_a and C_{pv} curves from a random vector, training data was created by first generating 1 million random C_a and C_{pv} pairs with corresponding T. This was performed for 5 holdout groups of patients corresponding to the training holdout groups described in 5.2.3 to ensure the learned sets were not influenced by testing patients' own data. For each of these sets of C_a and C_{pv} curves, k_1 and v_{dis} values were randomly selected from the relevant patient set (excluding holdout patients), while k_a, k_{pv}, T_a and T_{pv} were randomly selected from roughly physiologically reasonable ranges (see Table 5.1). C_t curves were then generated from the DITC model using the GAN generated C_a and C_{pv} functions along with the random parameters described in table 5.1 as inputs to the model. Finally, gaussian distributed noise was added such that the measured SNR would be 40 dB.

Table 5.1. *The values used for the generation of training data using the dual-input, two-compartment model. Note that U(a,b) is the uniform distribution from* a *to* b*, and* $N(\mu,\sigma^2)$ *is the normal distribution about* μ *with standard deviation* σ*. In this case the normal distribution was truncated to remove results outside the range [0,1].*

Parameter	Distribution	
k_1, v_{dis}	Randomly drawn from patient set	mL/100mL/min, mL/mL
$k_{pvp} + k_{ap}$	$U(50,300)$	mL/100mL/min
k_{pvp}	$N(0.75, (1/16)^2)(k_{pvp} + k_{ap})$,	mL/100mL/min

5.2.4.4 LSITC optimization from GAN data. Finally, consideration was given to minimize the error in LSITC analysis. The two obvious "tunable" parameters are t_0 and sampling time. The parameter t_0 refers to the first time point considered to satisfy the conditions of the LSITC model and thus used as the first point in the linear fit of the model. This is currently selected through a maximization of linearity as calculated by the ratio of singular values [22]. Determination of the sampling times is more complex, particularly if we implement irregular sampling as in LTR collection. This chapter uses the GAN simulated data to optimize t_0 and the sampling times, discretized in 30 second increments, for the LSITC analysis. Optimization is performed using a genetic algorithm to search for t_0 and sampling times. Breaking the signal into 30 seconds intervals increased the tractability of the problem for this discrete genetic algorithm. This resulted in each of the sampling points being chosen from 32 intervals of 30 seconds in the 16 minute datasets, where the first and last points are required. This was performed for 1 to 10

additional points, where the choice of points was optimized to minimize MSE error in a set of GAN based DITC generated synthetic voxels.

5.2.5 1D CNN

In addition to the application of a simple linear network to LTR data as discussed in 5.2.3, a reliable estimate of uptake rate from a 10 minute scan would be a meaningful improvement for patients and clinicians. With the expansion of the dataset complexity from LTR to HTR data it is likely that additional network capacity will be rewarded with superior results. Since the elements of the temporal sequence are highly correlated, a 1D convolutional neural network (CNN) approach is a logical extension that allows for expanded capacity without a needless ballooning of parameters.

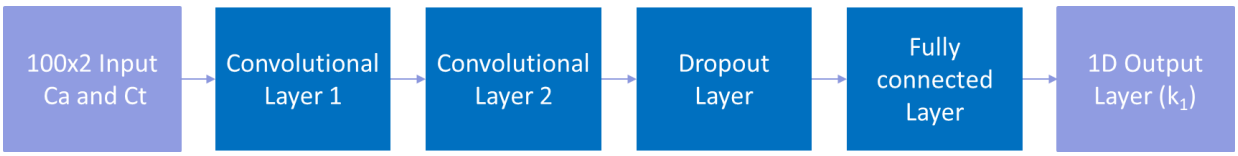

Figure 5.4. *The design of the 1D CNN applied to HTR data. Convolutional layer 1 contains 8 filters with width 9, convolutional layer 2 contains 16 filters with width 5, and the fully connected contains 10 neurons).*

To this end, a 1D CNN with 4 hidden layers (8 filters with width 9, 16 filters with width 5, A dropout layer, and a fully connected layer with 10 neurons) was trained on voxel-wise HTR data to predict k_1 (Figure 5.4). Training data was drawn from a similar pool as for LTR data in 5.2.3, but all training data was concatenated to 10 minutes. Final testing was likewise only performed for data concatenated to 10 minutes after the beginning of the arterial peak.

5.2.6 Error metric for evaluation of analysis methods and acquisition paradigms

For each method and dataset used to estimate k_1, the error was measured as NRMSE with the results of least squares fitting of the LSITC model for the full length (16-21 min) HTR dataset as the reference. NRMSE is defined here as RMSE normalized on an exam by exam basis by the interquartile range of the reference values as:

$$NRMSE = \frac{RMSE}{interquartile\ range} \tag{5.1}$$

Mean NRMSE is merely the mean across all exams analyzed.

The reference values were restricted to the values with a relative uncertainty below the 75th percentile. This minimizes the likelihood of performing the comparison with outliers and artifacts, such as those seen on some edges, but will also tend to decrease the denominator in the NRMSE calculation.

Relative uncertainty was measured as the expected standard deviation in k_1 estimation for the fit in a given voxel divided by the predicted k_1 for that voxel. Here the variance in k_1 is estimated by the Taylor expansion of the variation of K_1/v_{dis} (where K_1 is the slope in equation 2.10) as:

$$var(k_1) = var\left(\frac{K_1}{v_{dis}}\right) \approx \mu_{K_1}{}^2/\mu_{v_{dis}}{}^2\left(\frac{\sigma_{K1}{}^2}{\mu_{K1}{}^2} - \frac{2Cov(K_1,v_{dis})}{\mu_{K1}\mu_{v_{dis}}} + \frac{\sigma_{v_{dis}}{}^2}{\mu_{v_{dis}}{}^2}\right) \quad (5.2)$$

where

$$\sigma_{K_1} = \sqrt{\frac{\sum_{i:\, xi=x0}(\hat{y}_i - y_i)^2}{(n-2)\sum_{i:\, xi=x0}(x_i - \bar{x}_i)^2}} \quad (5.3)$$

$$\sigma_{v_{dis}} = \sqrt{\left(\frac{1}{n}\right)\sum_{i:\, xi=x0}(x_i)^2} \quad (5.4)$$

Where σ_a and μ_a are the respective standard deviations and means of any given measure a. x and y are defined in equation (2.10).

All results from five methods and datasets were compared to the k_1 estimated by fitting the LSITC model for HTR data at maximum length (at least 16 minutes and no more than 21 minutes), which are summarized in Table 5.2.

***Table 5.2**. The abbreviations used for each method and data pairing evaluated along with a description of the relevant method and data.*

Method/Data Abbreviation	Method Description	Input Data Description
LSITC-HTR	Fitting of LSITC model with t_0 chosen to maximize linearity	HTR data, with the data length truncated to a maximum length of 4 to 16 minutes
LSITC-LTR	Fitting of LSITC model with t_0=75 seconds	LTR data, with the data length truncated to a maximum length of 4 to 16 minutes. The initial points spaced at 25 second intervals.
NN-LTR	Application of the NN model trained by k_1 resulting from LSITC-HTR for full HTR datasets	LTR data, with the data length truncated to a maximum length of 4 to 16 minutes. The initial points spaced at 25 second intervals.

Augmented NN-LTR	Application of the NN model trained by DITC based data using input functions generated by GAN.	LTR data, with the data length truncated to a maximum length of 4 to 16 minutes. The initial points spaced at 25 second intervals.
OPT LSITC-LTR	Fitting of LSITC model with algorithmically chosen sampling times and t_0	8 points selected algorithmically to minimize error in augmented dataset. Truncated to a maximum length of 8 to 16 minutes.
LSITC HTR t0 = OPT	Fitting of LSITC model with HTR data but t_0 set to the optimum found in OPT LSITC-LTR	HTR data, with the data length truncated to a maximum length of 4 to 16 minutes

5.3 Results

5.3.1 Fitting of LSITC model

As expected, directly fitting the LSITC model to HTR data yielded more accurate k_1 values than fitting to LTR data. For both datasets the errors grew rapidly with a decrease in the acquisition length of the data (see Figure 5.5). At full acquisition length (16 minutes), LSTIC-HTR and LSITC-LTR resulted in an average NRMSE across exams of 0.60 (SD 0.38) and 1.77 (0.99), respectively. At an acquisition length of 10 min the average NRMSE increased to 2.59 (1.34) and 3.09 (1.54) for HTR and LTR datasets, respectively, as seen in Table 5.3. A visual comparison at 10 minutes can be seen in Figure 5.6.

