University of Wollongong

Research Online

University of Wollongong Thesis Collection 2017+

University of Wollongong Thesis Collections

2023

Scanner Specific Uncertainty of Quantitative MRI: Assessing Consistency for Clinical Implementation

Madeline E. Carr

Follow this and additional works at: https://ro.uow.edu.au/theses1

University of Wollongong

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following: This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of this work may be reproduced by any process, nor may any other exclusive right be exercised, without the permission of the author. Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Unless otherwise indicated, the views expressed in this thesis are those of the author and do not necessarily represent the views of the University of Wollongong.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Scanner Specific Uncertainty of Quantitative MRI: Assessing Consistency for Clinical Implementation

Madeline E. Carr

This thesis is presented as part of the requirements for the conferral of the degree:

Doctor of Philosophy

Supervisor:
Senior Professor P. Metcalfe

Co-supervisors:
Assoc. Professor L. Holloway & Dr M. Jameson & Dr A. Walker & Dr K. Keenan

The University of Wollongong
School of Physics

June, 2023

This work © copyright by Madeline E. Carr, 2023. All Rights Reserved.

No part of this work may be reproduced, stored in a retrieval system, transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of the author or the University of Wollongong.

This research has been conducted with the support of an Australian Government Research Training Program Scholarship.

Declaration

I, *Madeline E. Carr*, declare that this thesis is submitted in partial fulfilment of the requirements for the conferral of the degree *Doctor of Philosophy*, from the University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. This document has not been submitted for qualifications at any other academic institution.

Madeline E. Carr

June 21, 2023

Abstract

In the past, many of the steps involved in the radiotherapy workflow have been heavily reliant on X-ray based anatomical imaging. Incorporating additional imaging modalities in this workflow has been shown to be promising. This includes aiding in generating and modifying a patient's treatment regimen and also helping determine a patient's response to treatment. Magnetic resonance imaging (MRI) is one of these modalities which can provide both enhanced soft-tissue anatomical images and supplementary information relating to changes in tissue physiology (which occurs at a faster rate than anatomical changes). Quantitative imaging biomarkers (QIBs), derived from quantitative MRI (qMRI) techniques, are measurable quantities that relate to tissue physiology (e.g., diffusion or perfusion) and thus are of particular interest in radiotherapy.

Given these capabilities, there is potential for MRI to replace X-ray imaging in several steps of the radiotherapy workflow. However, until recent years there were no qMRI quality assurance (QA) guidelines available for departments to assess the technical performance of QIBs on their MRI scanners (e.g., accuracy and repeatability). This resulted in limited work being completed that investigates QIB performance metrics; ultimately limiting the widespread clinical utilisation of qMRI techniques. These investigations are essential to determine if the quantitative values derived can be accurately used to guide radiotherapy workflow decisions.

The overall aim of the work presented in this thesis was to provide a framework for departments to use and assess the accuracy and repeatability of qMRI QIBs intended for use

in their prospective single- and multi-centre clinical trials. This work included designing protocols that followed novel QA guideline recommendations, and also investigating methods to improve QA inefficiencies. The QA protocols designed were then monitored over short (daily) and long (monthly) periods to provide evidence for the need of this type of QA. All monitoring and testing was performed using a 3T radiotherapy dedicated MRI scanner, involved using standardized sequences and phantoms, and used semi-automated image analysis tools. This type of MRI scanner in particular is primarily designed for radiotherapy simulation and planning applications.

In Chapter 3, scanner performance metrics were assessed, including finding the long-term accuracy, linearity, and repeatability of T1- and T2-relaxation times. Image acquisitions were performed monthly over one year, and baseline accuracy errors (%bias) were determined. This involved calculating the relative differences in measured relaxation times compared to known reference values. Additionally, repeatability was assessed by calculating the coefficient of variation (%CV) of T1- and T2-time measurements across all monthly scans. Results included finding a strong linear correlation between measured and reference relaxation times. Further, the baseline bias and CV for all sequences tested was $\leq$ +8.5% and 4.2%, respectively. Signal saturation was reported to have the largest impact on image analysis and performance results.

In Chapter 4, scanner performance metrics for the apparent diffusion coefficient (ADC) QIB were assessed according to the Quantitative Imaging Biomarker Alliance (QIBA) Diffusion Profile. Image acquisitions were again performed monthly over one year, and comparison metrics included ADC %bias, %CV (for short-/long-term repeatability and precision), signal-to-noise ratio, and b-value dependency. Novel investigations into factors that could impact the technical performance of ADC in the clinic were also completed. This included changing the slice-encoded imaging direction and the reference sample ADC value. All performance metrics were found to be within Profile conformance tolerance limits and thus were considered sufficient. This included finding the bias

and long-term CV to be $\leq$ +0.83% and 0.9%, respectively.

In the last study of this thesis, as described in Chapter 5, the daily accuracy and repeatability of the T1-time and ADC was determined using anatomic-specific sequences. These sequences were selected from the imaging protocols routinely performed on patients enrolled in clinical trials run by the same department housing the 3T MRI scanner. In addition, an investigation seeking to improve QA efficiencies was completed. Results found that ADC measurements carried out at standard 0°C conditions and at ambient room temperature (AT) were comparable, and thus either could be incorporated into qMRI QA. Daily T1-time and ADC (at AT) results were also within acceptable %bias ($< 20\%$) and %CV ($< 2\%$) ranges for almost all sequences tested.

Recommendations derived from this thesis included that departments should first establish long-term QIB baseline performance metrics, and then routinely implement qMRI QA. This includes at annual frequencies, along with before and after any clinical trial, and also surrounding the time of any major system hardware and/or software upgrade. Further, constructing and utilizing phantom-positioning apparatuses was recommended to minimise setup time and phantom/coil positioning variation. Accurately quantifying these performance characteristics with routine QA will aid in progressing these techniques clinically. Employing this type of QA will help ensure that an observed change in a tissue's QIB value is due to a physiological response, and not measurement variability.

Acknowledgments

And finally, after almost four years, the end is here. I could not have completed this research without financial assistance from the Australian Government Research Training Program scholarship, and the South West Sydney Local Health District Cancer Therapy Centre top-up scholarship.

Over these past few years, I have met so many incredible and intelligent people... Firstly, a big thanks to all the team/students at Ingham Institute and Liverpool Hospital. This is especially to Phillip Chlap and Dr Sirisha Tadimalla for the coding assistance. Also to Dr Robba Rai, Doaa Elwadia, and Andy Truong - without their MRI training, expertise, and scheduling - I'd have no data and thus no thesis.

I need to majorly thank my diverse team of incredible supervisors. Thanks to my primary supervisor, Senior Professor Peter Metcalfe, for all of the key brainstorming coffee chats and travel opportunities provided. Thanks to Dr Michael Jameson - who threw two phantoms on my desk in first year which ultimately moulded my thesis into what it is today. Thanks also to Dr Amy Walker and Michael - the advice given to me both in personal and academic life have been invaluable, and I am so grateful for them both staying on my supervising team since my honours year.

I would like to thank my supervisor from across the globe - Dr Kathryn Keenan - for always offering her expertise and personal time to help propel our projects. I look forward to the day I can meet and thank Katy in person. And lastly, I would like to thank the

strongest, bravest, and smartest physicist I know - Dr Lois Holloway. After so many years of supervision, I am still in awe of the dedication and passion Lois brings to her teams daily, and I aspire to be a physicist like her.

I will forever be grateful for the opportunity to undertake a PhD, and am thankful to my friends and family for encouraging me to start the PhD and see it through to the end. Without them, it wouldn't have been possible. Specifically to my incredible partner Nathan - I thank him for always being optimistic and supporting me independent of the day/frustration. And finally, I'd like to thank my Mum and Dad, for always providing endless love and support through all of the ups and downs - of both the PhD and life.

Contents

List of Figures

List of Tables

Publication List

1. **Chapter 3:** M. Carr, K. Keenan, A. Walker, R. Rai, P. Metcalfe, L. Holloway. *Determining the Longitudinal Accuracy and Reproducibility of T1 and T2 in a 3T MRI Scanner*. Journal of Applied Clinical Medical Physics, vol. 22, no. 11, p. 143-150, 2021. DOI: 10.1002/acm2.13432

2. **Chapter 4:** M. Carr, K. Keenan, R. Rai, M. Boss, P. Metcalfe, A. Walker, L. Holloway. *Conformance of a 3T Radiotherapy MRI Scanner to the QIBA Diffusion Profile*. Medical Physics, vol. 49, no. 7, p. 4508-4517, 2022. DOI: 10.1002/mp.15645

Conference Presentations

1. **M. Carr**, L. Holloway, R. Rai, M. Jameson, P. Metcalfe, A. Walker. *Development of Novel MRI Susceptibility-Only Distortion Quantification Technique*. Oral presentation: Asia-Oceania Congress of Medical Physics (AOCMP) & Engineering and Physical Sciences in Medicine Conference (EPSM). Perth, Australia, 2019.

2. **M. Carr**, M. Jameson, C. Rumley, G. Liney, M. Lee, P. Chlap, P. Metcalfe, L. Holloway. *Parameter Induced Variations in DCE-MRI Analysis: Monitoring Treatment Response for Head and Neck Tumours*. Oral presentation: MedPhys19. Sydney, Australia, 2019.

3. **M. Carr**, M. Jameson, C. Rumley, G. Liney, M. Lee, P. Chlap, P. Metcalfe, L. Holloway. *Parameter Induced Variations in DCE-MRI Analysis: Monitoring Treatment Response for Head and Neck Tumours*. Electronic poster: International Society for Magnetic Resonance in Medicine (ISMRM). France (Virtual), 2020.

4. **M. Carr**, R. Rai, P. Metcalfe, L. Holloway. *Assessing the variability of MRI derived quantitative parameters used in radiotherapy treatment planning*. Oral presentation: Engineering and Physical Sciences in Medicine Conference (EPSM). Brisbane (Virtual), 2020.

5. **M. Carr**, R. Rai, P. Metcalfe, L. Holloway. *Assessing the variability of MRI derived quantitative parameters used in radiotherapy treatment planning*. Oral presentation: MedPhys20. Australia (Virtual), 2020.

6. **M. Carr**, K. Keenan, M. Boss, P. Metcalfe, A. Walker, L. Holloway. *Conformance of a 3T MRI Scanner to the QIBA Diffusion Profile*. Electronic Poster: Engineering

and Physical Sciences in Medicine Conference (EPSM). Brisbane (Virtual), 2021.

7. **M. Carr**, K. Keenan, A. Walker, P. Metcalfe, L. Holloway. *Quantitative MRI: High Risk or Potential?*. Oral presentation: MedPhys21. Australia (Virtual), 2021.

List of Abbreviations

ART	Adaptive Radiotherapy
ADC	Apparent Diffusion Coefficient
AT	Ambient Temperature
CA	Contrast Agent
CR	Complete Response
CT	Computerised Tomography
CTV	Clinical Target Volume
CV	Coefficient of Variation
DCE	Dynamic Contrast Enhanced
DFA	Dual Flip Angle
DG	Diffusion Gradient
DWI	Diffusion Weighted Imaging
EPI	Echo-Planar Imaging
FA	Flip Angle
FOV	Field of View
GRE	Gradient Echo
GTV	Gross Tumour Volume
ICRU	International Commission on Radiation Units and Measurements
IGRT	Image Guided Radiotherapy
IQR	Inter Quartile Ranges
IR	Inversion Recovery
LT	Long-Term

MRI	Magnetic Resonance Imaging
NIST	National Institute of Standards and Technology
OAR	Organs at Risk
PK	Pharmacokinetic
PTV	Planning Target Volume
PVP	Polyvinylpyrrolidone
QA	Quality Assurance
QIB	Quantitative Imaging Biomarker
QIBA	Quantitative Imaging Biomarker Alliance
qMRI	Quantitative Magnetic Resonance Imaging
RESOLVE	Readout Segmentation Of Long Variable Echo-trains
RF	Radio-Frequency
RO	Radiation Oncologist
ROI	Region of Interest
RTP	Radiotherapy Treatment Planning
SD	Standard Deviation
SE	Spin Echo
SNR	Signal to Noise Ratio
ST	Short-term
TE	Echo Time
TIC	Time Intensity Curve
TR	Repetition Time
VFA	Variable Flip Angle
VIBE	Volumetric Interpolated Breath-hold Examination

Chapter 1

Introduction

1.1 Overview

Cancer, a disease characterized by uncontrollable cell growth, is one of the leading causes of death in Australia [1]. Radiotherapy can be used to impede this growth by aiming high doses of radiation at cancer cells (tumours) to cause irreparable damage to their DNA. Imaging plays an important role in radiotherapy, including disease diagnosis and staging, radiotherapy planning (RTP), and assessing treatment response [2]. X-ray based imaging is often used to indicate responses via changes in tumour anatomy. However, such imaging has limitations. This includes the amount of time it takes for anatomical changes to become visible, and also in that X-rays deliver additional radiation doses to a patient [3].

Magnetic Resonance Imaging (MRI) has the potential to overcome these limitations. Historically, MRI has been used diagnostically to produce high-contrast images of a patient's soft-tissue anatomy without delivering an ionising radiation dose [4, 5]. In recent years, novel quantitative MRI (qMRI) techniques have been developed. These are capable of detecting changes in the physiological properties of the body and have the potential to indicate a tissue response to treatment faster than anatomical changes [2, 3, 6–9]. If a given property is extracted from an in vivo image and can be expressed quantitatively, it can be classified as a quantitative imaging biomarker (QIB) [10].

Examples of QIBs include the apparent diffusion coefficient (ADC) and longitudinal (T1) and transverse (T2) relaxation times. These QIBs have been shown to have potential as tools for cancer staging, and treatment response monitoring and predictions [2, 3, 11–16]. Such early treatment response indicators would thus allow clinicians to adapt treatment plans earlier than conventional response assessment methods. This would aid in enhancing treatment to better suit the patient's physiological needs, and thus improve their quality of life post-treatment [4, 17, 18].

To use these QIBs for clinical decision-making in radiotherapy, it is essential that their values have high accuracy and repeatability. This is especially the case given that the quantitative values derived could be used to guide the radiotherapy plan adaptation and thus affect patient outcomes [6, 13, 19]. However, there are minimal studies in the literature that report on assessing these performance metrics on MRI scanners which are routinely used to compute patient-based QIBs [7, 11, 18, 20, 21].

To validate the performance of QIBs, standardized imaging sequences and phantoms are required along with the development of standardized QIB quality assurance (QA) protocols [2, 7]. This thesis focused on developing such QA protocols for a 3T radiotherapy dedicated MRI scanner. These protocols were then routinely implemented on the scanner to assess its technical performance in deriving the ADC, and T1- and T2- relaxation times.

Specifically, Chapter 3 investigated the accuracy and long-term repeatability of T1- and T2-relaxation times. Chapter 4 investigated similar technical performance measures for ADC, along with delivering the outcome of novel ADC performance testing which used the 2019 Quantitative Imaging Biomarker Alliance (QIBA) Diffusion Profile. The short-term intra-scanner ambient temperature ADC and T1-time variability were also investigated in Chapter 5 using anatomy-specific sequences; simulating serial daily patient scanning as would occur during typical radiotherapy treatment.

This thesis overall aims to increase knowledge on the reliability of MRI-derived QIBs. A summary of this work and an overview of the potential future works from this thesis are provided in Chapter 6. The results from this thesis will be of particular use to the department housing the 3T MRI scanner, but also for external departments aiming to develop their own QA protocols and decide on the frequency in which they should be carried out. This aims to assist in translating qMRI techniques and their QIBs into the clinic to improve personalised cancer therapies and post-treatment outcomes for patients.

1.2 Aims and Objectives

The aims of this thesis are to:

Aim 1: Quantify the longitudinal intra-scanner accuracy and repeatability of T1 and T2 relaxation times.

Research Question (Chapter 3): How accurate and reproducible are the T1- and T2-relaxation times extracted from a clinical MRI scanner, when measured monthly, over a one-year period?

Aim 2: Assess the overall technical performance of a scanner measuring the ADC.

Research Questions (Chapter 4): How accurate, repeatable, and reproducible is the ADC value that is extracted from an MRI scanner, according to the Quantitative Imaging Biomarker Alliance (QIBA) Diffusion Profile? What impact does alternating the slice-encoding imaging direction have on ADC technical performance?

Aim 3: Monitor and quantify the daily fluctuations of measured T1-relaxation time and ADC values, when using anatomy-specific qMRI sequences on a clinical MRI scanner.

Research Questions (Chapter 5): Is it possible to simplify ADC qMRI QA by measuring ADC samples at ambient temperature? What intra-scanner variability exists in terms of accuracy and repeatability of T1-time and ADC when using anatomy-specific sequences daily?

Chapter 2

Literature Review

The purpose of this chapter is to provide an overview of radiotherapy and particular MRI techniques that have the potential to be used within the radiotherapy workflow. To start, the current radiotherapy workflow will be described, which will then be followed by an outline of MRI physics and specialised MRI techniques. The last section of this chapter will briefly cover the current state of quality assurance protocols available for MRI, and some of the prospective needs in this space for ensuring reproducibility; all of which will assist in the clinical implementation of specialised MRI techniques into radiotherapy.