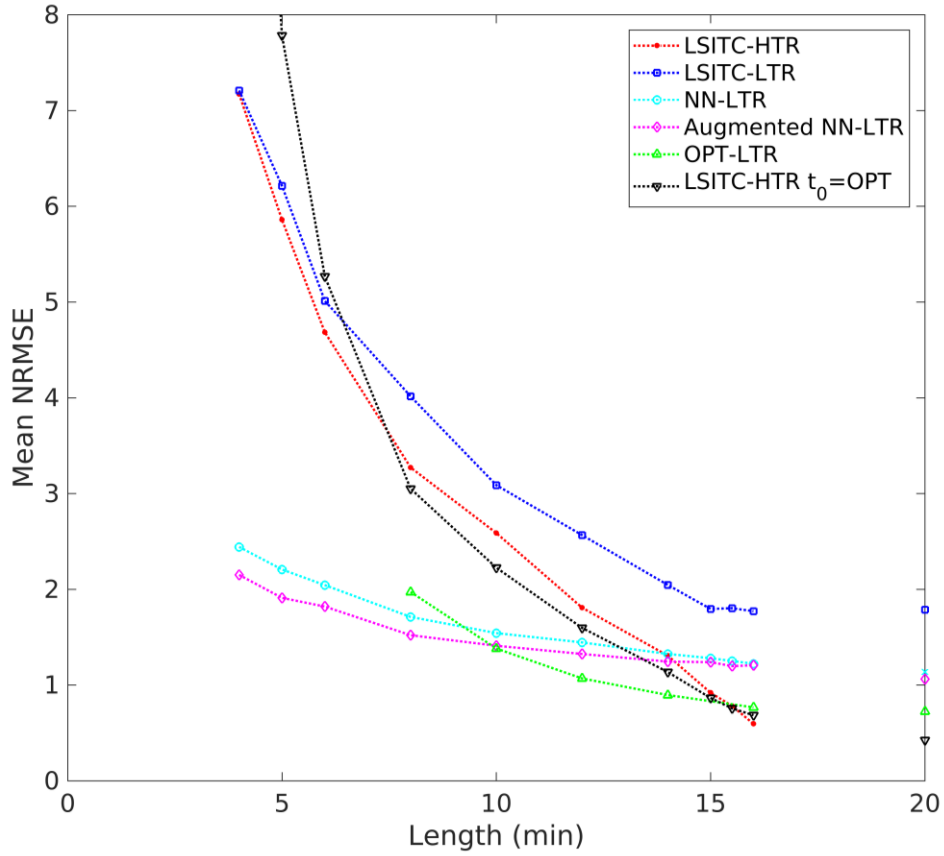

Figure 5.5. *Errors of estimated k_1 values with varied acquisition lengths for the tested methods.*

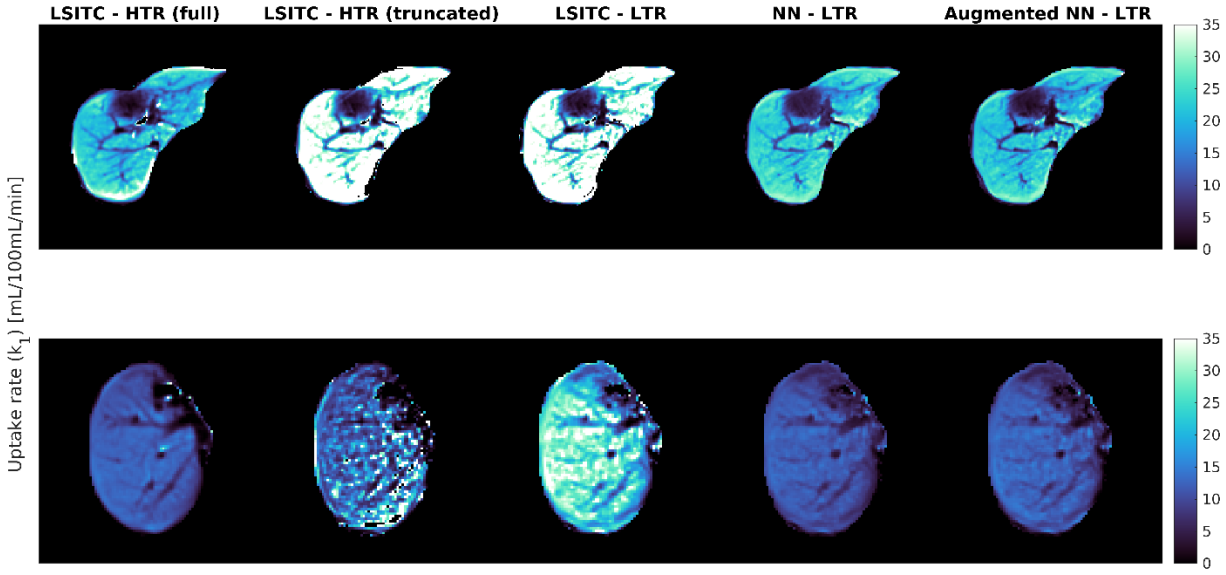

Figure 5.6*. The k_1 maps created using the HTR and LTR data truncated at 10 min both from directly fitting the LSITC model (second and third columns) and from the NN and GAN augmented NN models (fourth and fifth columns respectfully). The first column displays the reference k_1 images by fitting the LSITC model to full length HTR data acquired over approximately 20 min.*

5.3.2 NN model

The NN model yielded significantly reduced error rates in k_1 estimation over direct fitting of the LSITC model to the LTR data at all tested acquisition lengths (4-20 min). When the acquisition length was less than 14 min, the NN model applied to the LTR data resulted in the errors less than directly fitting of the LSITC model to the HTR data. This difference became significant for acquisitions of 10 minutes or less. The errors yielded by the NN model increased slowly with the acquisition length reduction, suggesting the NN model was resilient to data length. In contrast, direct fitting of the LSITC model yielded quickly increased errors with the data length reduction, regardless of the temporal resolution of the data (Figure 5.5).

5.3.3 GAN augmented NN model

On visual inspection randomly selected curves generated by the trained GAN seemed to replicate the basic features of the measured curves without being direct copies of individual examples. For randomly selected GAN generated C_a curves, the nearest normalized neighbor was found from the measured set of input curves. Three examples are shown in Figure 5.7. In each column the top plot is a randomly selected generated C_a and C_{pv} pair, and the bottom plot is the real C_a and C_{pv} pair whose normalized C_a curve is the nearest neighbor to the generated C_a curve based on a sum of squares difference. The comparisons did not show evidence of direct replication of the specifics of particular measured curves.

In addition to visual inspection, the distribution created by the GAN was assessed by producing histograms approximating the probability distribution of the pairwise Euclidean differences between examples within the measured data, as well as the pairwise differences in data generated by each GAN. Figure 5.8 displays these distributions of pairwise differences for each GAN, superimposed over the distribution of pairwise differences for the measured data. The difference between the mean distance for each GAN and the measured data is shown, along with the earth-movers-distance (EMD) to better represent the differences between distributions. In all cases the distribution of differences in GAN data visually mirrors that of the full dataset, with the smoothing we would expect from a larger number of samples from a similar dataset.

Augmentation with GAN generated data gave mixed results. Training on only synthetic data resulted in improvement over prediction error from training only on real data (Figure 5.5). With a statistically significant improvement in error over LSITC HTR for all datasets of length 12

minutes or less, and no significant drop in error up to 15 minutes. However, combining the real data with additional synthetic data did not meaningfully improve the prediction error. The results of augmented NN model trained by synthetic data only are shown in Figure 5.5 and Table 5.3 (Augmented NN-LTR).

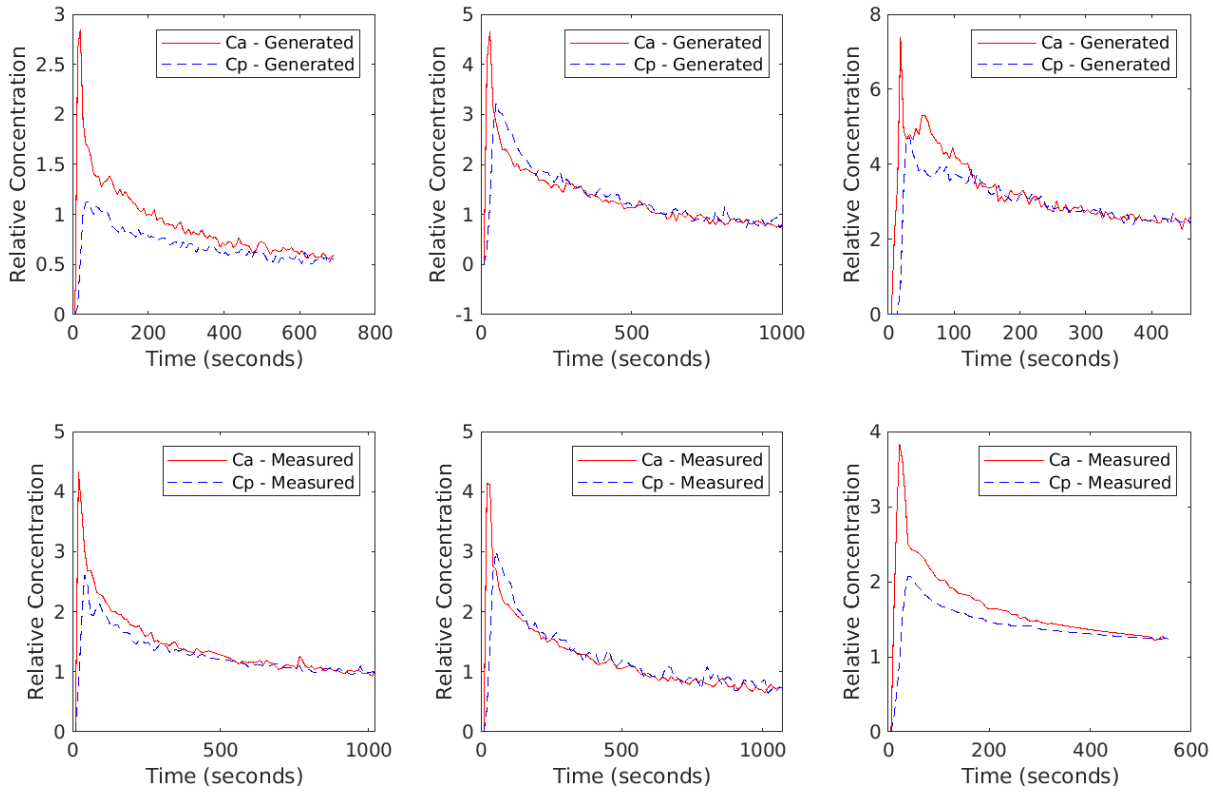

Figure 5.7. *Examples of generated (top row) and nearest neighbors from the measured (bottom row) C_a and C_{pv} curve pairs. Nearest neighbors were calculated based on the sum of squared differences in C_a alone.*

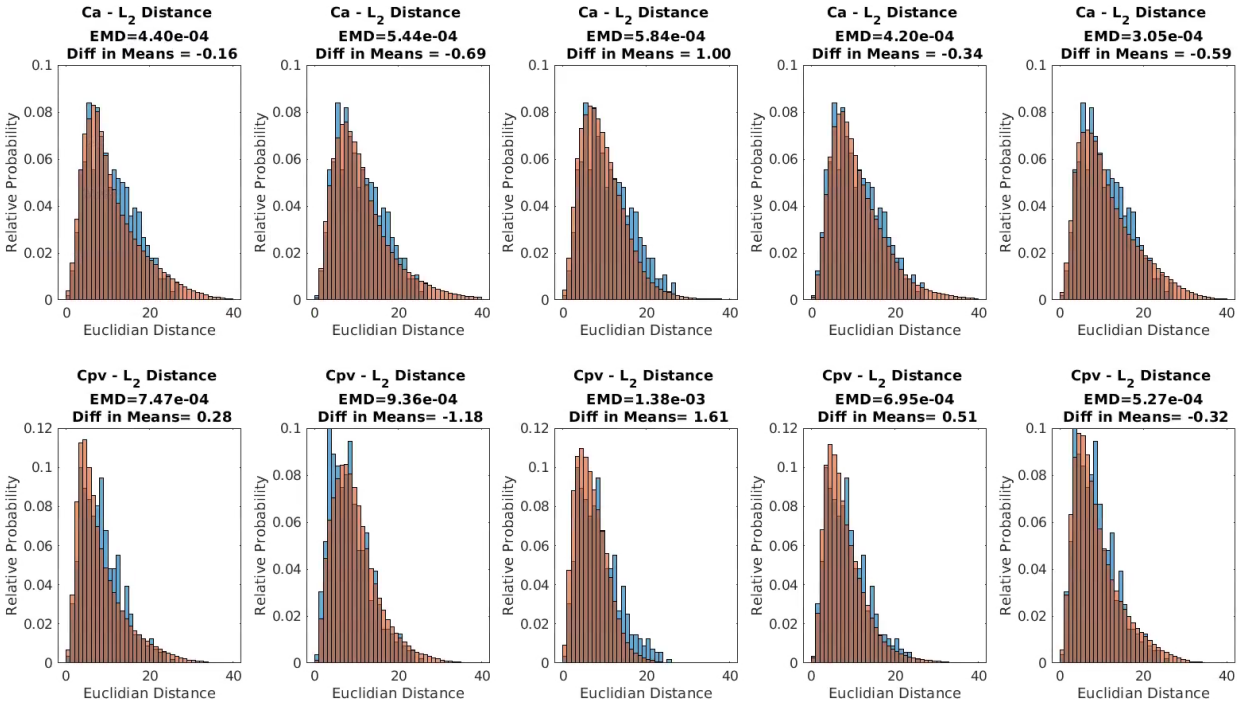

Figure 5.8. *For each of the 5 GANs used, the probability distributions for L_2 norm of the distance between randomly selected C_a and C_{pv} curves for GAN generated data is shown in red. The probabilities for the measured data are shown in blue as reference.*

5.3.4 Optimization of time points for the LSITC model fitting

When selecting the optimum sampling points for the LSITC model fitting, as additional points were selectively added to the set, optimization yielded a t_0 of 3 minutes in every case, without any sampling point prior to t_0. The sampling times chosen tended to group just after t_0, and near the end of the dataset. The error leveled off near 8 points in the simulated data, as seen in Figure 5.9. As a result, 8 points were used when testing this approach, apart from the pre-contrast and final points.

Implementation of the GAN data for LSITC optimization (OPT-LSITC LTR) yielded errors significantly lower than direct fitting of the LSITC model to HTR data with acquisition lengths of 12 min or less, and lower than NN models for data lengths greater than 10 minutes (Figure 5.5). This suggests that optimization of the time of data point acquisition could improve the performance of the LSITC model, but the NN model with non-optimized data still could perform better at a short acquisition length.

A further test of the optimal t_0 (3 min) was performed with full HTR data. As seen in Figure 5.5, the LSITC model fitting to HTR with a dynamic t_0 (LSITC-HTR) and an optimal t_0 (LSITC-

HTR t_0=OPT) yielded similar results, but worse results than the LSITC model fitting to the optimal 8-point LTR data (OPT LSITC LTR), indicating that the robustness of performance of the optimized LSITC is not merely due to the choice of t_0 but due to the particular set of points selected.