2.1 Cancer and Radiotherapy

Rapid and abnormal cell multiplication in the body occurs due to a mutation in the cell's DNA. When multiplying cells form groups, they are known as tumours which can be classified as malignant (cancer) or benign (not cancer). Malignant tumours can break away from their primary site, spreading through either the bloodstream or lymphatic system, and can have detrimental effects on a body if left untreated. Globally, cancer is one of the leading causes of death. In 2020, approximately 19.3 million new cancer cases were diagnosed and nearly 10 million deaths were caused by the disease [22].

There are continuous improvements made to cancer therapy treatments; some of which have led to the increase of cancer survival rate post-5 years of diagnosis in Australia.

This includes survival rates increasing from less than 50% in the 1980s to 70% today [1]. Radiation therapy, or radiotherapy, is one of the most common modalities used to treat cancer, with approximately 50% of newly diagnosed cancer patients being recommended radiotherapy to ensure an optimal outcome [23]. Radiotherapy directs radiation (e.g. X-rays or particle beams) at tumour sites, with the overall aim to damage beyond repair the cancer cell's DNA to try and stop the rapid cell multiplication from occurring.

Radiotherapy thus requires a high level of precision when both delineating the tumours and directing the radiation to successfully damage the cancer cells. This is whilst sparing the surrounding healthy tissues and organs at risk (OAR). Carefully conducting each step of the radiotherapy process is key to better treating and then improving the quality of life for a patient following their treatment.

The standard treatment pathway for radiotherapy patients commences once a diagnosis and treatment method are agreed upon by a radiation oncologist (RO) and the patient. Radiotherapy treatment planning (RTP) then involves using imaging to simulate the patient's anatomy in their intended treatment position. Using these images, contours of the target tumour volume and surrounding OARs will be created and calculations are performed over the entire contoured volume to simulate the dose delivery. To ensure the correct dose (as prescribed by the RO) will be delivered to the tumour, quality assurance (QA) is carried out by medical physicists.

Radiotherapy treatment then involves delivering the planned treatment dose (ionising radiation) to the patient. For the purpose of this thesis, delivery via external beam radiotherapy using a linear accelerator (linac) will be of focus. Depending on the RO's dose prescription, treatment on the linac will be carried out in a number of 'fractions' and over multiple days or weeks. At the end of standard radiotherapy treatment, follow-up assessments by the RO and their clinical team are performed.

These assessments generally involve anatomical imaging of the tumour site to assess the response of the tumour; dependent on anatomical changes observed according to a criterion of classification, as outlined in Table 2.1. Physiological imaging of the tumour can be acquired complementary to the anatomical imaging and can provide an RO with enhanced treatment response assessment tools [3, 16, 21, 24, 25]. The use of both anatomical and physiological imaging techniques in radiotherapy is briefly discussed below.

Table 2.1: Classification of treatment responses, defined by the World Health Organisation (WHO) [26].

Response Type	Acronym	Definition
Complete Response	CR	Complete disappearance of disease
Partial Response	PR	$\geq$ 50% decrease in total tumour size
No Change	NC	No PR nor $\geq$ 25% size increase in any lesions
Progressive Disease	PD	$\geq$ 25% lesion increase or new lesions appear

2.1.1 Adaptive Radiotherapy

Adaptive Radiotherapy (ART) is a specialized cancer therapy technique that adapts a radiotherapy plan during treatment, depending on early observations seen in the particular patient. Literature such as Li et al. (2011) provides a comprehensive review of this specialised technique [3].

As an overview of ART, a feedback loop is implemented at the time of treatment delivery (Figure 2.1). It is designed to detect and respond to signs of changes in the body during treatment. The type of feedback detected can be variable, including changes in anatomy (e.g., the tumour physically growing throughout treatment) or in the biology (e.g. tumour cellularity is increasing). Depending on the individualized feedback detected, clinicians can make an updated decision on the treatment regime and make changes accordingly, with an aim to improve outcomes for the patient.

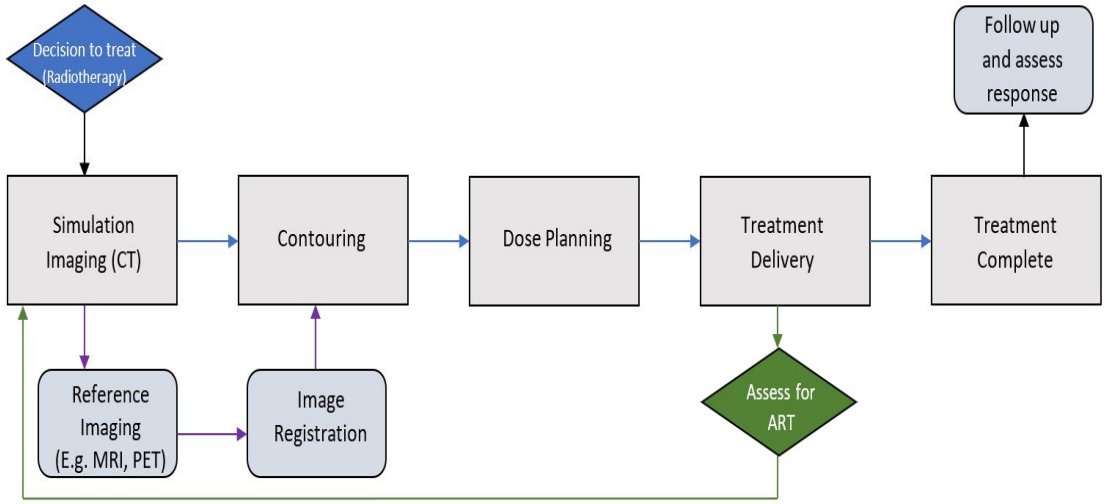

Figure 2.1: General radiotherapy workflow, with the addition of an adaptive pathway option. The choice of the selected adaptive pathway will be based on the observed change detected during imaging, which occurs just before treatment delivery (e.g. anatomical or biological change in the tumour). Simulation imaging is often performed using both computed tomography (CT) and magnetic resonance imaging (MRI).

2.1.2 Imaging for Radiotherapy

X-ray based computerized tomography (CT) is the most common imaging modality used to acquire images for RTP contouring. Multi-modal tumour localisation (and image registration) is occasionally completed using additional imaging modalities such as MRI when superior soft-tissue contrasting capabilities are required (e.g. in Brain) [3, 27]. According to the International Commission on Radiation Units and Measurements (ICRU) guidelines [28], planning volumes can be defined as:

- Gross Tumour Volume (GTV): Volume of primary tumour and metastatic sub-tumours.
- Clinical Target Volume (CTV): Volume of tissue considered to require radiotherapy (which contains an observable amount of GTV and/or sub-clinical malignant disease).
- Planning Target Volume (PTV): Geometrical concept of volume used to ensure the desired dose is delivered to the CTV with accepted clinical probability.
- Organ At Risk (OAR): Organs or tissues that are at risk of severe damage if irradiated.

As noted, accurately defining these volumes is one of the most important aspects of RTP;

ensuring sufficient dose is delivered to the tumour whilst sparing the healthy tissues surrounding [27]. The GTV and CTV are defined during the contouring process and are created relative to the volumes of known and alleged tumour locations, respectively (see Figure 2.2). In contrast, the PTV is defined during the dose planning stages and is designed to guarantee that the GTV and CTV receive their prescribed doses during treatment.

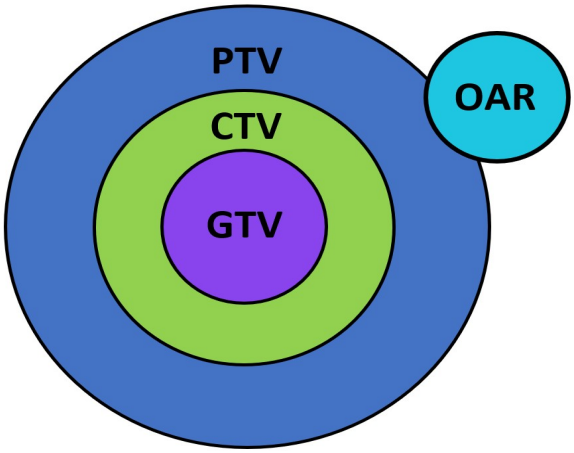

Figure 2.2: Simplified schematic diagram of different volumes delineated in the RTP process. This includes the Gross Tumour Volume (GTV), Clinical Target Volume (CTV), Planning Target Volume (PTV) and Organs At Risk (OAR).

In the past, only a single pre-treatment CT image was acquired for all planning and treatment; resulting in the requirement of very large PTVs. This was to account for the potentially large variability in tumour position, size and shape which could occur throughout treatment. Dose constraints on the surrounding healthy tissues however limited the dose that could be delivered to the tumour [4]. To accurately decrease the size of the PTV and boost tumour dose, high confidence in the tumour's position during treatment is required.

Image-guided radiotherapy, described below in section 2.1.3, employs imaging (usually X-ray based) in the treatment room to visualise the patient's anatomy on the day of each fraction delivery. This allows the clinician to gain confidence in tumour position (thus reducing the PTV margin) and perform image registration back to the reference simulation CT. Image registration thus examines if the patient's anatomy on the day of treatment matches that observed during simulation. This is completed to ensure that the dose is

correctly delivered to the planned volumes.

Cone-beam CT (CBCT) allows for this type of soft-tissue imaging. Specifically, CBCT has demonstrated improved target dose coverage compared to skin tattoos for example [4]. This is where tattoos have been used in the past to identify a tumour's corresponding surface location and thus assist in patient positioning. However, these are not always perfect surrogates for internal tumour location. Advanced radiotherapy techniques which deliver high doses in single fractions, such as intensity-modulated radiotherapy (IMRT), would not be possible without high levels of assurance that the patient is in the correct setup position [29].

Treatment response assessment, however, is still majorly assessed based on anatomical changes such as those outlined in Table 2.1. Using X-ray based imaging for all steps described has limitations, including additional radiation dose being delivered to the patient during imaging. Further, structural changes to anatomy take longer amounts of time to become observable than changes occurring at the cellular and molecular level [2, 3]. For these reasons, biological imaging is becoming increasingly of interest in the field of radiation oncology for disease diagnosis and response assessment [3, 4].

2.1.3 Image Guided Radiotherapy

Image-guided radiotherapy (IGRT) is a generalized term which includes any type of imaging completed in the radiotherapy treatment room. It can be defined for in-room imaging (e.g., patient alignment, adjusting for tumour motion or ART assessment) and for biological imaging [29]. In-room imaging can be performed immediately before treatment, which is useful to visualise inter-fraction anatomy changes. However, the information gained from these images acquired before treatment does not account for any intra-fraction anatomy movements (e.g. breathing, bladder filling, or muscular motion).

Fundamental developments in novel technologies now allow clinicians to acquire in-room

images during treatment, and adapt radiotherapy delivery in real-time [30]. This allows for increased accuracy in tumour targeting, reducing the dose to surrounding healthy tissues, and ultimately improving the quality of treatment [29]. There are several ways to implement IGRT to assist in real-time ART, and this is covered extensively in the literature [29, 30].

MRI is one of the imaging modalities that can be used to aid in the contouring and planning stages of the RTP pathway. This modality can supply high-quality soft-tissue anatomical images as well as physiological information through biological imaging. It also is starting to be used to improve ART treatments using real-time tumour and OAR tracking with hybrid MR-linear accelerators (MR-Linac).

2.2 The Physics of MRI

MRI is an imaging modality that exploits the known magnetic properties of tissues in the body to generate images of the underlying structures. A high level of complex quantum and electromagnetic physics is required to explain how these images are generated and can be found in detail in the literature (e.g. Tofts. et al. (2003)) [31, 32].

The overall design of an MRI scanner can be seen in Figure 2.3. Generally, clinical MRI scanners have main magnetic B_0 field strengths ranging between 0.3-4 T [33], and have gradient coils and radio-frequency (RF) coils to generate secondary magnetic fields. These are most commonly tuned to selectively interact with protons.

As a simplified overview of MRI physics, the human body has an abundance of protons (in the form of Hydrogen nuclei), and these are targeted in MRI. Protons have 'spin': a property enabling them to act like small bar magnets which are usually randomly aligned in the body due to thermal energy [33]. When protons are subject to a strong constant external magnetic field (B_0) such as in an MRI scanner, the axes' of the protons line up with the external field either in a parallel or anti-parallel direction.

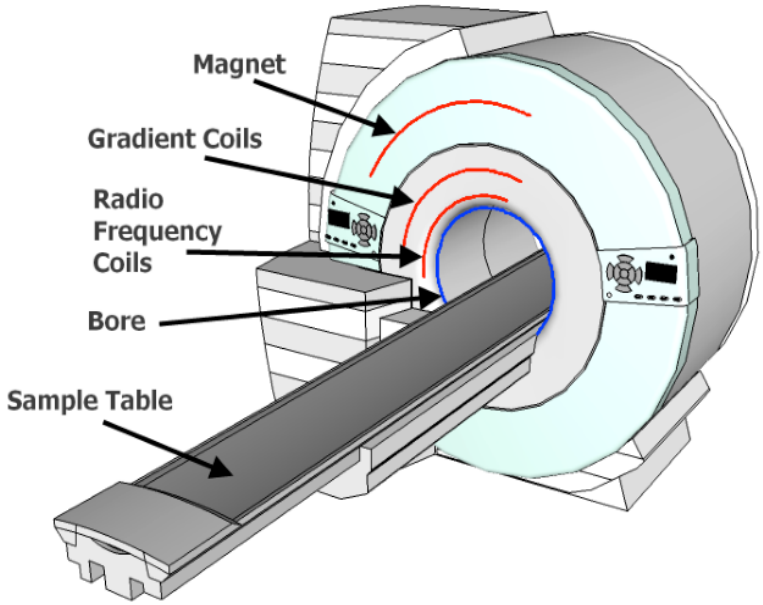

Figure 2.3: Schematic of a typical MRI scanner taken from Haynes et al. (2013) [34], showing the main-magnet (B_0), and gradient and radio-frequency (RF) coils position .

There are different components to the net magnetization ($\vec{M}$) that the protons experience: longitudinal magnetization (M_z), which is generated in line with B_0, and transverse magnetization (M_x and M_y, or M_{xy}), which is generated perpendicular to B_0. At equilibrium, when subjecting protons to B_0, the longitudinal magnetization parallel to B_0 (lower energy state) is at a maximum (M_0), whilst M_x and M_y are zero. This alignment is not static, but instead involves precessing in a circular motion around the external field at the Larmor frequency, ω_0. This frequency (Equation 2.1) is proportional to the strength of B_0 and the gyromagnetic ratio (γ) (which is measured in MHz/Tesla (T), and is unique for each element).

$$\omega_0 = \gamma B_0 (rad/s) \tag{2.1}$$

When energy in the form of RF pulses (B_1) is transmitted from the RF coils at the ω_0 for Hydrogen (42.57 MHz/T), protons in the lower energy state are able to absorb this energy and jump into higher energy states. This effectively rotates the direction of net magnetization ($\vec{M}$) in the tissue to be orientated in the transverse plane (Figure 2.4). The flip angle (FA = α) is usually given in degrees ($^\circ$) or radians (R), and represents the amount of rotation $\vec{M}$ experiences during the application of the RF pulse.

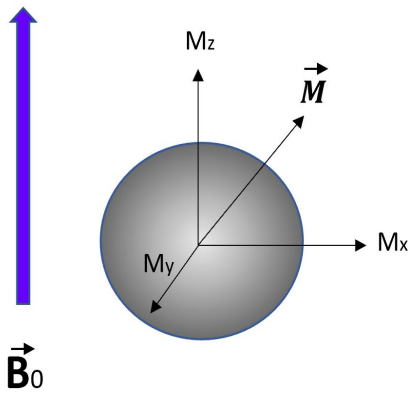

Figure 2.4: Excited proton with three associated magnetic field vector components (M_z, M_x, M_y), which add to give the net magnetisation $(\vec{M})$.

2.2.1 Characteristic Times

Immediately following the application of B_1, the net magnetization vector begins to return to its equilibrium state (M_0) through the process of relaxation. This process involves energy loss with the net magnetization vector gradually realigning with the z-axis (longitudinal relaxation), while the spins simultaneously lose their in-phase coherence in the transverse plane (transverse relaxation) and start to precess at their Larmor frequencies [33].

These changes in precession give rise to an oscillating magnetic field which induces a current in the scanners receiver coil; analogous to the effect of a vibrating string on an electric guitar. This current is registered as an MR signal. Additionally, gradient coils are used to modify the strength of the magnetic field as a function of spatial location within the scanner. This variation allows for the location of signal origin to be calculated from the frequency and phase of the detected current (allowing for spatial encoding of the MR signal) [33].