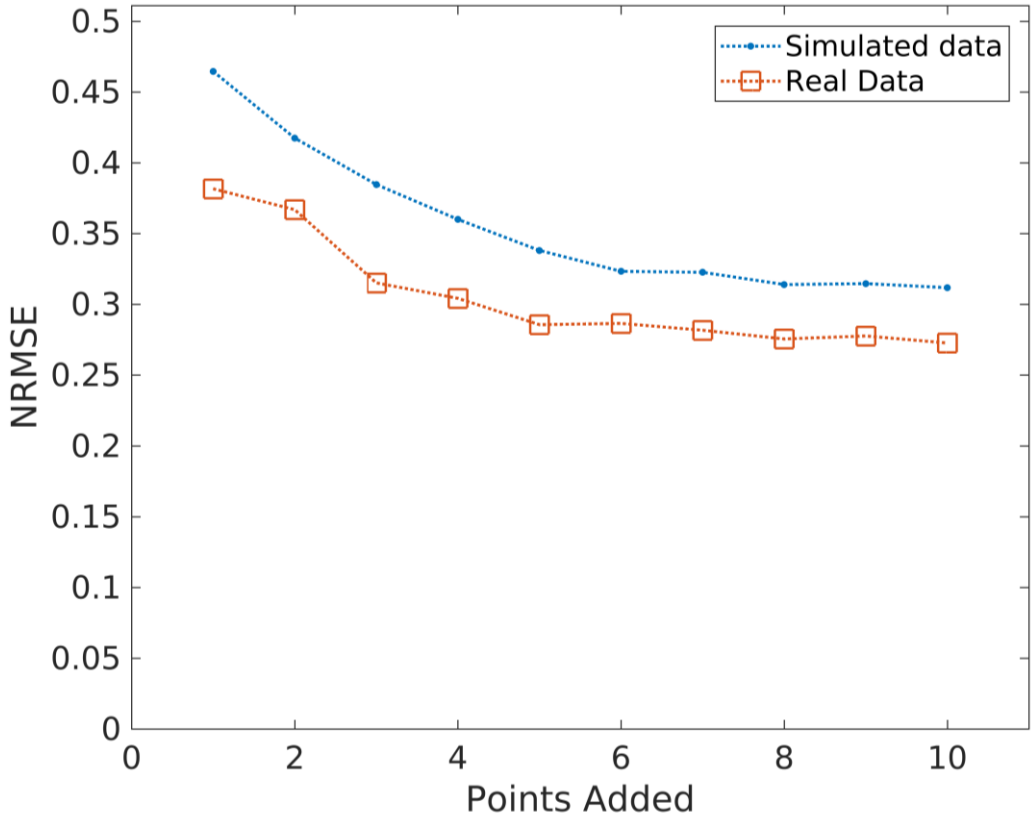

Figure 5.9. *The errors in simulated and real data as a function of the number of optimum sampling points using a procedure derived from the genetic algorithm. Note that error in the data leveled off after 8 points.*

Table 5.3. *Error rates (NRMSE) for each method as function of data length. Statistically significant improvements in NRMSE over LSITC HTR are indicated by an asterisk (*). Statistically significant increases in error are indicated by a negated asterisk. Significance was estimated based on a two sample t-test with a significance level of 0.05, except for the Max row, where a single sample t-test was used.*

Series Duration (min)	**NRMSE - mean (standard deviation)**					
	LSITC HTR	**LSITC LTR**	**NN LTR**	**Augmented NN LTR**	**OPT LSITC LTR**	**LSITC HTR t_0 = OPT**
4	7.17 (4.39)	7.21 (3.93)	2.44 (2.06)*	2.15 (1.78)*		14.64 (9.44) -*
5	5.86 (3.47)	6.21 (3.24)	2.21 (1.79)*	1.91 (1.43)*		7.78 (4.58)
6	4.68 (2.72)	5.01 (2.95)	2.04 (1.52)*	1.82 (1.16)*		5.27 (3.18)
8	3.27 (1.79)	4.02 (2.35)	1.71 (1.15)*	1.52 (0.85)*	1.97 (1.39)*	3.05 (1.78)
10	2.59 (1.34)	3.09 (1.54)	1.54 (0.93)*	1.41 (0.75)*	1.38 (0.72)*	2.23 (1.17)
12	1.81 (1.08)	2.57 (1.39) -*	1.44 (0.79)	1.32 (0.67)*	1.07 (0.57)*	1.60 (0.99)
14	1.31 (1.01)	2.05 (1.27) -*	1.32 (0.71)	1.24 (0.62)	0.90 (0.53)	1.14 (0.90)
15	0.92 (0.61)	1.79 (1.07) -*	1.28 (0.68) -*	1.24 (0.63)		0.86 (0.62)
15.5	0.78 (0.52)	1.80 (1.02) -*	1.25 (0.64) -*	1.20 (0.58) -*		0.76 (0.54)
16	0.60 (0.38)	1.77 (0.99) -*	1.22 (0.69) -*	1.21 (0.66) -*	0.77 (0.42)	0.68 (0.50)
Max	0.00 (0.00)	1.39 (0.80) -*	1.14 (0.58) -*	1.06 (0.56) -*	0.72 (0.33) -*	0.42 (0.26) -*

5.3.5 1D CNN

The error rate in terms of NRMSE relative to the full length LSITC-HTR was comparable to the error rate from merely altering t_0 for full length data as in 5.3.4, giving an NRMSE of 0.53 with a standard deviation of 0.26. This also compares favorably with the error rate of LSTIC-HTR after truncation to 16 minutes. Visually the maps seemed to replicate physiological patterns in the reference images with a reduction of outliers and edge effects (See Figure 5.10). The correlation with the reference was R = 0.93.

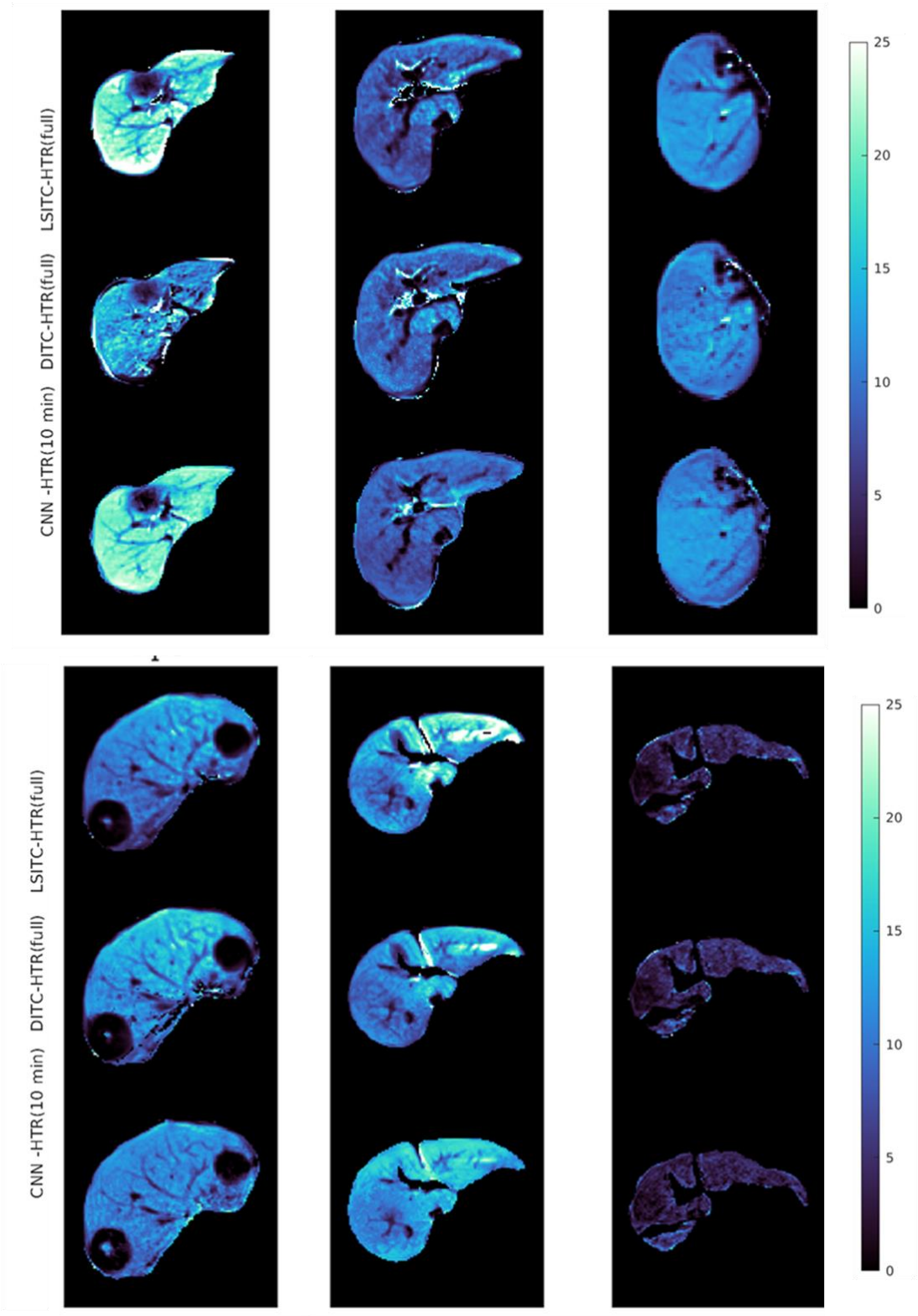

Figure 5.10. *Example maps of uptake rate based on (top to bottom in each block) LSITC and DSITC models for full length HTR data, and the CNN for only 10 minutes of data. The colormap depicts the value from 0 to 25 mL/100mL/min.*

5.4 Discussion

In this chapter, we developed NN models for estimation of k_1 and compared the results of the NN models to those from direct fitting of the LSITC model for various acquisition lengths and temporal resolutions of Gadoxetic acid enhanced dynamic MRI of the liver. Overall, the NN models are more resilient to the acquisition length reduction. The augmented input functions using GAN can further improve the performance of the NN models. For direct fitting of the LSITC model, ten optimized time points in the Gadoxetic acid enhanced dynamic data can significantly out-perform the HTR data (5-10 sec per volume) for acquisition lengths of 12 minutes or less, and the NN method for acquisition length not shorter than 8 minutes. This work suggests that the NN approach can be used to enhance the performance of k_1 estimation and optimize the data acquisition.

A key element of modeling liver pharmacokinetics is obtaining arterial and portal venous input functions. These input functions have been estimated using combinations of exponentials and other simplifications, but this involves either profound simplification or the development of models of increased complexity without a guarantee of successfully capturing the relevant features of the input functions. Use of measured input functions has notable advantages in capturing the true empirical characteristics of these input functions. However, when employing data driven methods this will practically limit the researcher to a relatively small number of example cases. When machine learning methods are applied to millions of voxels but the guiding input functions consist of a few dozen examples, we may fear overlearning these limited underlying examples, rather than a more useful learning of the underlying relationships between our relevant parameters and input functions in general. Addition of noise or variation in sampling time may make this underlying granularity less starkly memorable. However, a more ideal solution would be the construction of arbitrary or random input functions from the feature space the input functions inhabit. A promising means for this generative task is a generative adversarial neural network.

One difficulty in generative networks, where the network is not cyclic (generating corresponding examples in another space rather than arbitrary or random examples in the desired space) is assessment of the quality of the generative model. One approach is the usage of these examples as augmentation data for a relevant learning task. If the augmentation helps, it is more reasonable

that the generative model is representing the variation in the underlying set appropriately, or at least in a way that helps the trained network to better understand the relevant relationships. Here we used a generative adversarial neural network to generate arterial and portal venous input functions for gadoxetic acid kinetics in the liver.

The augmented NN that was trained only on GAN generated data resulted in superior results as compared to the NN trained using any fraction of the measured data with HTR-LSITC as the reference. There are various possible causes of the decrease in performance with the addition of real data. It is likely that the very few input functions were not useful in further generalizing the solution over the training from the GAN and DITC generated data. It may also have skewed the solution towards those measured input functions. It should be noted that since the GAN itself is trained from measured data, the generated examples will include characteristics caused by sampling noise, movement and other variations in the data. Because of this the input to the DITC model generated from this GAN has variation that would not be expected in the underlying input functions in reality.

In addition to the already mentioned benefits, the GAN derived data and DITC model defined reference values allowed the simulated dataset to be used to evaluate independent models relative to the DITC model. This allowed us to use references of not only our best estimate (whether DITC or LSITC) to complete (16+ min) real datasets, but also to the ground truth inputs to the DITC model without fitting error in the reference k_1 values. This helps quantify possible error in these estimates and gives a parallel reference measure for restricted methods. This is of particular interest when attempting to assess optimum, or at least improved, acquisition times for the image volumes used to estimate k_1. Use of these model defined input parameters made this optimization less susceptible to a mere reproduction of the linear fit of the LSITC model (along with any limitations or errors in this method), and helped to assess the best timing (giving the variability observed in the input functions) to acquire points for LSITC without bias to the timing used in the measured reference set.

The optimal sampling points for OPT-LSITC LTR essentially followed the expected weights for a linear regression, in that points near the end were preferred, with successive trials adding points closer to the center as those at the ends were already included. The selection of t_0 is perhaps more salient, indicating that the addition of a point near the 3 min mark would aid LSITC

accuracy when applied to LTR data. This time roughly corresponds to the equilibrium phase [53], which would logically initiate the portion of the data where the assumptions of the LSITC model hold true. This approach resulted in lower error than even LSITC applied to HTR data from 15.5 to 8 min, for the real dataset, even though the reference was used HTR data with a variable t_0. This also casts doubt on the use of 75 seconds as t_0 in LTR data. If 3 minutes is the location of the equilibrium phase, then voxel-wise LSITC analysis of most LTR data has only 2 data points to work with, since none of the arterial or portal venous phase points will fall after that point. Without an overdetermined fit the error rates will likely be large, and concurrent error quantification will rely on assumptions regarding the similarity of nearby points. However, the selection of t_0 was not the primary factor in the improvement over other LSITC methods. This is apparent from the small difference between LSITC-HTR and LSITC-HTR where t_0 = OPT. This indicates that the specific selection of points was helpful in improving the fit. It is possible that some of the improvement came from selecting no points prior to t_0. This does not change which points are fit, but does change the x and y vectors since the integral of C_a will differ in equation 2.10. It may be that the discrepancy of C_a from C_{pv} increases the error in datasets that include pre-t_0 sampling points.

Regardless of the method used the error was greater for shorter datasets. Data length was especially significant for LSITC analysis, for both LTR and HTR data. With a fixed best t_0 and careful choice of sampling points this was reduced somewhat, perhaps making acquisitions as short as 12 minutes practical. Below this level the NN methods worked best, showing relatively little change in error with data length in time. This indicates that the underlying information is sufficient for a comparatively accurate prediction even with relatively short collection time used by the NN. However, the results did not outperform LSITC-HTR for long datasets. In each of these cases it is important not to interpret the error in absolute terms, particularly near the maximum length. Remember that the error measures will be impacted by error in the results of LSITC applied to HTR.

Use of the LSITC model as the reference allowed rapid analysis and comparison with regard to k_1, even for LTR data. In a previous study, k_1 values estimated from the LSITC and DITC have been compared and the results are very similar[22]. However, this model does omit parameters present in the DITC model, notably k_a and k_{pv}. Previous studies have correlated portal venous

perfusion to liver function[54] and arterial perfusion to tumor presence[20]. Theoretically, simultaneous quantification of k_1, k_a and k_{pv} from a single dynamic MRI acquisition using the DITC is advantageous. Practically, there are some limitations. The FDA approved standard dose of Gadoxetic acid only contains a quarter of the Gadolinium in a standard dose of Gd-DTPA or Multihance. This results in a weak contrast enhancement and a low signal-to-noise ratio in the arterial phase signals, thereby challenging reliable quantification of arterial perfusion. Therefore, in practice, if tumor diagnosis and assessment are the primary interest, Gd-DTPA or Multihance is used. If liver function measurement is the primary interest, Gadoxetic acid is used. If both tumor assessment and liver function are of interest, a trade-off has to be made. Compared to the DITC and LSITC models, the Tofts model only considers the contrast transport between the intra-vascular and the extra-cellular space, which can only be applied for an extra-cellular contrast agent, but not an intra-cellular agent, like Gadoxetic acid.

5.5 Conclusions

Data length is significant for LSITC analysis as applied to DCE data for standard temporal sampling. With a fixed best t_0 and careful choice of sampling points this can be reduced somewhat, particularly for acquisitions at least 12 minutes in length. Below this level the NN worked best, indicating that NN methods may be helpful in improving the robustness of uptake analysis in temporally short datasets. Combination of a GAN with DITC model created data contributed to the training of the NN, indicating the variation in input functions was being appropriately represented. Further work should assess the impact on functional avoidance therapy dependent on the means used to create functional maps.

CHAPTER 6
Conclusions, Summary, and Extensions

This work has introduced a method, LSITC, for the robust quantification of voxel-wise liver function using DGAE MRI. This method was validated through various means in chapters 2 and 3, with voxel-wise validation most pronounced in chapter 2 and global validation most pronounced in chapter 3. LSITC can be used in the place of DITC to give more robust and rapid estimates of gadoxetic acid uptake rate. Furthermore, chapter 3 establishes this measure can be used to predict global function, based on comparisons with both ICGR and ALBI. The validation of uptake rate based measures simultaneously casts doubt on portal venous perfusion as an accurate and robust measure of liver function in some patient populations. In this instance the patient population had a very high prevalence of HCC and Cirrhosis, with ICGR and ALBI scores depicted many cases with profoundly compromised liver function. The inclusion of best fits between ALBI, ICGR and uptake derived measures also creates an initial conversion between these measures. In cases where guidelines, recommendations, and significant prior protocols rely on ALBI or ICGR this enables use of DGAE imaging alone to obtain the needed information and adhere to the desired protocol. Ideally, DGAE MRI is be able to simultaneously allow clinicians to 1) characterize intrahepatic lesions, 2) quantify total liver function, and 3) guide regional function based RT optimization.