The characteristic time it takes for M_z to reach 63% of its original magnitude (M_{z_0}) is called T1. During the regrowth of M_z, the protons dissipate their energy into the surrounding lattice environment and thus T1 is also known as the longitudinal or "spin-

lattice" relaxation time. Without additional RF pulses applied, the regrowth of M_z occurs exponentially (see Equation 2.2).

$$M_z(t) = M_0(1 - e^{-\frac{t}{T1}}) \tag{2.2}$$

In terms of the transverse magnetisation (M_{xy}), as noted it begins to decay simultaneously to M_z, but at a faster rate (Figure 2.5). T2 is the characteristic time for transversal or "spin-spin" relaxation, and is the time associated with the transverse magnetic field decaying to 37% of its original strength. The decay of M_{xy} is of mono-exponential nature and can be described by Equation 2.3.

$$M_{xy}(t) = M_0(e^{-\frac{t}{T2}}) \tag{2.3}$$

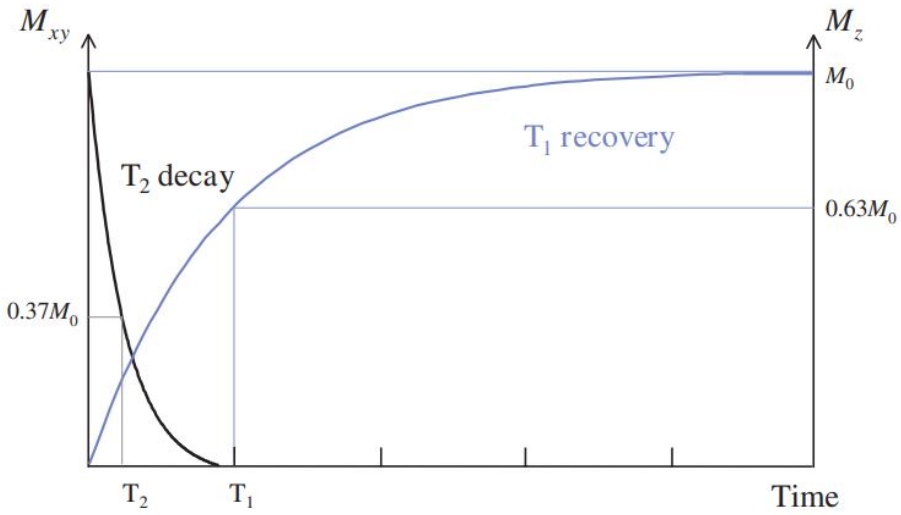

Figure 2.5: Different durations of MRI characteristic times, T1 and T2, which occur simultaneously, however at different rates. T1 refers to 63% of the recovery of the longitudinal magnetization (M_z), whilst T2 refers to 37% the decay of the transverse magnetization M_{xy}, following the stimulation from an RF pulse. Image taken from McRobbie et al. (2006) [35].

There is a third characteristic time defined as T2* (the 'spin-spin dephasing' time), which accounts for the fact that M_{xy} decays at a much faster rate than that mathematically predicted by Equation 2.3. This faster decay rate can be due to either external or internal factors, including inhomogeneities in the external magnetic field or the susceptibility properties of the tissue and surrounds.

It is known that T1- and T2-times differ for different tissues, and these properties are exploited to gain better contrasted images of tissues and organs of interest [36]. The way MRI pulse sequences are designed with specific imaging parameters to take advantage of the tissues properties is well summarized in the literature [31, 33, 36].

2.3 Quantitative MRI

In the past, clinical MRI has predominantly been used as a tool for qualitative and diagnostic imaging [3]. Quantitative MRI (qMRI) on the other hand can be defined as the extraction of a physical or biological feature from an image obtained from an MRI scanner, that can be expressed as a metric with units (or given as a relative quantity compared to a reference material) [7, 37].

This technique has been explored within the research domain for several decades and is starting to transition into routine clinical applications. This has been with the aid of significant developments in MRI hardware, image sequences, analysis, and reconstruction methods in recent years, enabling improvements of such advanced techniques [3, 7].

Recall from section 2.1.2 that biological imaging is becoming of increased interest in radiotherapy. Specifically, a quantitative imaging biomarker (QIB) can be defined as *"an objective characteristic derived from an in vivo image measured on a ratio or interval scale as an indicator of normal biological processes, pathogenic processes or a response to a therapeutic intervention"* [10].

Although QIBs can be found using many imaging modalities, qMRI techniques such as Diffusion-Weighted Imaging (DWI) and Dynamic Contrast Enhanced (DCE) MRI have demonstrated significant potential to produce QIBs for radiotherapy purposes and are extended upon below. These purposes include assisting in target definitions [3, 38], disease detection [21, 25], and assessing and monitoring treatment responses [3, 16, 24].

Magnetic resonance fingerprinting (MRF) is an alternative and novel MRI technique, which has significant potential for advancements in qMRI. This technique allows for the simultaneous measurement of multiple tissue properties, often referred to as multi-parametric maps; reducing image acquisition time. Essentially, MRF matches the signals emitted by different tissue types in the body to a pre-calculated 'dictionary' of potential tissue signal patterns (like T1 and T2 relaxation times) [2, 39].

Successfully implementing MRF into the radiotherapy workflow could improve patient comfort and minimize motion artefacts; offering a significant advantage over other qMRI techniques [39, 40]. However, as Ding et al. (2021) highlighted in their review, there is a need for additional research to be conducted before MRF can be confidently used by clinicians for assessing tumour QIB response [39]. Thus, the focus for the remainder of this thesis will be on established qMRI techniques, namely DWI- and DCE-MRI.

2.3.1 Diffusion-Weighted Imaging (DWI)

Diffusion describes the process of particles moving along a concentration gradient. Each fluid has an individualized diffusion coefficient, D (m^2/s or μm^2/ms), that represents the fluid's molecular mobility (or Brownian motion). Given the movement of water can be represented as the movement of protons (situated within the Hydrogen nuclei), it then becomes evident that MRI can be used to measure diffusion-related properties of tissues.

In many fluids, the diffusion of molecules acts equally in every direction giving them isotropic properties, which can be characterized by a scalar D. However, in biological tissues, the diffusion properties of the water molecules are affected by various structures, such as cell membranes, nerve fibres, and macromolecules [31, 38]. This causes the diffusion in biological tissues to be directionally dependent (i.e., have anisotropic properties).

It is assumed in DWI-MRI analysis that the signal intensity in diffusion images will experience a mono-exponential decay with increasing b-values (see below, including Equation

2.5). However, non-mono-exponential decay may be observed for biological tissues with such anisotropic properties. Thus, the term 'apparent diffusion coefficient (ADC)' is instead used to describe water mobility and diffusivity within tissues.

In DWI-MRI, a b-value (b) is an acquisition parameter that determines the amount of diffusion-weighting applied to a sequence. Differences in image signal intensities are caused by changing the b-value, whereby the number and magnitude of the b-values selected are generally tissue-dependent and require a compromise between the desired diffusion-weighted (DW) contrast, signal-to-noise (SNR) and acquisition times [21, 41, 42].

The b-value is measured in s/mm^2, and is dependent on γ, the time duration of the diffusion pulse (δ), and the time interval between (Δ) and amplitude ($|g|$) of the two diffusion gradients (see Figure 2.6). Mathematically, this parameter can be defined by equation 2.4 [31, 43]:

$$b = \gamma^2 \delta^2 |g|^2 (\Delta - \frac{\delta}{3}) \tag{2.4}$$

One method proposed by Stejskal and Tanner in 1965 used the addition of two high-amplitude gradients in a spin-echo (SE) sequence to introduce diffusion weighting into an MRI acquisition [31, 43]. Utilizing these principals, echo-planar imaging (EPI) is a common sequence type used for DWI-MRI clinical applications [31]. An EPI can encode the k-space and capture an entire 2D planar image using multiple or a single shot of RF-excitation pulses. In order to measure the ADC, at least two different DW images are needed, which are obtained at two distinct b-values [31].

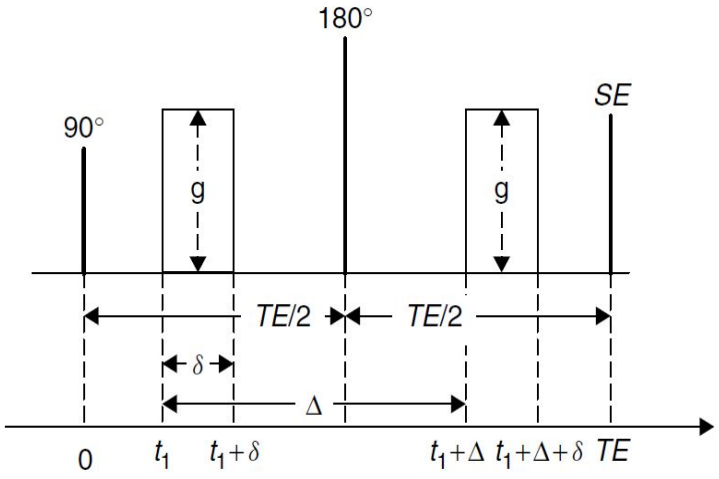

Figure 2.6: Schematic diagram of a Stejskal–Tanner pulsed field gradient (PFG) diffusion weighted spin-echo scheme. This scheme is used by pulsed gradient spin-echo (PGSE) sequences to measure diffusion. Image taken from Tofts et al. (2003) [31].

The first signal, (S_0), is generated with very low (or no) DG switched on, creating 'b_0 images' which are T2-weighted. Following, the DG's are switched on using a non-zero b-value to generate a second signal (S(b)) and corresponding image with a particular diffusion-weighting. This can be repeated with multiple non-zero b-values. In the case of 2 b-values, the ADC is calculated for each voxel in an image according to equation 2.5.

$$S(b) = S_0 e^{-b.ADC} \rightarrow ADC = -\frac{1}{b}\ln(\frac{S(b)}{S_0}) \tag{2.5}$$

Another ADC-mapping sequence is called ZoomIt [44]. As an overview, ZoomIt comprises of a Siemens-specific EPI sequence which is combined with a reduced volume excitation technique. This sequence is often used clinically since it can produce high-quality images without long acquisition times. Shorter acquisition times are achieved by meticulously only exciting (or zooming in to) regions of interest in the phase-encoded (PE) (critical) direction. This reduces the amount of encoding information collected, and thus the computation time required to produce an image [44, 45].

Similarly, Readout Segmentation Of Long Variable Echo-trains (RESOLVE), is another common Siemens manufactured sequence [46]. As the name suggests, the technique uses a read-out segmented EPI to reduce the sampling time. This in effect minimizes

susceptibility distortions and T2* blurring effects, which are known to affect DWI-EPI sequences [46–48]. Sequences such as these are often used for imaging specific anatomy in the body which is prone to susceptibility effects or requires shorter acquisition times (e.g., lungs). The composition of RESOLVE and ZoomIt sequences are out of the scope of this thesis, however can be found in the literature [44, 46, 47].

2.3.2 Clinical Potential of DWI-MRI in Radiotherapy

The ADC of tissues is impacted by factors such as tissue cellularity and membrane permeability [21, 38]. In general, malignant tissues have lower ADC values due to their increased cellularity and structural disorder, as caused by the rapid cell multiplication [21, 38]. Measuring the ADC value of tissues in past studies has been found to assist in tumour characterization, such as in renal lesions [49] and breast tumours [50].

Measuring the changes in ADC of tissues over time can also indicate cellular structure and physiological changes, as shown in Figure 2.7. These changes would otherwise be invisible using other imaging methods [21]. When radiotherapy causes sufficient damage to the cancer cell's DNA for example, cellular membranes break down. This in turn reduces the cell density and increases the ADC value. These changes can be detected earlier than anatomical changes during treatment, and can thus allow for earlier adaption to the treatment plan [14, 18].

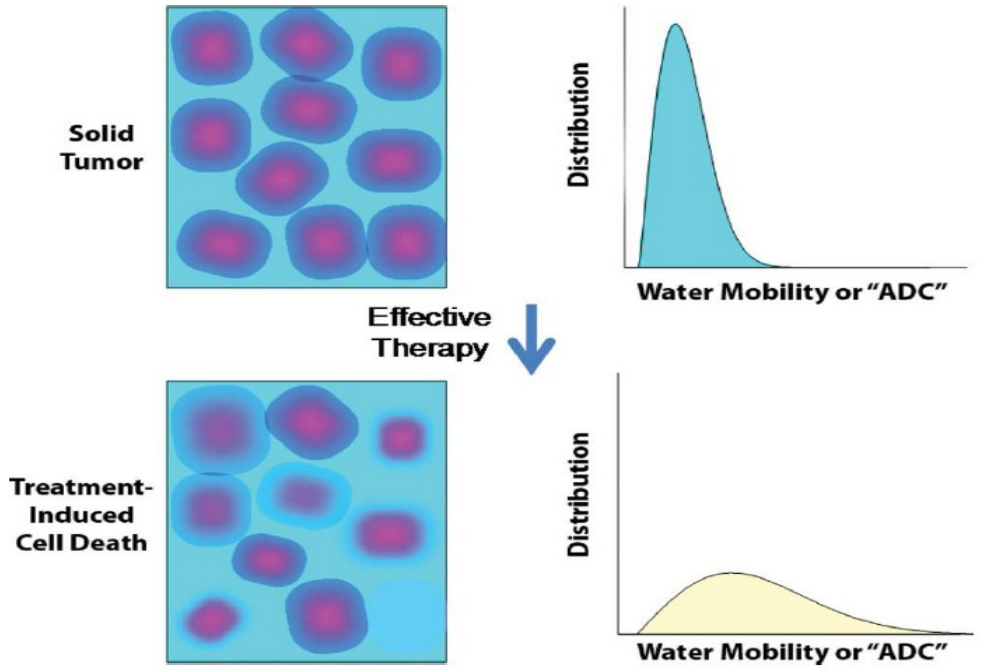

Figure 2.7: Schematic, taken from Thoeny et al. (2010) [8], displaying the effect of successful therapy on the distribution of ADC within a tumour: Cellular density reduces, increasing the diffusivity and thus increasing the mean tumour ADC value.

Monitoring ADC changes over the course of treatment has shown potential to assist in outcome prediction for tumours in a variety of regions in the body [14, 38, 51–55]. This includes predicting primary rectal cancer tumour response to neoadjuvant (delivered before the main treatment) combined radiotherapy- and chemotherapy (CRT) [54, 55]. For example, in a study by Lambrecht et al. (2012), patients with pathologic complete response (pCR) had significantly lower pretreatment tumour ADC values, compared to patients with no pCR [54]. Further, the change in ADC (ΔADC) was found to be larger for patients with pCR.

Similar results have been found for head and neck squamous cell carcinoma (HNSCC) cancer patients. For example, ADC data acquired throughout treatment found that primary tumours with lower %ΔADC predicted local failure when treated with CRT, compared to those with locoregional control [14, 56]. This finding was confirmed in several other HNSCC predictive response studies that were included in a systematic review by Chung et al. (2018) [53].

Given approximately 30-40% of HNSCC cancer patients still have locoregional disease failure [56], finding a reliable QIB that can indicate failure before treatment is of high clinical interest. Such indications would give clinicians the confidence required to decide to adapt a treatment plan earlier, or to use alternative treatment methods which would spare the surrounding healthy tissue from unnecessary radiation toxicity [53, 56].

Methods providing early indications of recurrence for cervical cancers are also desired. This is due to approximately 40% of this patient cohort being at risk of a recurrence developing [57]. A systematic review by Schreuder et al. (2014) concluded that of all cervical cancer ADC studies completed between 2005 and 2014, a threshold ΔADC within the tumour between pre- and post-treatment MRI scans could predict either CR or no-response [58]. With several studies since reporting on similar observed changes in cervical cancer tumour ADC values with clinical outcomes [9,57,59,60], ADC shows promising potential to aid in methods for monitoring and/or treating cervical cancers.

Correlating observed early changes in tumour ADC with response has also been investigated in the brain [31, 61], breast [52], prostate [16] and lungs [62]. The majority of radiotherapy response studies report larger increases in ADC post-treatment for responders compared to non-responders [18]. Such findings highlight the potential of ADC to assist in early response detection.

When determining the exact ADC thresholds for tissue characterisation and responses, the imaging parameters used for the data acquisition need to be taken into account [21]. This makes it increasingly difficult when there are other study limitations including biases in the patient cohort selection [49, 58, 62], or the fact that some studies use multiple scanners to make conclusions [49]. Several additional factors can impact the accuracy of DWI-MRI, which are summarised by Winfield et al. (2014) [6]. This includes EPI-based sequences' insufficient SNR and non-linearities in diffusion gradient field strengths; all leading to biased ADC values.

Additional factors that can influence ADC values includes the presence of magnetic field inhomogeneities. Such field variations can originate from several causes, such as susceptibility effects at the interface of different structures (e.g., tissue and air), or the influence of eddy currents; both of which are common in DWI sequences [2, 6]. These field inhomogeneities can lead to geometric image distortions: as the proton spin frequency is impacted by changes in the magnetic field, this can result in the output signal being mapped to the wrong location [63]. This in effect can alter the perceived location of the signal and measured ADC value.

There is hence a significant need to validate the repeatability of ADC before it can be used in widespread clinical applications. Further, ensuring a standardized approach to data acquisition and image analysis between centers/scanners is essential [6, 7, 21, 64]. QA methods attempt to address these needs and noted challenges and will be discussed in detail in section 2.4.