Chapter 4 explored the impact of using uptake maps as opposed to perfusion maps for functional avoidance RT planning. The results of these disparate means of determining function had a real impact on treatment planning, though, perhaps surprisingly, it may not always be as pronounced as we would suppose. This is partially because there are already intense constraints on the planned dose distribution. After ensuring the organ at risk constraints are met, and the target dose is being delivered, much freedom has already been lost, such that while it remains possible to alter the dose delivered to the highly functional regions, the results will be strongly impacted by the position of the functional regions relative to the target. Furthermore, despite relatively weak ranked correlations between k_1 and k_2 (or k_{pv}), there are many ways for the spatial distributions

of relatively high k_1 and k_2 regions to motivate a similar ultimate dose distribution. Both plans succeed in somewhat reducing dose to the specified regions. It was only in individual cases where the mismatch is most extreme that the differences were more substantial. In cases like this in particular we would expect treatment planning optimization to differ significantly, failing to reap the full benefits of functional avoidance treatment planning if we use perfusion as a surrogate for function. We might also expect additional gains for more sophisticated functional avoidance plans that account for expected functional loss based on the predicted dose response of the tissue. With increasingly granular and finely tuned optimization, we should expect the detailed differences between methods of function to increasingly matter to the function ultimately retained. An additional relevant step towards maximizing function guided RT effectiveness and minimizing RT risk is therefore a detailed statistical model of the functional response of gadoxetic uptake rate as a function of radiation dose. An analysis of pre and post RT uptake maps for the enacted RT plan would enable development of this model.

Chapter 5 has both probed the error rates of the LSITC model for very limited cases, and investigated the application of machine learning approaches to these limited datasets.
The examination of error across various analysis methods and suboptimal datasets in chapter 5 provides some insight into possible extensions. LTR data can be used to get more accurate assessments than direct application of LSITC to conventional LTR data would allow. This is easily seen in the relative stability of the error rate when using a NN. The application of a CNN based approach to 10 minutes of HTR data is a simple inclusion in the chapter with potentially large clinical implications. A reliable estimate of uptake rate from a 10 minute scan would be a meaningful improvement for patients and clinicians. However, the application of machine learning to DGAE scans in Chapter 5 only scratches the surface of the potential in this area. An additional architecture of interest for this problem is a long short-term memory (LSTM) network which would better enable the interaction of relevant information across temporally distant regions of the signal. This architecture has been applied to the determination of pharmacokinetic parameters of Gadobenic acid perfusion in head and neck scans[55].

There are additional implications on data acquisition as well. The reduction of total acquisition length achievable via the CNN approach is attractive, but there are also indications that the utility

of the timeseries is not equally spread across the timespan of the scan. This is a very intuitive result based on the structure of the characteristic concentration curves, but finds validation in the relative benefit of the "optimized" placement of points seen for the "OPT LSITC LTR" approach. Since the fundamental relationship governing the uptake rate can be viewed as a linear problem through the LSITC model, it is expected that data from the very beginning of the, and as late as possible in the hepatobiliary phase would be most beneficial in acquiring a robust fit. This model based assessment gives one answer for the characterization of which portions of the data range are most important. This could also be assessed without an assumed static model. A neural network with attention mechanisms would not only provide an interesting solution to the problem, but examination of the relative attention in different areas of the timeseries could also give an indication of the temporal location of the information most statistically salient in determining the uptake.

This work has real and immediate clinical applications, where robust measures of gadoxetic acid uptake rate allows for a simplification of overall clinical workflow through a reduction in the number of tests needed, while simultaneously allowing for more advanced tailoring of radiation therapy plans and progression.

APPENDIX A
Formulation of dual-input two-compartment equations

The change in the total amount of contrast in the distribution volume in a voxel is:

$$V_{dis}\frac{dC_{dis}(t)}{dt} = F_{ap}C_{ap}(t-\tau_a)+F_{pvp}C_{pvp}(t-\tau_{pv})-(F_{ap}+F_{pvp}+K_1)C_{dis}(t) \quad (1A)$$

where F_{ap}, and F_{pvp} are the total amounts of arterial blood plasma flow and portal vein blood plasma flow in the distribution space in the voxel, respectively, and K_1 is the amount of contrast taken up by cells per second in the voxel. F_{ap}, F_{pvp} and K_1 have a unit of vol/s.

The change in the total amount of agent in the intracellular volume in the voxel is:

$$V_i\frac{dC_i(t)}{dt} = K_1C_{dis}(t) \quad (2A)$$

Where V_i is the intracellular volume in the voxel.

Equation (1A) can be re-written as:

$$V_{dis}\frac{dC_{dis}(t)}{dt} = F_{ap}C_{ap}(t-\tau_a)+F_{pvp}C_{pvp}(t-\tau_{pv})-\left[\frac{F_{ap}+F_{pvp}+K_1}{V_{dis}}\right]V_{dis}C_{dis}(t)$$

$$= F_aC_a(t-\tau_a)+F_{pv}C_{pv}(t-\tau_{pv})-(k_2+k_1)V_{dis}C_{dis}(t) \quad (3A)$$

Where $k_2 = (F_{ap}+F_{pvp})/V_{dis}$, and $k_1 = (K_1)/V_{dis}$, both of which are rates in units of inverse seconds. F_a and F_{pv} are the total amounts of arterial blood and portal vein blood flow, respectively. Here, we use the relationship: $F_tC_t = F_pC_p+F_{red}C_{red}$, which means the total amount of blood flow times total concentration of agent in blood is equal to the amount of blood plasma flow times the concentration of agent in plasma plus the amount of blood red cell flow times the concentration of agent in red cells. Since C_{red} is zero, $F_tC_t = F_pC_p$ (agent does not enter red cells). Also, we cannot measure the concentration of agent in plasma, but can measure the concentration of agent in the total blood.

Using hematocrit *Hct*, we have $k_2 = ((1-Hct)(F_a+F_{pv}))/V_{dis}$.

The solution of (3A) is:

$$V_{dis}C_{dis}(t)= \int_0^t \left[F_a C_a(\tau-\tau_a)+F_{pv}C_{pv}\left(\tau-\tau_{pv}\right)\right] e^{-(t-\tau)(k_2+k_1)}d\tau \quad (4A)$$

The solution of (2A) is:

$$V_i C_i(t) = K_1 \int_0^t C_{dis}(\tau)d\tau \quad (5A)$$

The total account of contrast in the voxel is:

$$V_t C_t(t) = V_{dis}C_{dis}+V_i C_i+V'_e C'_e = V_{dis}C_{dis}+V_i C_i \quad (6A)$$

Where C'_e is zero.

The final solution of the total amount of agent in a voxel or volume of interest is:

$$V_t C_t(t) = \int_0^t \left[F_a C_a(\tau-\tau_a)+F_{pv}C_{pv}\left(\tau-\tau_{pv}\right)\right] e^{-(t-\tau)k_2}d\tau + K_1 \int_0^t C_{dis}(\tau)d\tau \quad (7A)$$

The unknown parameters in (7A) are F_a, F_{pv}, k_2 (or V_{dis}), K_1, and τ_a and τ_p.

Conventionally, we present blood flow in density so that after quantification of F_a, F_{pv} *and* K_1 per voxel or per volume of interest we need to calculate F_a *(*F_{pv} *or* K_1*)*$/V_{dis}$.

We can re-write the solution of (4A), and (5A) as:

$$\begin{aligned} V_{dis}C_{dis}(t) &= \int_0^t \left[F_a C_a(\tau-\tau_a)+F_{pv}C_{pv}\left(\tau-\tau_{pv}\right)\right] e^{-(t-\tau)k_2}e^{-(t-\tau)k_1}d\tau \\ &= V_{dis}\int_0^t \left[k_a C_a(\tau-\tau_a)+k_{pv}C_{pv}\left(\tau-\tau_{pv}\right)\right] e^{-(t-\tau)k_2}e^{-(t-\tau)k_1}d\tau \end{aligned} \quad (8A)$$

$$V_i C_i(t) = K_1 \int_0^t C_{dis}(\tau)d\tau = k_1 \int_0^t V_{dis}C_{dis}(\tau)\,d\tau \quad (9A)$$

Therefore, we can combine (8A) and (9A) to re-write (6A) as (2.1) and (2.2):

$$\overbrace{V_t C_t(t)}^{\text{Agent in Tissue}} = \overbrace{v_{dis}C_{dis}(t)}^{\text{Exctracellular Agent}} + \overbrace{k_1 \int_0^t v_{dis}C_{dis}(\tau)d\tau}^{\text{Intracellular Agent}} \quad (2.1)$$

$$V_{dis}C_{dis}(t)=V_{dis}\int_0^t \left(k_a C_a(\tau-\tau_a)+k_{pv}C_{pv}\left(\tau-\tau_{pv}\right)\right) e^{-(t-\tau)(k_2+k_1)}d\tau \quad (2.2)$$

APPENDIX B
Solution to three parameter linearization

We can recast equation (2.8) in the form $y = ax_1+bx_2+cx_3$, where

$$y = (1 - Hct)C_t(t),$$

$$\vec{x} = (x_1, x_2, x_3) = (C_a(t), \int_0^t C_a(\tau)\, d\tau, \frac{dC_a(t)}{dt}),$$

$$a = v_{dis}\frac{k_2}{k_1+k_2}\left(1 - \frac{k_1}{k_1+k_2}\right),$$

$$b = v_{dis}\frac{k_1k_2}{k_1+k_2},$$

$$c = -v_{dis}\frac{k_2}{(k_1+k_2)^2}.$$

Through algebraic manipulation of the final three terms we can then solve for k_1, k_2, and v_{dis} in terms of the coefficients of this model.

$$k_2 = -\frac{a}{c} \tag{1B}$$

$$k_1 = -\frac{k_2}{2} + \sqrt{\frac{k_2^2}{4} - \frac{b}{c}} \tag{2B}$$

$$v_{dis} = \frac{b(k_1+k_2)}{k_1k_2} \tag{3B}$$

REFERENCES

1. Suresh K, Owen D, Bazzi L, et al. Using Indocyanine Green Extraction to Predict Liver Function After Stereotactic Body Radiation Therapy for Hepatocellular Carcinoma. *Int J Radiat Oncol Biol Phys*. 2018;100(1):131-137. doi:10.1016/j.ijrobp.2017.09.032

2. Dawson LA, Normolle D, Balter JM, McGinn CJ, Lawrence TS, Ten Haken RK. Analysis of radiation-induced liver disease using the Lyman NTCP model. *Int J Radiat Oncol Biol Phys*. 2002;53(4):810-821. doi:10.1016/s0360-3016(02)02846-8

3. Koay EJ, Owen D, Das P. Radiation-Induced Liver Disease and Modern Radiotherapy. *Seminars in Radiation Oncology*. 2018;28(4):321-331. doi:10.1016/j.semradonc.2018.06.007

4. Ge P-L, Du S-D, Mao Y-L. Advances in preoperative assessment of liver function. *Hepatobiliary & Pancreatic Diseases International*. 2014;13(4):361-370. doi:10.1016/S1499-3872(14)60267-8

5. Johnson PJ, Berhane S, Kagebayashi C, et al. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach-the ALBI grade. *J Clin Oncol*. 2015;33(6):550-558. doi:10.1200/JCO.2014.57.9151

6. Reeder SB. Quantification of Liver Function with MRI: Is It Ready? *Radiology*. 2018;290(1):134-135. doi:10.1148/radiol.2018182251

7. Bennink RJ, Cieslak KP, van Delden OM, et al. Monitoring of Total and Regional Liver Function after SIRT. *Front Oncol*. 2014;4. doi:10.3389/fonc.2014.00152

8. Wu VW, Epelman MA, Wang H, et al. Optimizing global liver function in radiation therapy treatment planning. *Phys Med Biol*. 2016;61(17):6465-6484. doi:10.1088/0031-9155/61/17/6465

9. Wang H, Feng M, Frey KA, Ten Haken RK, Lawrence TS, Cao Y. Predictive Models for Regional Hepatic Function Based upon 99mTc-IDA SPECT and Local Radiation Dose for Physiological Adaptive RT. *Int J Radiat Oncol Biol Phys*. 2013;86(5):1000-1006. doi:10.1016/j.ijrobp.2013.04.007

10. Sørensen M, Mikkelsen KS, Frisch K, Villadsen GE, Keiding S. Regional metabolic liver function measured by 2-[18F]fluoro-2-deoxy-d-galactose PET/CT in patients with cirrhosis. *J Hepatol*. 2013;58(6):1119-1124. doi:10.1016/j.jhep.2013.01.012

11. Zou W, Dong L, Kevin Teo B-K. Current State of Image Guidance in Radiation Oncology: Implications for PTV Margin Expansion and Adaptive Therapy. *Seminars in Radiation Oncology*. 2018;28(3):238-247. doi:10.1016/j.semradonc.2018.02.008

12. Chin S, Eccles CL, McWilliam A, et al. Magnetic resonance-guided radiation therapy: A review. *Journal of Medical Imaging and Radiation Oncology*. n/a(n/a). doi:10.1111/1754-9485.12968

13. Cao Y, Wang H, Johnson TD, et al. Prediction of liver function by using magnetic resonance-based portal venous perfusion imaging. *Int J Radiat Oncol Biol Phys*. 2013;85(1):258-263. doi:10.1016/j.ijrobp.2012.02.037

14. Yamada A. Quantitative Evaluation of Liver Function Within MR Imaging. In: El-Baz AS, Saba L, Suri J, eds. *Abdomen and Thoracic Imaging*. Springer US; 2014:233-251. doi:10.1007/978-1-4614-8498-1_9

15. Sourbron SP, Buckley DL. Tracer kinetic modelling in MRI: estimating perfusion and capillary permeability. *Phys Med Biol*. 2012;57(2):R1-33. doi:10.1088/0031-9155/57/2/R1

16. Wang H, Feng M, Jackson A, Ten Haken RK, Lawrence TS, Cao Y. A Local and Global Function Model of the Liver. *Int J Radiat Oncol Biol Phys*. 2016;94(1):181-188. doi:10.1016/j.ijrobp.2015.09.044

17. Cao Y, Pan C, Balter JM, et al. LIVER FUNCTION AFTER IRRADIATION BASED UPON CT PORTAL VEIN PERFUSION IMAGING. *Int J Radiat Oncol Biol Phys*. 2008;70(1):154-160. doi:10.1016/j.ijrobp.2007.05.078

18. Bormann RL, da Rocha EL, Kierzenbaum ML, Pedrassa BC, Torres LR, D'Ippolito G. The role of gadoxetic acid as a paramagnetic contrast medium in the characterization and detection of focal liver lesions: a review. *Radiol Bras*. 2015;48(1):43-51. doi:10.1590/0100-3984.2013.1794

19. Nilsson H, Blomqvist L, Douglas L, Nordell A, Jonas E. Assessment of liver function in primary biliary cirrhosis using Gd-EOB-DTPA-enhanced liver MRI. *HPB*. 2010;12(8):567-576. doi:10.1111/j.1477-2574.2010.00223.x

20. Sourbron S, Sommer WH, Reiser MF, Zech CJ. Combined quantification of liver perfusion and function with dynamic gadoxetic acid-enhanced MR imaging. *Radiology*. 2012;263(3):874-883. doi:10.1148/radiol.12110337

21. Verloh N, Haimerl M, Zeman F, et al. Assessing liver function by liver enhancement during the hepatobiliary phase with Gd-EOB-DTPA-enhanced MRI at 3 Tesla. *Eur Radiol*. 2014;24(5):1013-1019. doi:10.1007/s00330-014-3108-y

22. Simeth J, Johansson A, Owen D, et al. Quantification of liver function by linearization of a two-compartment model of gadoxetic acid uptake using dynamic contrast-enhanced magnetic resonance imaging. *NMR in Biomedicine*. 2018;31(6):e3913. doi:10.1002/nbm.3913

23. Yoon JH, Choi J-I, Jeong YY, et al. Pre-treatment estimation of future remnant liver function using gadoxetic acid MRI in patients with HCC. *Journal of Hepatology*. 2016;65(6):1155-1162. doi:10.1016/j.jhep.2016.07.024

24. Stenmark MH, Cao Y, Wang H, et al. Indocyanine Green for Individualized Assessment of Functional Liver Reserve in Patients Undergoing Liver Radiation Therapy. *International Journal of Radiation Oncology • Biology • Physics*. 2013;87(2):S26-S27. doi:10.1016/j.ijrobp.2013.06.073

25. Van Beers BE, Pastor CM, Hussain HK. Primovist, Eovist: what to expect? *J Hepatol*. 2012;57(2):421-429. doi:10.1016/j.jhep.2012.01.031

26. Patlak CS, Blasberg RG. Graphical Evaluation of Blood-to-Brain Transfer Constants from Multiple-Time Uptake Data. Generalizations. *J Cereb Blood Flow Metab*. 1985;5(4):584-590. doi:10.1038/jcbfm.1985.87

27. Johansson A, Balter J, Feng M, Cao Y. An Overdetermined System of Transform Equations in Support of Robust DCE-MRI Registration with Outlier Rejection. *Tomography: A Journal for Imaging Research*. 2016;2(3):188-196.