2.3.3 Dynamic-Contrast Enhanced MRI

Dynamic-Contrast Enhanced (DCE) MRI is a qMRI technique capable of quantifying blood flow, tissue perfusion, and microvascular wall permeability [12, 19]. The technique involves using a dynamic and highly T1-weighted sequence to observe contrast agent (CA) up-take into a tissue after a CA is administered into the peripheral vein [65]. A pre-contrast T1-map is also commonly acquired so that changes in T1-time (caused by the CA) can be tracked from its baseline value ($T1_0$).

As the CA is being taken up, a sufficient temporal resolution is required to effectively determine the signal enhancement rate as caused by the CA [65, 66]. This signal is mapped as a time-intensity curve (TIC) and is then converted to a concentration. This concentration is then used in mathematically fitted and calculated pharmacokinetic (PK) parameters which are capable of characterizing particular tissue cell properties (e.g., plasma volume

fraction- see Figure 2.8). Yankeelov et al. (2009) and Tofts et al. (2003) provide a comprehensive review of DCE-MRI and PK-modelling theory [65].

2.3.4 T1 Mapping for DCE-MRI

For reliable quantitative measurements to be obtained from DCE-MRI, pre-contrast T1-times are required to be either known or able to be accurately measured ($T1_0$) [19, 31, 65, 67–69]. This is in order to have adequate conversions from the TIC to concentration, and thus accurate PK parameter generation. For regions of increased homogeneity, the simplest method of T1-time estimation for a tissue is to assign its value to the tissue's T1-time reported in the literature [68, 70].

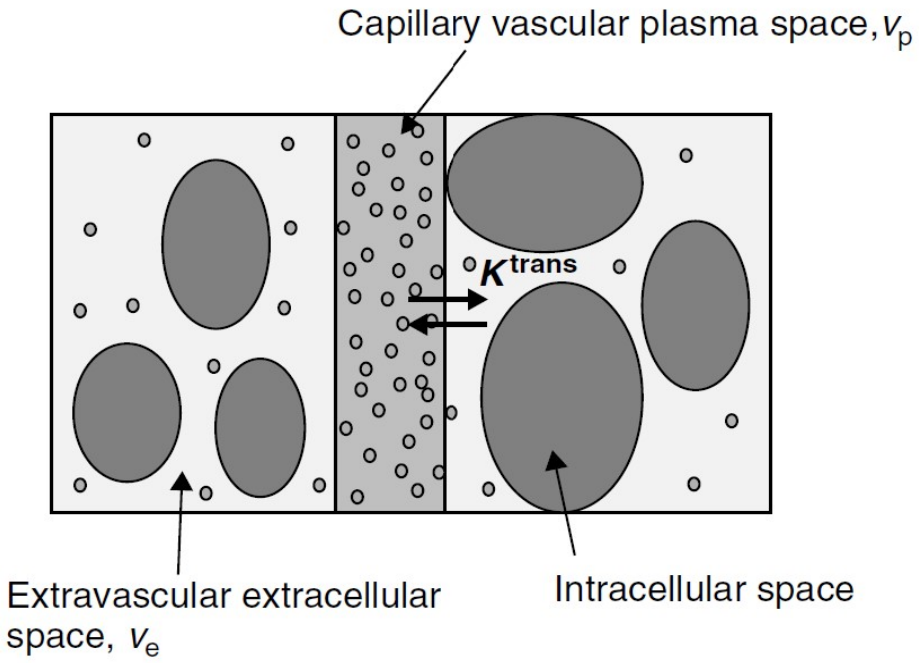

Figure 2.8: Graphical depiction of the major compartments involved in contrast agent movement for DCE-MRI, taken from Tofts et al. (2003) [31]. The small grey circles represent the contrast agent particles, while large ovals identify the cells. Here, K_{trans} is another PK-parameter called the intravascular-extravascular transfer rate constant. It is displayed along with v_e and v_p, and is dependant on the capillary wall's surface and permeability.

The prostate for instance, has a broadly accepted T1-value of 1400-1700ms [68, 70, 71]. The problem with such an assumption is it lacks accounting for patient individuality, including age and inhomogeneity between regions of the organ/tissue (e.g., glands in the prostate) [68]. Thus, it has been suggested that in order to achieve the most accurate

PK-parameter output results, patient-specific T1-mapping should be employed [68]. For example, the head and neck (HN) region has increased heterogeneity, low SNR, and larger variations in magnetic susceptibilities, which can make assumptions of the T1-time to one value increasingly inaccurate [72].

The current gold standard method for measuring T1-time is using an inversion recovery (IR) sequence [31, 72]. This approach includes changing the time of inversion (TI) in a modified SE sequence to generate 3D T1-maps. This method however requires extensive acquisition times and is thus difficult to implement clinically [7].

Instead, gradient-echo (GRE) based sequences are generally used to generate the T1-maps for patient-based imaging due to shorter acquisition times [7, 73]. This includes using the variable flip angle (VFA) method [74], which uses several short spoiled GRE (SPGR) sequences sensitive to T1-times, whereby the FA's (α_n) are changed between repetitions [74–77].

Most FA's utilized in VFA sequences' are less than 60°. Due to this smaller FA, some longitudinal magnetization still exists between repetitions, and hence the TR required to wait for the original magnetization to return to its original amount is reduced for following pulses. This allows for the desired shorter image acquisition time [31, 75, 76]. Another benefit to using an SPGR sequence is that it obtains the images using a similar acquisition (imaging type and parameters) to the dynamic sequence. This allows for cross-image registration and calibration errors to be mitigated [68, 76].

The dual flip angle (DFA) method involves using only two FA's. The reduction in the number of FAs shortens the time of acquisition even further and simplifies image post-processing [75, 78]. However, there is often a compromise between reduced VFA acquisition time and the accuracy of the T1-time estimation, which can have consequential effects on the output PK-parameter maps [68, 72].

When using the VFA method, sometimes the prescribed and applied FA magnitude can differ due to B_1-inhomogeneities [79, 80]. This can cause errors in the T1-time calculation, and have significant impacts on the PK parameters derived [67–69]. Further, this effect increases with higher strength B_0 fields [79]. Thus, B_1-maps are often acquired during the same imaging session as the VFA acquisition, and are used in post-processing to correct the T1-times calculated [80].

One Siemens vendor-specific T1-mapping sequence is a volumetric interpolated breath-hold examination (VIBE), which was developed in 1999 by Rofsky et al. (1999) [81]. VIBE uses fast 3D GRE sequences to generate high-quality T1-weighted and 3D-volumetric images of the patient's anatomy in a "breath-hold" (less than 30-seconds) time-frame to minimise respiratory motion effects [82].

Another variant of this sequence is called StarVIBE. It is commonly used to compensate for forms of involuntary motion in the body and reduce artefacts [83]. This allows patients to continue free-breathing during imaging, which is beneficial for patients who may not be able to hold their breaths for extended periods of time (E.g., for lung cancer patients). StarVIBE has been shown to robustly be able to produce sufficient anatomical and functional images of structures in the lung [83, 84].

2.3.5 Clinical Potential of DCE-MRI in Radiotherapy

DCE-MRI is often used clinically in cancer therapies for qualitative analysis (e.g. aiding in visualizing tumour location) [12]. The technique's non-invasive and quantitative ability to assess tumour blood flow and vessel permeability has shown promise for it to be used in tissue characterisation and treatment response assessment. This includes for tumours located in the breast [15,52,85,86], brain [87], head and neck [88,89], and prostate [12,38].

This technique has seen the potential to be particularly useful in breast cancer diagnosis; a

disease where often traditional diagnostic techniques, such as mammography, especially can fail for women with dense breasts [85]. Several studies have reported on its use to improve the diagnostic ability to characterize breast lesions [85, 86].

One such study composed of 124 patients was completed by Cheng et al. (2018). It aimed to use DCE-MRI to help differentiate between benign and malignant lesions, and the results were very promising [85]. The study showed that the quantitative DCE parameters derived alone had high diagnostic accuracy (specificity) of 94.9%. Further, several of the PK parameters (inc. K_{trans} and v_p) were significantly higher for lesions which were malignant than those classified as benign.

Other studies highlight the potential of DCE-MRI in predicting treatment response for cervical cancer patients [90]. One study conducted by Zahra et al. (2008), found that both semi-quantitative and quantitative PK-parameters significantly aligned with the tumour regression trends which were observed over the course of treatment.

Anatomy-specific technical limitations have been reported to exist in the literature; inhibiting conclusive correlations between PK-parameters and diagnosis and/or treatment outcomes to be made. This includes tumours in the head and neck region due to the large susceptibilities and tumour motion that can be present [72, 88]. Similarly, tumours located in the cervical and lung regions often require increased temporal resolution for the dynamic sequence due to rapid CA uptake; this presents a compromise to spatial resolution, sometimes at the expense of adequate PK-parameter assessment [90, 91].

There are however a number of broad factors which can impact the output PK parameters accuracy and precision, which are ultimately limiting DCE-MRI's clinical implementation [15]. This includes the noted inaccuracies that can be present in the pre-contrast T1-map, image artefacts, and lack of standardization in used image acquisition parameters and image processing/analysis techniques [15, 89, 90, 92, 93].

For the technique to become translated clinically, imaging standardization is required. This is along with the development of internationally recognized T1-mapping QA methods, in order to validate the accuracy and repeatability of the T1-maps subsequent PK-parameters. This will be further addressed in section 2.4.

2.4 MRI Quality Assurance (QA)

QA refers to the process of ensuring that a product (e.g. an MRI scanner) is producing an output (e.g. accurate ADC) of its intended quality. It does this by assessing if the process is operating adequately in generating the output [31, 94]. There exist two types of QA. The first includes QA services completed by a vendor to ensure the machine is performing within the manufacturer's specifications (without user interaction). The second is that carried out by the user to quantify any other machine performance characteristics [31].

MRI QA often involves manual and department-specific processes. This is due to there being only very few national/international MRI QA guidelines available when compared to other imaging modalities which use ionizing radiation [95]. There are some QA guidelines available for departments to use and test specific standard MRI performance characteristics. For example, the National Electrical Manufacturers Association (NEMA) has published specifications and procedures to verify a scanner's SNR ratio [96] and slice thickness [97] performance.

For MRI to be used in radiotherapy however, images need to have higher spatial accuracy and image reproducibility compared to diagnostic MRI images in order to achieve accurate image registration and tissue delineation. Recently, the American Association of Physicists in Medicine (AAPM) Task Group 284 Report was published. As an overview, it outlined the additional MRI QA tasks needed for the effective implementation of MRI into radiotherapy [98]. Suggested testing frequencies (e.g., monthly or daily) were noted for certain QA tests, along with methods needed to carry out the tests such as characteris-

ing B_0 and B_1 inhomogeneities [98].

At current, there are no qMRI performance characteristics listed in an MRI scanner's vendor specification list. Further, because of the novelty of qMRI compared to standard MRI techniques, there are even fewer resources and literature available for departments to use to develop and complete their own qMRI QA. It is for these reasons that the known level of reliability for the quantitative values derived from scanners is limited, and is thus restricting the clinical translation of the techniques [3, 7, 19, 37].

Ideally, qMRI assessments would involve measuring the technical performance characteristics of each QIB, including its accuracy, repeatability, and reproducibility [7, 17]. A definition of terms such as these, as per adopted from Sullivan et al. (2015) [99], will serve to be standard references for the remainder of this thesis:

- Reference Value: A value used as a basis for comparison with values of quantities of the same kind; derived through utilizing reference methods.
- Linearity: The ability to provide measured values directly proportional to the value of the measure (in the same experimental unit). .
- Precision: The closeness of agreement between measured quantity values obtained by replicate measurements of the same (or similar) experimental units.
- Repeatability: The precision; given that the conditions remain unchanged between replicate measurements (e.g., same measurement procedure, operator and operating conditions, and physical location).
- Reproducibility: The precision; given that the conditions vary between replicate measurements.
- Bias: An estimate of a systematic measurement error; often expressed as a percentage difference when comparing the measured value to a reference value.
- Accuracy: Does not have a single unambiguous definition. It is often used interchangeably with bias (as per this thesis) and is used in a way that combines both bias and precision.

Determining such technical performance characteristics becomes especially important

when the QIB is planned for use in cancer therapy, and a patient's treatment plan is to be potentially adapted based on the QIB value measured. There are developments currently occurring in this space, including international task groups developing qMRI QA profiles and dedicated phantoms to assess QIB reliability on a scanner (as discussed in sections 2.4.1 and 2.4.2 below).

2.4.1 qMRI QA Profiles

Although not yet recognised internationally as a standard for qMRI QA, there exist lists of recommended qMRI imaging protocols, testing samples, and analysis methods for performing the QA [7, 100]. The Quantitative Imaging Biomarkers Alliance (QIBA), a task group created by the Radiological Society of North America (RSNA), has spent significant time developing DWI- and DCE-MRI QA 'Profiles' [20, 101]; incorporating tolerances and QA recommendations from over 1000 publications in the literature [7].

These Profiles were released to aid in quantifying the reliability of DWI- and DCE-MRI derived QIBs on MRI scanners to increase the QIBs clinical utility in single- and multi-site trials. All Profiles, analysis software, and computer-generated DICOM datasets known as digital reference objects (DROs) QIBA created are publicly available [7, 20, 101, 102].

The Profiles and literature suggest that standardized imaging sequences, protocols, phantoms for QA, and image analysis methods are key for translating qMRI QIBs into clinical trials, and then clinical practice [7, 20]. A similar agreement was found in an overview of Imaging Biomarkers in Oncology by Dregely et al. (2018) [2]. The following sections will focus on MRI phantoms available for MRI QA purposes.

2.4.2 MRI QA Phantoms

A phantom refers to any object that acts to mimic human tissue for applications in medical imaging. They are generally designed with known properties and characteristics that have been standardized using a source external to the medical imaging system (e.g. Nu-

clear Magnetic Resonance (NMR) methods vs an MRI scanner). They have less variation within their samples and allow for the removal of many of the variable properties and interactions that occur naturally in human tissue. For these reasons, phantoms are largely used for measuring the functional quality of medical imaging systems. However, due to often oversimplifying the complexity of human tissue and having incomparably large SNRs, in vivo imaging should also be used for QA validation [18].

There are several phantoms and reference objects designed for user-based clinical MRI QA purposes [100, 103–105]. Clinical scanner properties like SNR, geometric accuracy, spatial resolution, slice thickness and position, image uniformity, ghosting artefacts, and central frequency can be assessed using phantoms like the American collage of Radiology (ACR) MRI phantom [103], Eurospin Test Objects [104], or in part, the Alzheimer's Disease Neuroimaging Initiative (ADNI) MagPhan phantom [105].

All phantoms have specific usability limitations. For example, the Eurospin's complexity with multiple components and higher cost reduces its ability to be clinically implemented [103]. Also, the ACR phantom is the most widely used clinical phantom in the world, however, it's known to degenerate over time and is limited in its set-up positioning for performance assessments [100].

There are now specific qMRI phantoms and reference objects available; able to measure both clinical and QIB related MRI scanner technical performance [7, 100]. A summary by Keenan et al. (2018) describes the phantom characteristics required for such assessments and provides an overview of several phantoms currently available to complete this testing [100]. qMRI QA using these phantoms can thus be used to ensure the changes measured in patient QIBs are due to a physiological response, and not measurement variability caused by scanner performance fluctuations.

2.4.3 ISMRM/NIST qMRI Phantoms

Two phantoms recommended for qMRI QA are the National Institute of Standards and Technology (NIST)/International Society of Magnetic Resonance in Medicine (ISMRM) System phantom, and the NIST/RSNA/ National Cancer Institute (NCI) Isotropic Diffusion phantom [100].

ISMRM/NIST System Phantom

The System phantom was designed by NIST in collaboration with the ISMRM, and is currently commercially manufactured by CaliberMRI (Colorado, United States). Its features include similarities to the ADNI (geometric distortion array) and ACR (resolution and slice profile additions) phantoms for clinical MRI QA applications. Additionally, this phantom enables the measurement of some qMRI QIBs like T1, T2 (similar to EuroSpin) and proton density (PD) [73, 100, 106]. The phantom is SI-traceable, low cost, has stable vial solutions (for up to 4 years), and allows for simple and reproducible setups on MRI scanners [73, 106].