28. Yoneyama T, Fukukura Y, Kamimura K, et al. Efficacy of liver parenchymal enhancement and liver volume to standard liver volume ratio on Gd-EOB-DTPA-enhanced MRI for estimation of liver function. *Eur Radiol*. 2014;24(4):857-865. doi:10.1007/s00330-013-3086-5

29. Chandarana H, Block TK, Ream J, et al. Estimating liver perfusion from free-breathing continuously acquired dynamic gadolinium-ethoxybenzyl-diethylenetriamine pentaacetic acid-enhanced acquisition with compressed sensing reconstruction. *Invest Radiol*. 2015;50(2):88-94. doi:10.1097/RLI.0000000000000105

30. Armbruster M, Zech CJ, Sourbron S, et al. Diagnostic accuracy of dynamic gadoxetic-acid–enhanced MRI and PET/CT compared in patients with liver metastases from neuroendocrine neoplasms. *J Magn Reson Imaging*. 2014;40(2):457-466. doi:10.1002/jmri.24363

31. Sommer WH, Sourbron S, Huppertz A, Ingrisch M, Reiser MF, Zech CJ. Contrast agents as a biological marker in magnetic resonance imaging of the liver: conventional and new approaches. *Abdom Imaging*. 2012;37(2):164-179. doi:10.1007/s00261-011-9734-9

32. Simeth J, Aryal M, Owen D, Cuneo K, Lawrence TS, Cao Y. Gadoxetate uptake rate as a measure of global and regional liver. In: 2020 ASTRO Annual Meeting.

33. Simeth J, Aryal M, Owen D, Cuneo K, Lawrence TS, Cao Y. Gadoxetic Acid Uptake Rate as a Measure of Global and Regional Liver Function as Compared to Indocyanine Green Retention, Albumin-Bilirubin Score, and Portal Venous Perfusion (to be submitted).

34. Yoon HI, Koom WS, Lee IJ, et al. The significance of ICG-R15 in predicting hepatic toxicity in patients receiving radiotherapy for hepatocellular carcinoma. *Liver Int*. 2012;32(7):1165-1171. doi:10.1111/j.1478-3231.2012.02784.x

35. Oellerich M, Burdelski M, Lautz HU, Binder L, Pichlmayr R. Predictors of one-year pretransplant survival in patients with cirrhosis. *Hepatology*. 1991;14(6):1029-1034.

36. Stenmark MH, Cao Y, Wang H, et al. Estimating functional liver reserve following hepatic irradiation: Adaptive normal tissue response models. *Radiotherapy and Oncology*. 2014;111(3):418-423. doi:10.1016/j.radonc.2014.04.007

37. Chandarana H, Block TK, Rosenkrantz AB, et al. Free-Breathing Radial 3D Fat-Suppressed T1-Weighted Gradient Echo Sequence: A Viable Alternative for Contrast-Enhanced Liver Imaging in Patients Unable to Suspend Respiration. *Investigative Radiology*. 2011;46(10):648-653. doi:10.1097/RLI.0b013e31821eea45

38. de Graaf W, Häusler S, Heger M, et al. Transporters involved in the hepatic uptake of 99mTc-mebrofenin and indocyanine green. *Journal of Hepatology*. 2011;54(4):738-745. doi:10.1016/j.jhep.2010.07.047

39. Abdalla EK, Denys A, Chevalier P, Nemr RA, Vauthey J-N. Total and segmental liver volume variations: implications for liver surgery. *Surgery*. 2004;135(4):404-410. doi:10.1016/j.surg.2003.08.024

40. Haimerl M, Schlabeck M, Verloh N, et al. Volume-assisted estimation of liver function based on Gd-EOB-DTPA–enhanced MR relaxometry. *Eur Radiol*. 2016;26(4):1125-1133. doi:10.1007/s00330-015-3919-5

41. Yoon JH, Lee JM, Kang H, et al. Quantitative Assessment of Liver Function by Using Gadoxetic Acid–enhanced MRI: Hepatocyte Uptake Ratio. *Radiology*. 2018;290(1):125-133. doi:10.1148/radiol.2018180753

42. Yamada A. Quantitative Evaluation of Liver Function within MR Imaging. *OMICS J Radiology*. 2012;01(03). doi:10.4172/2167-7964.1000e109

43. Ba-Ssalamah A, Bastati N, Wibmer A, et al. Hepatic gadoxetic acid uptake as a measure of diffuse liver disease: Where are we? *Journal of Magnetic Resonance Imaging*. 2017;45(3):646-659. doi:10.1002/jmri.25518

44. Saito K, Ledsam J, Sourbron S, et al. Measuring hepatic functional reserve using low temporal resolution Gd-EOB-DTPA dynamic contrast-enhanced MRI: a preliminary study comparing galactosyl human serum albumin scintigraphy with indocyanine green retention. *Eur Radiol*. 2014;24(1):112-119. doi:10.1007/s00330-013-2983-y

45. Leen E, Goldberg JA, Anderson JR, et al. Hepatic perfusion changes in patients with liver metastases: comparison with those patients with cirrhosis. *Gut*. 1993;34(4):554-557. doi:10.1136/gut.34.4.554

46. J Simeth, N Baughan, J Dow, R Ten Haken, A Johansson, M Aryal, D Owen, K Cuneo, T Lawrence, Y Cao, M Matuszak. Impact of Mis-Match Between Liver Function and Hepatic Perfusion On Functional Avoidance Treatment Planning. In: *Medical Physics*. Vol 45. ; 2018:e120-e706. doi:10.1002/mp.12938

47. Son SH, Jang HS, Jo I-Y, et al. Significance of an increase in the Child-Pugh score after radiotherapy in patients with unresectable hepatocellular carcinoma. *Radiation Oncology*. 2014;9(1):101. doi:10.1186/1748-717X-9-101

48. Simeth J, Cao Y. GAN and dual-input two-compartment model-based training of a neural network for robust quantification of contrast uptake rate in gadoxetic acid-enhanced MRI. *Med Phys*. 2020;47(4):1702-1712. doi:10.1002/mp.14055

49. Hartmann KG, Schirrmeister RT, Ball T. EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals. *arXiv:180601875 [cs, eess, q-bio, stat]*. Published online June 5, 2018. Accessed June 19, 2019. http://arxiv.org/abs/1806.01875

50. Emami H, Dong M, Nejad-Davarani SP, Glide-Hurst CK. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. *Med Phys*. Published online June 14, 2018. doi:10.1002/mp.13047

51. Ren Y, Zhu Z, Li Y, et al. Mask Embedding for Realistic High-Resolution Medical Image Synthesis. In: Shen D, Liu T, Peters TM, et al., eds. *Medical Image Computing and Computer*

Assisted Intervention – MICCAI 2019. Vol 11769. Springer International Publishing; 2019:422-430. doi:10.1007/978-3-030-32226-7_47

52. Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification. *Neurocomputing*. 2018;321:321-331. doi:10.1016/j.neucom.2018.09.013

53. Niendorf E, Spilseth B, Wang X, Taylor A. Contrast Enhanced MRI in the Diagnosis of HCC. *Diagnostics (Basel)*. 2015;5(3):383-398. doi:10.3390/diagnostics5030383

54. Cao Y, Wang H, Johnson TD, et al. Prediction of Liver Function by Using Magnetic Resonance-based Portal Venous Perfusion Imaging. *Int J Radiat Oncol Biol Phys*. 2013;85(1):258-263. doi:10.1016/j.ijrobp.2012.02.037

55. Zou J, Balter JM, Cao Y. Estimation of pharmacokinetic parameters from DCE-MRI by extracting long and short time-dependent features using an LSTM network. *Medical Physics*. n/a(n/a). doi:10.1002/mp.14222