In particular, the phantom has a spherical geometry with a 200 mm inner diameter (ID) shell. This design was selected to mimic the approximate dimensions of a human head and thus fit into most MRI receiver head coils [106]. The phantom design can be observed in Figure 2.9 and includes five polyphenylene sulfide (PPS) plates (for low water absorption and thermal expansion), which are connected together internally with PPS rod structures. Anteriorly to posteriorly, the phantom comprises of [106]:

- Plate 5: 5 fiducial spheres (grey vials) and a T1 array (green vials)
- Plate 4: 13 fiducial spheres, resolution insert and a T2 array (red vials)
- Plate 3: 26 fiducial sphere and proton density array (yellow vials)
- Plate 2: 13 fiducial spheres and a slice profile (slice thickness insert)
- Plate 1: 5 fiducial spheres

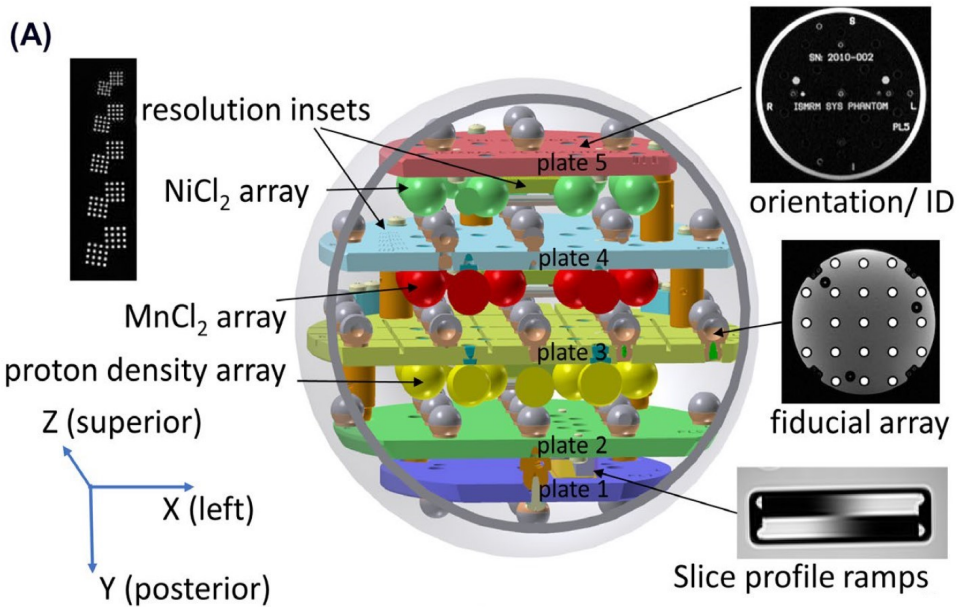

Figure 2.9: ISMRM/NIST System phantom schematic, taken from Stupic et al. (2021) [106]. Observe there are five different plates embedded to facilitate the various tests able to be performed using the phantom. These tests include geometric accuracy (grey fiducials), slice thickness, spatial resolution, T1 (green vials), T2 (red vials) and proton density (yellow vials) measurements.

The System phantom fulfils many of the phantom property requirements listed in the recommendations by Keenan et al. (2018) [100]. Further details on this phantom can be found in a review on the phantom by Stupic et al. (2021) [106], and selected prior studies which have used this phantom are outlined within chapters 3 and 5.

NIST/RSNA/NCI Isotropic Diffusion Phantom

NIST, NCI, and the RSNA collaboratively developed a commercially available isotropic diffusion phantom, with 13-vials embedded with well-characterised reference ADC values [107, 108]. This phantom provides the ability to assess a scanner's technical performance over a wide range of physiologically relevant ADC values, which could affect ADC reliability [107–109].

The Diffusion phantom is also manufactured by CaliberMRI, and its design has been extensively described in the literature (see chapters 4 and 5) [7, 100, 107, 109]. As an overview, it has an exterior spherical (194 mm diameter) plastic shell to again mimic the shape of a human head [108]. To provide temperature stability, it has two opening points,

located anteriorly and posteriorly, that can be opened to fill the phantom with an ice bath.

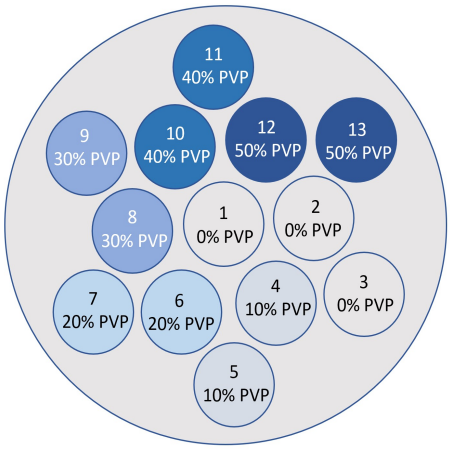

Figure 2.10: Vial arrangement within the Isotropic Diffusion phantom. PVP concentrations are shown (by mass fraction (%)) for inner- and outer-ring vials. Centralized vial 1 (and vials 2-3) contain distilled water.

Within the phantom, there are 13 30mL plastic vials that encase aqueous solutions with known reference ADC values. Figure 2.10 shows the phantom's inner- and outer-rings of vials surrounding the central distilled-water vial (positioned at isocenter). The surrounding vials contain a range of concentrations (by mass fraction) of Polyvinylpyrrolidone (PVP) in an aqueous solution including (%): 0, 10, 20, 30, 40, and 50. Reference ADC values in each of the vials cover a wide range of physiologically relevant ADC values [108].

2.5 Summary

Over the years, there have been many advances in radiotherapy: many of which have led to improved post-treatment patient outcomes [1]. The use of both anatomical and physiological imaging could further improve these outcomes by assisting ROs in ART-based decision-making by providing enhanced treatment response assessment tools [2, 3, 16, 21, 24, 25]. In particular, QIBs derived using qMRI techniques have been shown to have the potential to aid in this decision-making. For example, some qMRI QIBs have been shown to help assess treatment response for a wide-range range of anatomical sites [2, 3, 11–16].

A key limitation associated with the past investigations that report on the potential of qMRI QIBs includes the unknown accuracy and repeatability of the QIB values being derived using clinical MRI scanners [7, 37]. To overcome this limitation, qMRI QA is needed to be performed on scanners to assess the inherent technical performance of each QIB [7, 18]. This includes for MRI scanners being used in both single- and multi-site clinical trials investigating QIBs. Only in recent years have developments occurred in the space of qMRI QA; including the evolution of dedicated qMRI phantoms and Profiles (see sections 2.4.3 and 2.4.1).

Incorporating such QA will allow for the accuracy and repeatability of a QIB to be known, and assist in overcoming some of the technical challenges of qMRI in single- and multi-site clinical trials [7, 21, 64]. It is only then can the MRI-derived QIBs be considered for clinical implementation and ultimately help improve treatment outcomes for patients [7].

The key focus of this thesis was to use the newly available resources to develop efficient methodologies for assessing the accuracy and repeatability of qMRI biomarkers on MRI scanners. These methods were then used to monitor a 3T radiotherapy dedicated MRI scanner's performance over short- (daily) and long-term (monthly) periods in order to provide evidence for the need of this type of QA. Overall, this was aimed to assist other departments in designing and implementing their own qMRI QA protocols.

Chapter 3

Determining the Accuracy and Repeatability of T1 and T2 in a 3T MRI Scanner

The results presented in this chapter have been published in the Journal of Applied Clinical Medical Physics:

M. Carr, K. Keenan, A. Walker, R. Rai, P. Metcalfe, L. Holloway. *Determining the Longitudinal Accuracy and Reproducibility of T1 and T2 in a 3T MRI Scanner*. Journal of Applied Clinical Medical Physics, vol. 22, no. 11, p. 143-150, 2021. DOI: 10.1002/acm2.13432

Contributions: M. Carr performed the measurements, analyzed the data, and wrote the manuscript with support from all the authors. K. Keenan, R. Rai, P. Metcalfe, A. Walker, and L. Holloway were involved in planning experiments and supervising the work.

3.1 Introduction

Recall from section 2.2.1 that T1- and T2-relaxation times become properties of tissues when the body is exposed to an external magnetic field, such as inside an MRI scan-

ner. These parameters are a function of the underlying quantum mechanical properties of the tissues in a magnetic field and can be measured using several different approaches [19, 40, 110].

T1-time is important in DCE-MRI for quantifying blood perfusion (see section 2.3.4), and T2-time has been used in applications such as disease diagnosis [25, 31] and detecting cartilage degeneration [111]. Quantifying the changes in T1- and T2-relaxation times, over time, has seen the potential to aid in monitoring patient treatment responses to a variety of cancer therapies [16, 24, 25, 31].

To translate T1- and T2-time into clinically utilized QIBs, it is essential they are accurate, repeatable, reproducible and independent of scanner performance [7, 37]. As noted in section 2.4.2, phantom-based QA programs can assist in determining the QIBs technical performance on a specific scanner. The ISMRM/NIST System phantom is commercially available for this purpose; with embedded vial arrays having well characterized T1- and T2-reference times (see Figure 2.9).

In past longitudinal studies, this phantom was scanned repeatedly (up to 100 days) to monitor changes in T1 and T2 over time using MRF [40, 110]. Variability in T1-time derived from traditional spin echo (SE) methods has also been assessed using the phantom, with single centre results finding that accuracy and repeatability of T1-time varied pre- and post- a scanner upgrade [112].

Furthermore, a multi-site study found performance was dependent on sample T1-relaxation time, magnetic field strength and sequence choice [19]. Previous longitudinal studies assessing T1- and T2-time accuracy and repeatability acquired measurements at infrequent and/or over short time periods [19, 112, 113], or assessed relaxation times relevant to specific anatomy at different magnetic field strengths [114].

To address these noted shortfalls in the literature, the work presented in this chapter investigates T1- and T2-relaxation time accuracy and repeatability by uniquely imaging the System phantom at high-frequency intervals. This included monthly image acquisitions over one year, which were used to determine the scanners' baseline performance metrics.

Using the System phantom also allowed for this assessment to be performed over a wide-range of relaxation times instead of being limited to single relaxation time samples. Quantifying the longitudinal changes in parameters such as T1- and T2-relaxation times is essential for advancing the use of qMRI clinically (e.g., in long-term treatment response monitoring) [7].

3.2 Methods

3.2.1 Data Acquisition

A System phantom was scanned on a monthly basis over the course of a year, ensuring each scan was separated by a minimum interval of two weeks, using a 3T MRI scanner (Siemens Healthineers, MAGNETOM Skyra, Erlangen, Germany). Imaging was completed using a 20 channel Head/Neck coil (Figure 3.1). T1-weighted inversion recovery (T1-IR) and T1-VFA sequences were utilized for T1-mapping (see section 2.3.4), whilst a multi-echo spin echo (T2-SE) sequence was used for T2-mapping.

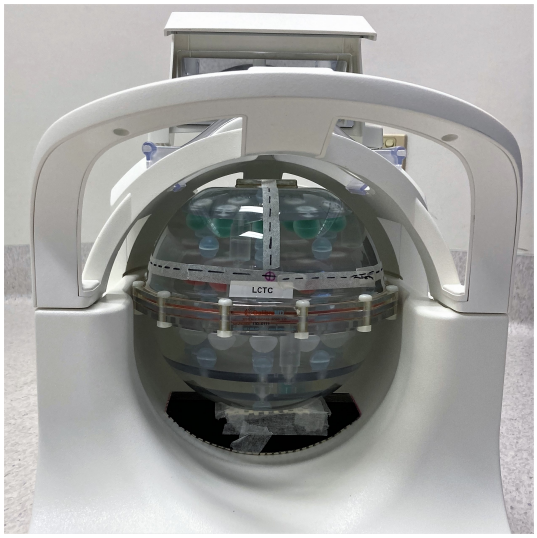

Figure 3.1: The System phantom on top of a 3D-printed holder, fitted to the Head/Neck coil. 3D-orthogonal markings were drawn to assist with external laser alignment.

Sequence and parameter selections (outlined in Table 3.1), were based on recommended protocols in the phantom manual [115]. All phantom setups, image acquisitions and analyses were completed by one user (physicist with 3 years of MRI experience).

Table 3.1: Acquisition parameters utilized for the three sequences tested. This includes T1-inversion recovery (T1-IR), T1-variable flip angle (T1-VFA) and multi-echo spin echo (T2-SE). Parameters encompassed: FA = flip angle, TE = echo time, TR = repetition time, TI = time of inversion, FOV = field of view, Acq. time = acquisition time, FE/PE/SE = frequency/phase/slice encoding respectively.

Parameter	**T1-IR**	**T1-VFA**	**T2-SE**
Sequence	2D/TSE-IR	3D/SGRE	2D/SE-ME
Orientation	Coronal	Coronal	Coronal
FA (degrees)	N/A	2, 5, 10, 20, 25, 30	N/A
TE (ms)	6.9	2.44	10 to 320 by 10 ms intervals
TR (ms)	4500	6.6	5000
TI (ms)	35, 75, 100, 125, 150, 250, 1000, 1500, 2000, 3000	N/A	N/A
FOV (mm^2) (FE x PE)	250 x 250	250 x 250	250 x 250
# Slices	1	32	1
Slice Thickness (mm)	6	6	6
Matrix (FE/PE/SE)	256/192/1	256/192/32	256/192/1
# of averages	1	1	1
Acq. Time (min)	40	8	16

3.2.2 ISMRM/NIST System Phantom

The T1 and T2 sequences imaged the T1- and T2-arrays (as described in section 2.4.3) embedded within the System phantom (serial: 130-0111: CaliberMRI, Colorado, United States), respectively [106]. NMR IR methods have been used by NIST in the past to characterize these solutions under 3T and 20°C conditions, and the reference values are provided in Table 3.3 [116, 117]. Note that vials 1 and 5 from the T2-array were not included in the analysis at the recommendation of the manufacturer.

The phantoms' arrays at 20°C covered a large range of relaxation times, including those

found in the human body at 3T: T1 = 121 ms to 1884 ms and T2 = 30 ms to 79 ms [71, 106, 117, 118]. The physiologic range of relaxation times was of particular interest, and thus results were separated into two categories: full vial range and, a subset, human vial (physiologic) range. Temperature of the surrounding deionized water in the phantom was measured both before and after each scanning session using a NIST-traceable thermometer (Traceable® Extreme Accuracy Thermometer, 1227U09, Thomas Scientific, Swedesboro, U.S.A).

3.2.3 Image Analysis

All image processing was completed using a consistent software platform and analysis method. ImageJ (National Institutes of Health, Maryland, USA) was initially used to manually identify the centre pixel locations of each array vial on the shortest inversion time (TI) (35 ms) image of the T1-IR dataset. These locations along with all datasets were imported into an in-house developed Python script (see Appendix A). This automatically positioned a 10-pixel (approx. 9.8 mm) diameter circular region of interest (ROI) to be at the centre of each vial and on the central slice of the respective T1 and T2 array.

The average signal (S) for each vial ROI was calculated and fitted to corresponding signal Equations 3.1, 3.2, or, 3.3, depending on the sequence: T1-IR, T1-VFA or T2-SE, respectively. This process included using a non-linear least-squares Levenberg-Marquardt algorithm to calculate the respective relaxation times, as implemented by LMFIT in Python [19,31,110,112,119]. Specifically, this was a mono-exponential fit (3-parameters) for the T2-SE signal whereby the first echo was omitted due to expected and observed inconsistent phase coherences [120, 121]. For T1-IR and T1-VFA, this used a 4- and 2-parameter fit model, respectively (see Table 3.2).

$$T1-IR: S(TI) = |M_0(1+(invF-1)e^{\frac{-TR}{T_1}} - invF \bullet e^{\frac{-TI}{T_1}})| + n \tag{3.1}$$

$$T1-VFA: S(\alpha) = \frac{\sin(\alpha) \bullet M_0 \left[1 - e^{\frac{-TR}{T_1}}\right]}{1 - cos(\alpha) \bullet e^{\frac{-TR}{T1}}} \tag{3.2}$$

$$T2-ME: S(TE) = M_0 e^{\frac{-TE}{T2}} + n \tag{3.3}$$

In equations 3.1, 3.2, or, 3.3, α denoted the nominal FA used in the VFA acquisition, TR was the repetition time, TI was the time of the inversion pulse and TE was the echo time. These parameters were read from the DICOM header in the imported images using Python. Further, $invF = (1 - \cos(\emptyset))$ was the inversion factor where $\emptyset$ is the angle of inversion, M_0 was the equilibrium magnetization and n was the noise factor.

Table 3.2: Initial values and bounds of parameters implemented in the signal fitting equations for each of the three sequences. Note that TI_{null} was estimated to be the TI which had the minimum signal (S(TI)) value registered to the T1-IR data.

Sequence	Parameter	Initial Value	Min	Max
T_1-IR	M_0	max(S(TI))	0	-
	InvF	2	1	2
	n	0	0	-
	T1 (ms)	TI_{null}/ln(2)	0	5000
T1-VFA	M_0	max(S(α))	0	-
	T1 (ms)	800	10	3000
T2-SE	M_0	max(S(TE))	0	-
	n_0	0	-	-
	T2 (ms)	median(S(TE))	0	2000

These three parameters were all unknown and required derivation using the non-linear least-squares Levenberg-Marquardt algorithm, with the initial parameters and bounds given in Table 3.2. These were based on the expected range of minimum and maximum T1 and T2 times in the phantom for each sequence, as well as to closely align with the initial parameters reported in the literature [19].

Note for T1-VFA signal fitting, vials 1-8 were fit with all six FA's in the image sequence. Vials 9-14 however, due to the large amount of signal saturation detected for FAs 20-30^o, were only fit with FAs: 2^o , 5^o and 10^o. This was a similar filtering method to that employed in the literature when similar saturation effects were observed [112].

To assess the accuracy error of the measured T1- and T2- times, the %bias was calculated (as per Equation 3.4) for each monthly acquisition and each vial using a comparison to the NIST measured reference vial value.

$$\%bias = \left(\frac{Measured\ T_1\ -\ NIST_{(NMR)}T_1}{NIST_{(NMR)}T_1} \right) \times 100 \tag{3.4}$$

To assess the repeatability over the range of reference T1- and T2-times in the phantom, a coefficient of variation (%CV) was calculated as per stipulated in Equation 3.5. This was using the individual vials' mean (μ) and standard deviation (σ), calculated over the 12 monthly repetitions.

$$\%CV\ = \frac{\sigma}{\mu} \times 100 \tag{3.5}$$

3.3 Results

The phantoms' T1 and T2 arrays were imaged monthly using the respective T1-IR/T1-VFA and T2-SE imaging sequences described in Table 3.1. The average interval between imaging sessions was 4 weeks. The mean T1- and T2-times derived for each vial and their standard deviations (SD) are listed in Table 3.3, and were calculated using all months' measurements. Table 3.3 also includes the NIST reference vial values, as reported by the manufacturer [116, 117].

3.3.1 Full Vial Range

A wide range of accuracy errors in T1 and T2 measurements existed over the full vial range. Visualization of this variability can be seen in Figure 3.2. T1-IR had the smallest bias when all vials were included (median= +3.6%), compared to T1-VFA (+5.0%) and T2-SE (+5.8%) sequences.

In terms of repeatability for T1-IR, T1-VFA and T2-SE sequences, respective CV's over the full vial range were 2.5%, 4.3% and 2.2%. The largest CV's and accuracy deviations were found to occur in the shortest reference time vials, as seen in Figure 3.3. For example, the T1-IR bias in vials 13 and 14 was respectively +21% and +71%, and -14% and -35% for T2-SE. Similarly, T1-VFA measurements in vials 10-14 had biases in the range of -17% to +14%.

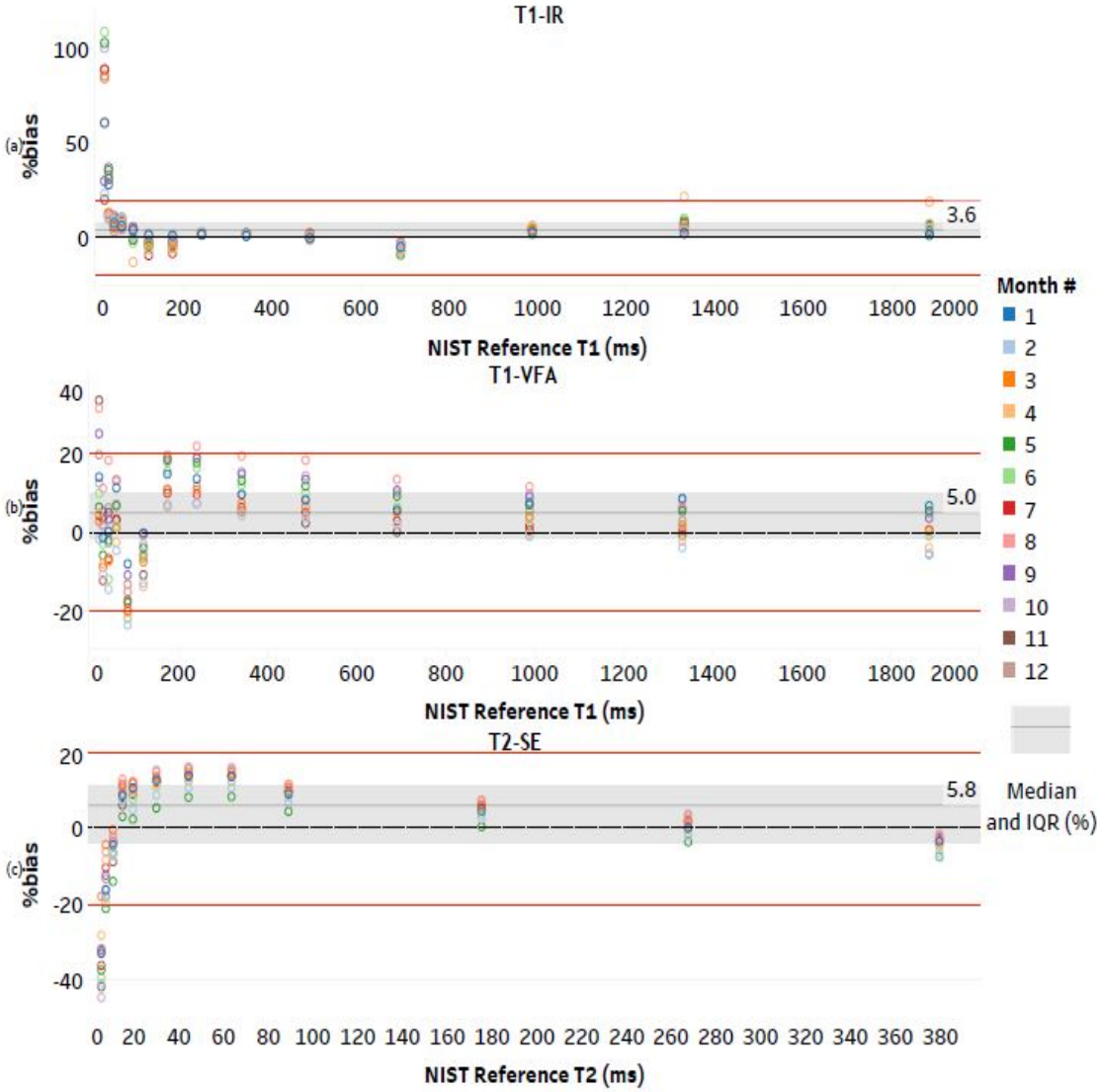

Figure 3.2: Bland-Altman plots for (a) T1-IR, (b) T1-VFA and (c) T2-SE (bottom), show the difference (%) between measured and reference T1 and T2 times for vials in the full vial range. Median bias (and lower – upper quartiles) are displayed and include: +3.6 % (+0.8 – +7.7), +5.0% (-1.46 – +9.96) and +5.8% (-3.8 – +11.1) for T1-IR, T1-VFA and T2-SE respectively.

Table 3.3: T1- and T2-times both measured by NIST at 20°C and experimentally derived using T1-IR, T1-VFA or T2-SE methods. The NIST times were obtained from the phantom manual [117], and experimental times have been presented as an average of the 12 monthly measurements (with respective SD's). Vials 1-14 refer to the full vial range, whilst only vials 1-9 refer to the human vial range (left).

Vial Number:		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
T1 (ms) NIST	Value	1884.0	1330.2	987.3	690.1	485.0	341.6	240.9	175.0	121.1	85.8	60.2	42.9	30.4	21.4
	SD	30.3	20.4	14.2	10.1	7.1	5.0	3.5	2.5	1.8	1.2	0.9	0.4	0.6	0.3
T1-IR (ms)	Value	2007.2	1430.7	1032.9	655.0	489.4	348.3	246.8	170.2	117.9	86.7	64.9	46.4	36.7	36.8
	SD	77.8	63.7	12.0	13.9	5.5	1.2	1.0	4.7	3.9	4.2	1.1	0.9	3.3	6.5
T1-VFA (ms)	Value	1894.1	1370.2	1036.0	732.7	528.3	376.6	273.5	198.1	113.9	71.1	63.3	42.2	29.7	24.4
	SD	74.7	52.4	39.8	27.1	22.7	15.8	11.9	8.4	5.4	3.9	3.4	3.7	2.1	2.4
T2 (ms) NIST	Value	*	379.5	267.3	175.1	*	88.9	63.4	44.2	29.9	19.4	14.7	10.5	7.3	5.3
	SD	*	3.6	2.5	1.7	*	0.8	0.6	0.4	0.3	0.2	0.1	0.1	0.1	0.1
T2-SE (ms)	Value	*	364.5	269.1	183.9	*	97.1	71.9	50.2	33.4	21.2	16.0	10.0	6.3	3.5
	SD	*	7.4	5.6	3.5	*	1.8	1.3	1.0	0.8	0.6	0.4	0.4	0.4	0.4

*Vials removed from analysis at the recommendation of the manufacturer.

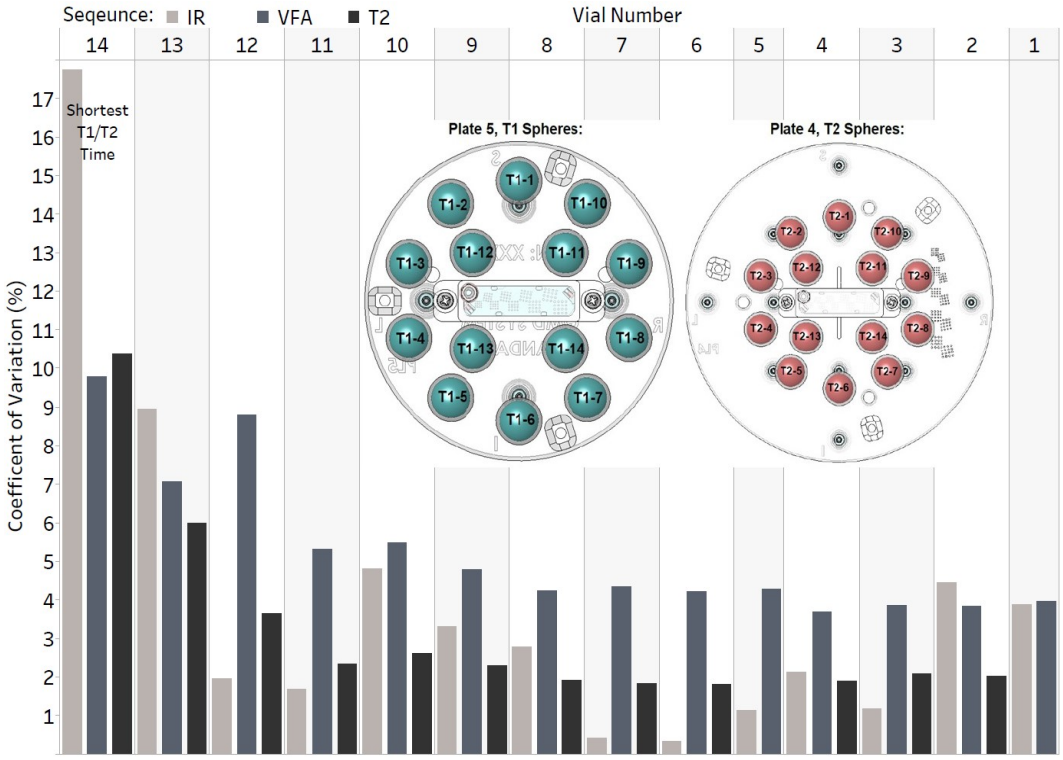

Figure 3.3: Coefficient of Variation (CV) calculated for each vial and each sequence from all monthly acquisitions. This is where vial 14 corresponded to the shortest vial reference T1- and T2-times, while vial 1 had the longest reference time (vial diagrams, not to scale, provided from the System phantom manual [115]).

3.3.2 Human Vial Range

The human vial range omitted results from the vials with the shortest reference times. Consequentially, the median bias fluctuated between approximately -20% and +20% (Figure 3.4), and IQR's and CV's were reduced compared to the full vial range (Table 3.4).

Table 3.4: Summary of the baseline accuracy (bias) and repeatability (CV) measured for the 3 sequences investigated at 3T: Full phantom vial range (top) and human vial range (bottom). Values are stated as the median of all months/vials' acquisitions and their respective inter-quartile ranges.

	(%)	**T1-IR**	**T1-VFA**	**T2-SE**
Full Vial Range:	bias	+3.6 (6.9)	+5.0 (11.4)	+5.8 (14.9)
T1: 1884-21 ms &	CV	2.5 (3.0)	4.3 (1.4)	2.2 (0.9)
T2: 380-5 ms				
Human Vial Range:	bias	+2.0 (6.11)	+6.5 (10.0)	+8.5 (10.7)
T1: 1884-121 ms &	CV	2.1 (2.2)	4.2 (0.4)	1.9 (0.2)
T2: 380-30 ms				

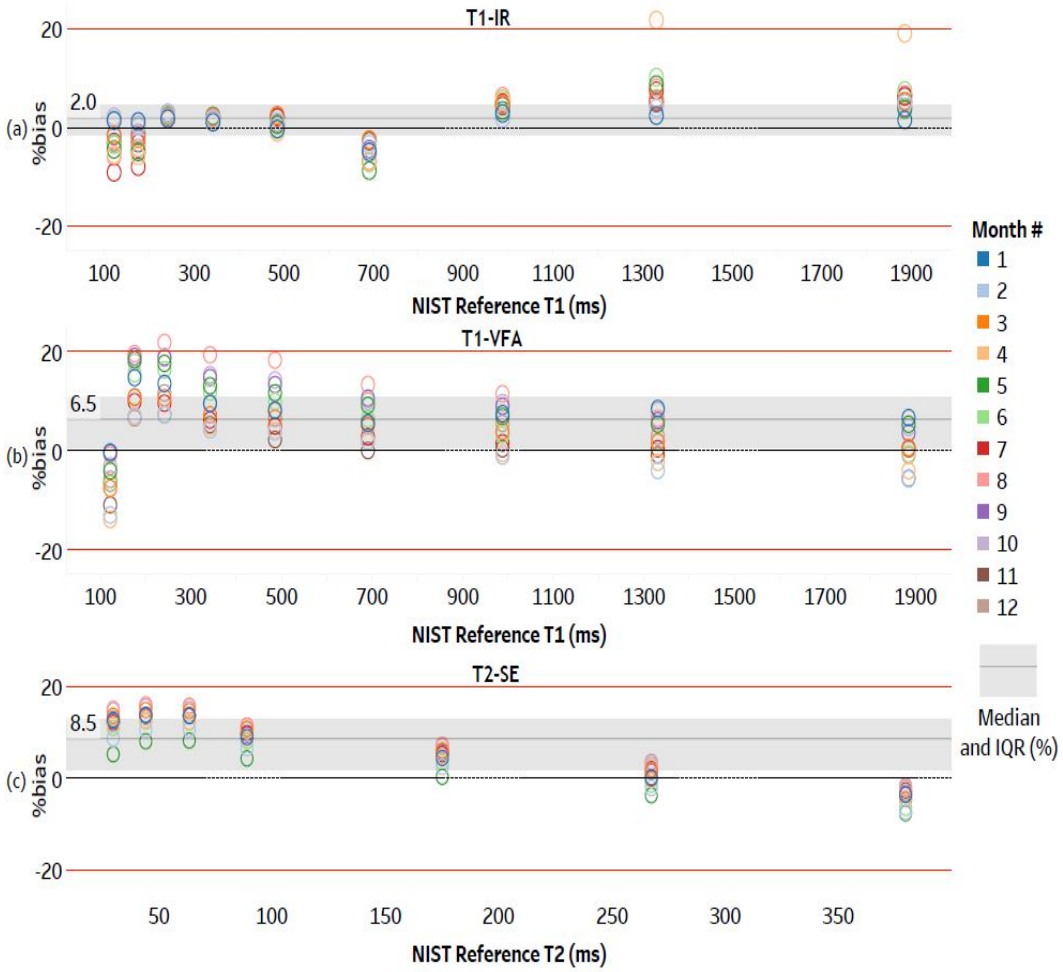

Figure 3.4: Bland-Altman plots for (a): T1-IR, (b): T1-VFA, and (c): T2-SE. The %difference can be observed between measured and reference T1 and T2 times for vials within the human range. Median biases (and lower – upper quartiles) are displayed and include: +2.0 % (-1.6 - +4.5), +6.5 % (+0.7 – +10.8) and +8.5 % (+2 - +12.7) for T1-IR, T1-VFA and T2-SE respectively.

3.3.3 Signal Fitting

The resulting fitted signals from month 1 data acquisitions are shown in Figures' 3.5, 3.6, and 3.7. Similar fits were observed over all months. Observations of the errors introduced by keeping the 1st echo in the T2 fitting can be seen in Figure 3.8. Further, Figure 3.9 shows the T1-VFA results for all months (over the full vial range) for the case that the filtering process recommended was not utilized (i.e., using 6 FAs for all vials).

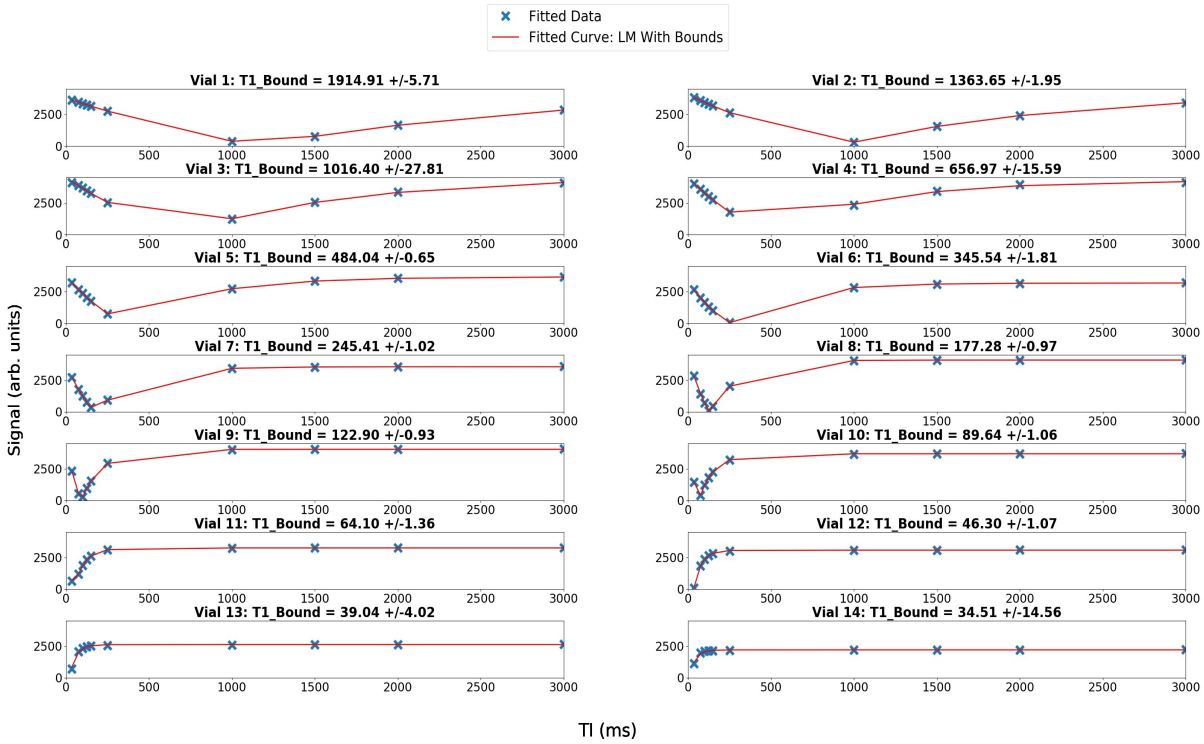

Figure 3.5: T1-IR fit produced for vials 1-14 for month 1. Values are given as T1 $\pm$ SD (of the fit). No signals approached saturation points (magnitude = 4095).

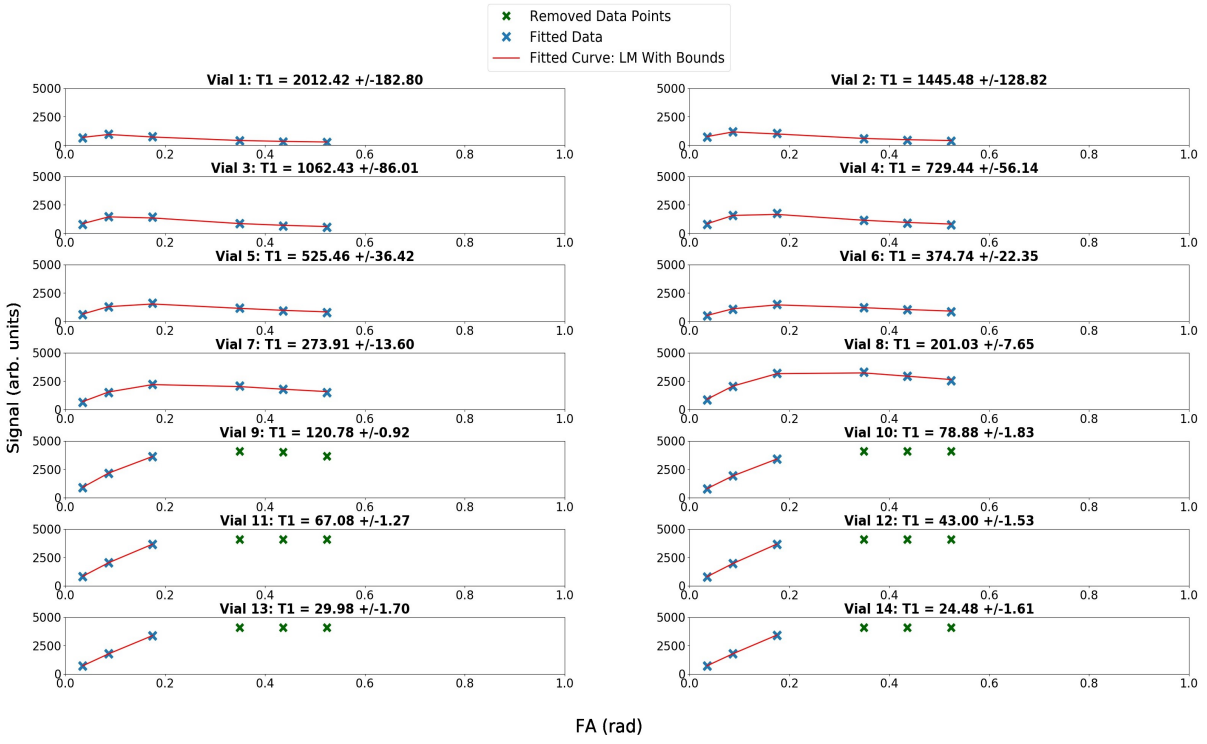

Figure 3.6: T1-VFA fit produced for vials 1-14 for month 1. Values are given as T1 $\pm$ SD (of the fit). Notice the green data points approaching saturation (magnitude = 4095) at shorter reference T1 vials (9-14) and for larger flip angles. These were removed from the fit as seen.

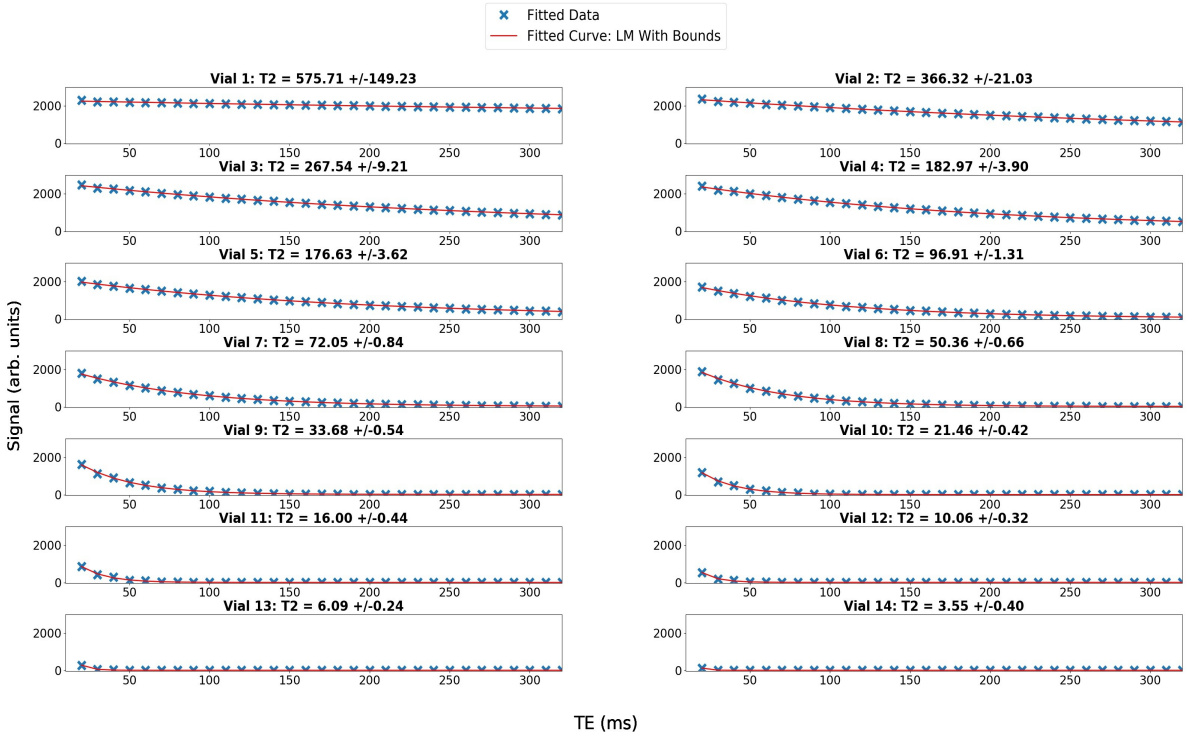

Figure 3.7: T2-SE fit produced for vials 1-14 for month 1, with the 1st echo removed. Values are given as T2 $\pm$ SD (of the fit). Notice that signals were approaching the noise floor, especially for shorter reference T2 vials.

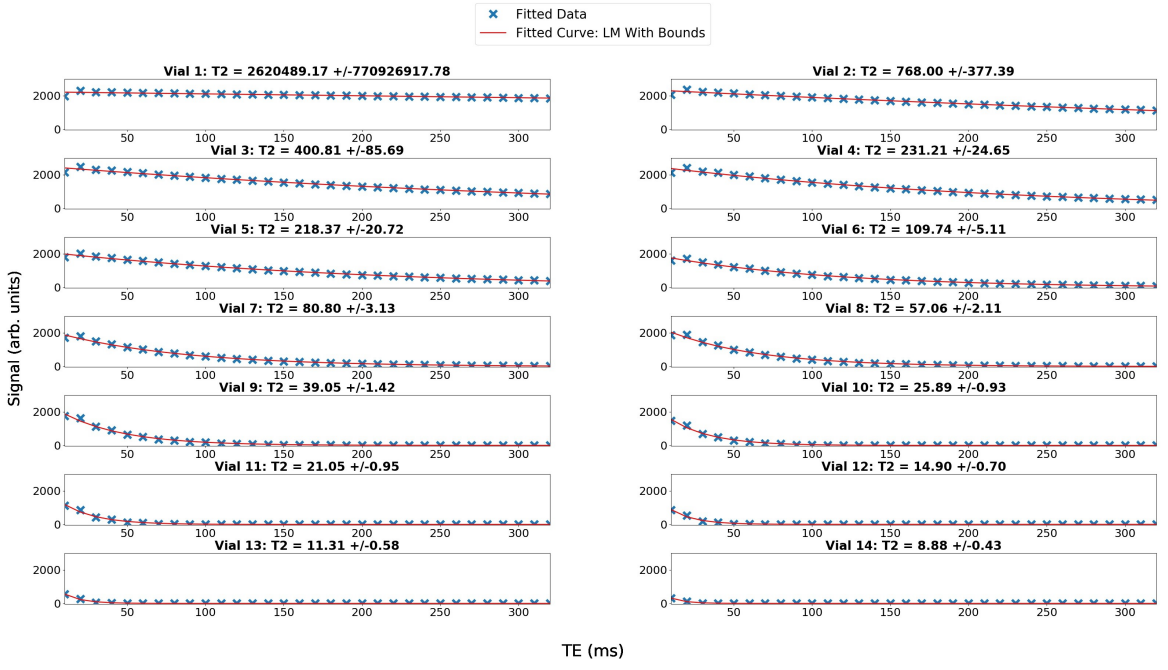

Figure 3.8: Original T2-SE fit produced for vials 1-14 for month 1 whilst preserving the 1st echo. Values are given as T2 $\pm$ SD (of the fit). Note the difficulty in producing the mono-exponential fit; large SD error were produced along with illogical overestimations of T2, especially for higher reference time vials (1-4).

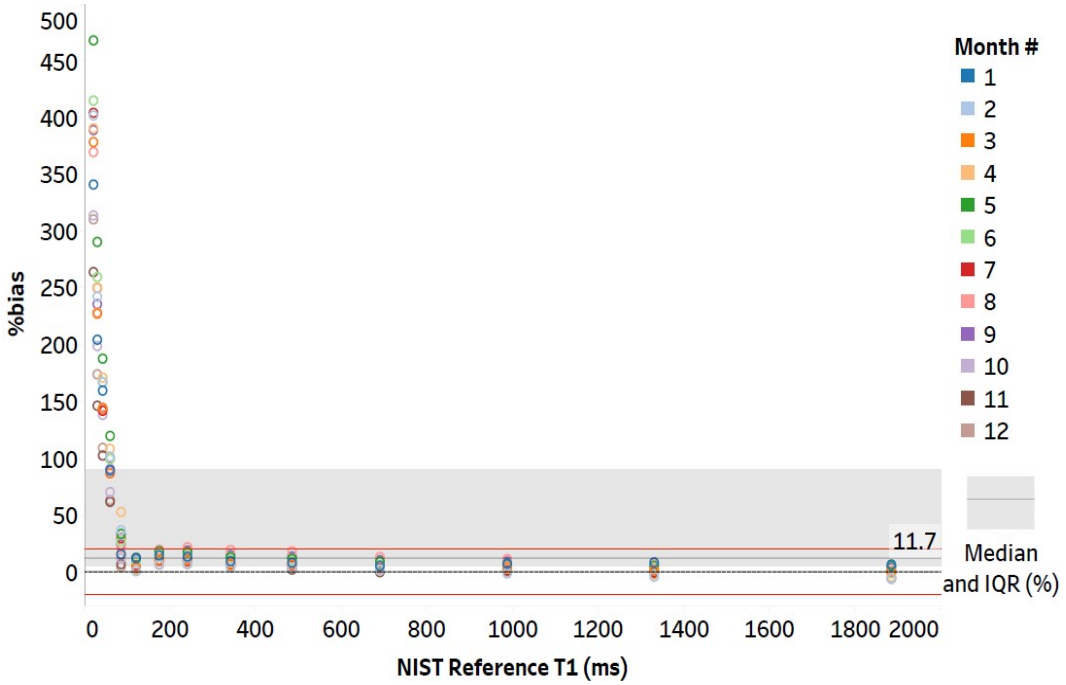

Figure 3.9: Non-filtered Bland-Altman plots for T1-VFA highlights the difference (%) between measured and reference T1 times for vials in the full range. Notice the large variability in %bias for shorter T1-times. Median biases (and lower – upper quartiles) are displayed and include: +11.7 % (+5.5 – +89.9).

3.3.4 Linearity, Time and Temperature Dependencies

There was a strong linear correlation between NIST reference times and all measured relaxation times. The coefficient of determination, R^2, was calculated by plotting the reference times against those measured (Figure 3.10). R^2 for T1-IR, T1-VFA and T2-SE was found to be 0.999, 0.999 and 0.998, respectively.

The monthly changes in the T2 measurements over the 12 months can be seen in Figure 3.11, along with recorded temperature fluctuations. Each month/vial is presented with its respective errors, generated from the SD of the fit. This was calculated using the square root of the diagonals of the covariance for the parameter. Similar results for T1-IR and T1-VFA were observed.

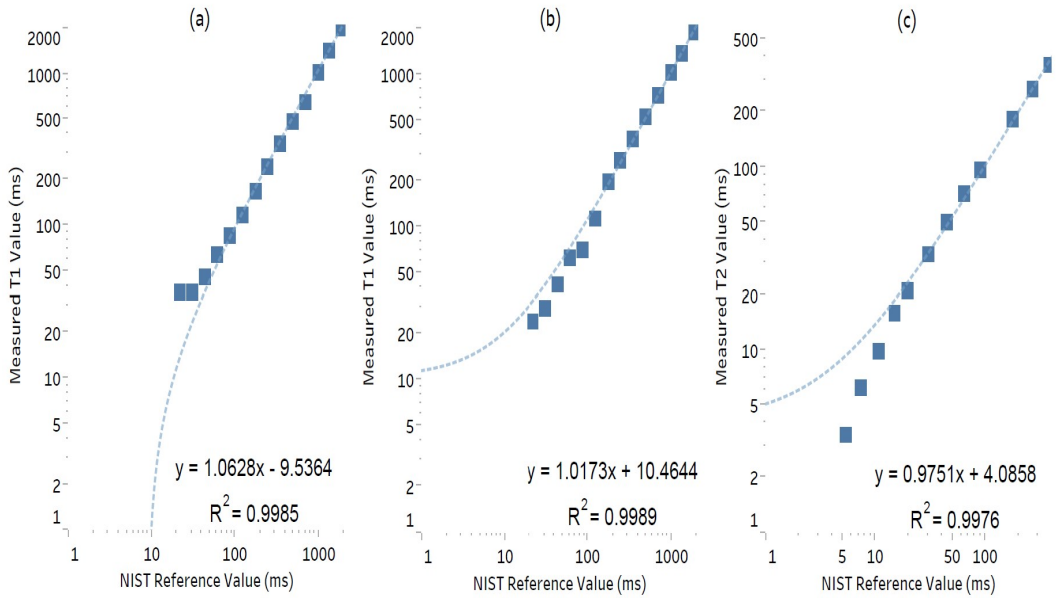

Figure 3.10: A strong linear correlation was found between (full vial range) reference and measured T1 and T2 times. This was true for (a) T1-IR, (b) T1-VFA and (c) T2-SE sequences. Note: All axes have employed a logarithmic scale.

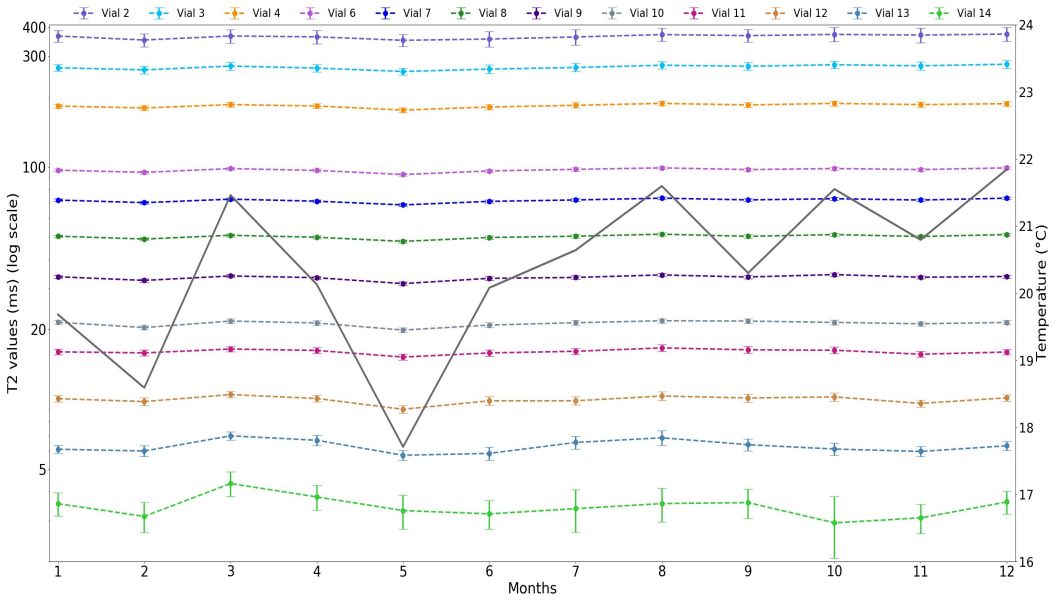

Figure 3.11: Monthly fluctuations observed in T2-SE measurements, with overlaid average temperature readings. Error bars were generated from the standard deviation of each vial (calculated from the parameter fit).

On average, the initial and final temperatures recorded each month was 20.1 $\pm$ 1.5°C, and 20.8 $\pm$ 1.0°C, respectively. The change in temperature over individual imaging sessions was generally less than $\pm$0.5°C. Correlation coefficients (ρ) were calculated between recorded temperature and measured T1 for IR and VFA sequences (ρ = 0.003 and ρ =

-0.001, respectively), and for T2 (ρ =0.007). Similar calculations showed that there was no clear relationship between systematic variations over time with the T1 or T2 measurements ($\rho < |0.001|$).

3.4 Discussion

In this chapter, the reliability of T1 and T2 relaxation time parameters derived using a 3T MRI scanner was assessed. Accuracy and repeatability results presented in this chapter were comparable to previous studies completed using the same phantom type and similar sequences, when including all vials [19, 110, 112].

The advice to remove specific vials from the T2 analysis was at the recommendation of the manufacturer. They believe there were probable mixing or labelling errors that occurred during the manufacturing of vial 5. The issues with vial 1 most likely derived from the storage of the vials' $MnCl_2$ solution prior to manufacturing the phantom: it was stored in glass, and it is suspected that the Mn within plated onto its glass storage bottle [122].

Also, vial 1's solution has a low concentration of Mn, and reducing this further would result in an anomalously longer T2 than expected. These issues have since been resolved by the manufacturer; however, this highlights the need for monitoring qMRI phantoms as well as systems.

3.4.1 Accuracy and Repeatability

Repeatability was improved for T1 and T2 measurements in the human vial range of the phantom compared to the full vial range. The CV of T1 and T2 in this range for all sequences tested was less than 5%. Further, Bland-Altman plots in Figure 3.4 showed the bias of these parameters ranged between approximately -20 to +20%, with an average parameter over-estimation. The average biases measured for T1-IR, T1-VFA and T2-SE sequences were +2.0%, +6.5% and +8.5%, respectively. These bias magnitudes were sig-

nificantly lower compared to thresholds reported in literature (of 20% to 40%) that could lead to erroneous outcomes if used in applications such as tissue discrimination, including distinguishing malignant from benign tissues [68, 118].

Although the average across all three sequences recorded positive biases, drifts in bias were observed for vials with longer reference T1 and T2 times. As shown in Figure 3.4, the T1-IR sequence showed an upward drift from -5.1% for vial 4 (reference T1=690ms), to a mean positive bias of +6.5% for vial 1 (T1=1884ms). In contrast, the T1-VFA sequence demonstrated a downward trend for vials with reference T1 $\geq$ 241ms, moving from a mean positive bias of +13.5% for vial 7 (T1=241ms) towards the zero-line, recording +0.5% for vial 1 (T1=1884ms). These findings are in line with observations made in past research using similar sequences [19, 110, 123, 124].

For the T2-SE sequence, a similar downward trend was seen for reference vials with T2 $\geq$ 88ms (vial 6). This included progressing from a positive bias of +9.2% for vial 6 to a negative bias of -4.0% for vial 2 (T2=380ms). Comparable findings have been reported in the literature for T2 measurements using this phantom [110, 125]. This observation could be due to vials with long T2 times being more susceptible to inaccurate flip angles and eddy currents [110].

The T1-IR method showed superior accuracy and repeatability when compared to T1-VFA. This is consistent with existing literature [19, 118]. Such a finding was expected given VFA methods have been known to overestimate T1-times, and show increased susceptibility to B1-inhomogeneity effects compared to T1-IR; frequently requiring a B1-correction [19]. This comparative approach was employed as T1-IR sequences are widely accepted as a gold-standard in the field (see Section 2.3.4), and thus provide a valuable point of reference for the T1-VFA sequence [19]. Such methodology is common in the literature when evaluating alternative or new T1-mapping sequences, such as novel MRF methods [19, 110, 112, 126].

This study aimed to follow QIBA guidelines by utilizing a common imaging protocol that was open source and could allow for prospective multi-site investigations [7, 115]. Note that no B1-corrections were implemented in this study as there is currently no commonly used correction technique available [127]. A future study would utilise department-specific patient imaging protocols for T1-mapping and compare scanner baseline %bias and repeatability.

3.4.2 Signal Fitting

For the T2-array, signals for shorter T2 vials often approached the noise floor (Figure 3.7). Also, the mono-exponential fitting applied to the T2-SE signal, replicating methods used in the majority of clinical and preclinical studies, is known to be susceptible to inaccuracies generated by B1-inhomogeneities [40, 118, 121]. This can lead to imperfect refocusing flip angles, especially for the first echo (Figure 3.8) and can contribute noise [111, 120, 121]. For these reasons, the first TE and signal were discarded from the fit and a noise factor was introduced in the model fitting procedure [Equation 3.3].

During post-processing of the T1-VFA magnitude images, the average signal from ROI's in vials with shorter reference times (9-14) were observed to have signal saturation. This was especially the case for larger flip angles (20° to 30°). Thus, only magnitude images for FA's 2°, 5° and 10° were used to calculate T1 for the saturated vials, similar to Keenan et al. (see Figures 3.6 and 3.9) [112].

It should be noted that several methods exist for calculating qMRI parameters like T1 and T2, with each algorithm compromising between efficiency, accuracy, and ability to handle low SNR values [128]. The Levenberg-Marquardt algorithm is commonly used in the literature and has reported satisfactory efficiency and accuracy performance, and thus was used in this study [19, 114, 121, 128, 129]. However, this method can be computationally demanding and sensitive to the initial guesses for parameters; if these are far from

the true parameter value, it may lead to highly inaccurate results [130]. The impact of algorithm choice on T1 and T2 quantification, while out of the scope of this study, should be considered in future investigations.

3.4.3 Linearity, Time and Temperature Dependencies

Due to limited scanner time availability, imaging could not be completed on the same day of each month. Instead, a time constraint of at least 2 weeks between imaging sessions was implemented, achieving a 4-week average spacing. Temperature variations between 18-22°C had no observed effect on the measured T1 and T2 times. This was expected for the $NiCl_2$ solutions in the T1-array, with known minimal fluctuations within these temperature ranges [106].

There was a 1.6%/ °C linear dependence expected for the $MnCl_2$ T2-array solutions [106]. However, due to the small temperature fluctuations recorded in this study (averaged within 1°C of the NIST reference conditions), no significant relationship was observed (ρ = 0.007).

There was a strong linear correlation between the reference and measured vial relaxation times ($R^2 < 0.997$). It can be seen in Figures 3.10 and 3.9 that the largest deviations in %bias and repeatability occurred for vials with the smallest relaxation times. This can be partially explained by the acquisition parameters utilized. For example, in the T1-array, vials 13 and 14 had reference times of approximately 30 ms and 21 ms respectively; shorter than that of the first TI (35 ms) used in the T1-IR pulse sequence.

Similarly, for the T2-array, vials 13 and 14 had reference times of 7 ms and 5 ms; less than the shortest TE (10 ms) used in the T2-SE sequence. Detecting shorter T1 and T2 times is often a challenge to scanner's gradient hardware and available sequence acquisition parameters [110]. However, these sequences were utilized as they are commonly available and designed to capture the wide range of relaxation times in the phantom.

There were no trends in variability for T1 or T2 accuracy measurements over the course of the 12-month study (see Figure 3.11). According to the literature, major system upgrades can cause large changes to occur in T1 measurements [112, 131, 132]. During this study, two hardware replacements of the Transmit-Box (containing RF transmitters) occurred between months 7 and 8 and also months 10 and 11. In the months immediately following these replacements, particularly for months 8 and 11, the percentage of pixels with signal saturation reduced by -3.4% and -2.3%, respectively. However, these reductions were not significantly different when compared to other months measurements, and hence the cause of the reduction was not determined.

3.5 Conclusion

High accuracy and long-term repeatability in physiologically relevant T1 and T2 times on a radiotherapy dedicated 3T MRI scanner was found in this chapter. Baseline bias (and CV(%)) for T1-IR, T1-VFA and T2-SE sequences were +2.0% (2.1), +6.5% (4.2) and +8.5% (1.9) respectively. Shorter sample relaxation time vials had increased measurement instability; however, no systematic variations in accuracy over time were observed. These longitudinal and frequent assessments of qMRI scanner technical performance fluctuations are especially important in the case of treatment response monitoring [7].

With the high repeatability of the accuracy measurements observed, similar to Ihalainen et al., it is predicted that future measurements using this scanner would yield similar results [113]. Consequentially, QA frequency recommendations to the department involved conducting testing annually and surrounding the time of any major scanner upgrades. Future work will involve determining deviations in T1- and T2-times when using department-specific sequences (see Chapter 5), and to find the cause of the signal saturation fluctuations in T1-VFA acquisitions for shorter reference time vials.

Chapter 4

Conformance of a 3T Radiotherapy MRI Scanner to the QIBA Diffusion Profile

The results presented in this chapter have been published in Medical Physics:

M. Carr, K. Keenan, R. Rai, M. Boss, P. Metcalfe, A. Walker, L. Holloway. *Conformance of a 3T Radiotherapy MRI Scanner to the QIBA Diffusion Profile*. Medical Physics, vol. 49, no. 7, p. 4508-4517, 2022. DOI: 10.1002/mp.15645

Contributions: M. Carr performed the measurements, analyzed the data, and wrote the manuscript with support from all the authors. K. Keenan, R. Rai, M. Boss, P. Metcalfe, A. Walker, and L. Holloway were involved in planning experiments and supervising the work.

4.1 Introduction

qMRI is increasingly being used in radiation oncology, and especially in the case of DWI-MRI. Recall from Section 2.3.1, DWI-MRI can be used to study a patient's tissue cellularity via calculating the ADC. This measured value has potential to aid in disease diagnosis

and monitoring/predicting treatment responses [8,21]. Further detail regarding the clinical potential of this QIB is outlined in Section 2.3.2. It is however the technical performance uncertainties associated with measuring ADC that are currently limiting the wide-spread clinical implementation of the technique [7, 37, 41, 133].

In 2019, QIBA released their Diffusion Profile [7, 20]. Specifically, this Profile lists the precise methods and specifications that need to be performed/met for a site to ensure reproducible ADC in multi-center clinical trials. It incorporates recommendations based on literature to use standardised phantoms and sequences to establish a scanner's baseline performance levels (e.g., ADC %bias, repeatability, precision, signal to noise ratio (SNR)) [2, 20, 21, 37]. The Profile requires testing to be performed on a pure water ADC sample [20], and there are several phantoms available for this purpose [107, 133–135].

The NIST/RSNA/NCI Diffusion phantom, is a commercially developed phantom that can also be used for ADC QA. It has been used in the past to test the accuracy and repeatability of ADC on/or between MRI scanners [63, 64, 136], MR-linacs [11, 41], and Diffusion Tensor Imaging (DTI) parameters on MRI scanners [109].

The literature is however lacking long term and frequent system stability measurements of ADC (see section 2.4). This is essential for simulating serial scanning in treatment response monitoring [20]. Further, most studies investigating ADC reliability using the NIST/RSNA/NCI phantom report only on using coronal [64, 107, 124] or axial [41, 133] slice-encoded image acquisitions. Although the Profile requires only axial imaging for conformance testing, multi-directional DWI imaging is routinely performed in the clinic, depending on the anatomical site [137–139]. Thus, it is important to determine any imaging directional dependencies on ADC reliability [20].

The main aim of the work presented in this chapter was to assess the long-term technical performance of ADC on a 3T radiotherapy dedicated MRI scanner. Factors important

for the clinical imaging of different anatomical sites with uncertain effects on scanner performance were investigated, including multi-directional imaging and ADC linearity.

4.2 Methods

4.2.1 Image Acquisition and Phantom setup

The design of the diffusion phantom used in this study (phantom serial number: DP128-A-03-0113), has been described in Section 2.4.3. The phantom was imaged at monthly intervals over one year (at a minimum of 2 weeks apart) using the same 3T MRI scanner as was used in Chapter 3.

The system-specific phantom scan protocol and parameters used in this study are outlined in the QIBA Profile [20]. This included the use of a 2-D single shot echo-planar imaging (SS-EPI) sequence (scan time of approx. 2 minutes), with a 3-scan trace and four *b*-value diffusion weightings (s/mm^2): 0, 500, 900, and 2000. The echo-time (TE) and repetition-time (TR) used for imaging were 10,000 ms and 106 ms, respectively.

Additionally, the QIBA Profile required standard shimming procedures be employed in the imaging protocol to improve the uniformity of the magnetic fields of the scanner [20]. For the Siemens scanner used in this study, the standard shimming methods for EPI sequences involved setting the B_0 shim mode to 'standard', and B_1 shim mode to True-Form. This is where TrueForm refers to the use of TimTX Trueform RF transmitter coils (Siemens Healthineers, Erlangen, Germany) [44].

Prior to scanning, the phantom was filled with an ice-water bath and refrigerated for a minimum of two hours to achieve thermal equilibrium at 0°C [20, 108]. Immediately prior to scanning, the phantom was refilled with ice and the temperature was measured using the same NIST-traceable thermometer as described in Chapter 3. Temperature was measured again immediately after scanning.

For phantom alignment, the central water vial was aligned to isocenter within a 20 channel Head/Neck coil 4.1. The phantom was manually repositioned from its axial orientation to coronal and then sagittal within the Head/Neck coil (as described in the phantom manual) [108].

Figure 4.1: Diffusion phantom in its axial orientation within a head/neck coil. A 3D-printed base holder and stabilizing aids were used to assist in reproducible positioning

Simultaneously to each physical rotation, the slice-encoding (and phase-encoding for coronal) direction was changed to match the respective phantom orientation, maintaining Figure 2.10's vial arrangement in the generated ADC-maps. Each long-term (LT) monthly acquisition included repeating the SS-EPI sequence four times to acquire the short-term (ST) measurements in each phantom orientation, as per QIBA Profile guidelines.

Note that precautionary methods were implemented to minimize the likelihood of susceptibility-induced distortions caused by air-gaps in the ice bath (see Section 2.3.2). Firstly, crushed ice was used to reduce any initial air-gaps in the phantom. Secondly, the phantom was removed from the scanner and topped-up with more crushed ice and cold water between physical phantom rotations if required. The latter was only performed if substantial air-gaps were detected in the first set of images collected per phantom orientation, as viewed

on the console. All four repetitions were then additionally acquired once the phantom had been returned to its correct position within the scanner.

Both DW images and scanner generated (inline) ADC maps were exported from the Siemens Syngo Workstation to preserve DICOM meta data. The inline maps were calculated using a linear regression analysis by fitting the signal for all b-values, S(b), to the mono-exponential model described in Equation 2.5 [49].

4.2.2 Region of Interest Analysis

Using the first repetition ADC map measured for each imaging direction, the central pixel location was manually identified for each of the vials using ImageJ v1.53c (National Institutes of Health, Maryland, USA). These locations along with all four repetitions of DW images and ADC maps were imported into an in-house developed Python analysis script (see Appendix A).

Circular ROIs of 1.2 cm diameter, covering approximately 109 pixels, were positioned over the center of each of the 13 vials on three central phantom slices for statistical analysis (average pixel intensity calculated over the volume of interest (VOI)). Phantom slices found to have major artefacts occurring near any of the 13 vials were excluded from analysis via shifting the entire VOI selection away from the affected slice.

4.2.3 QIBA Profile Analysis

The Profile required the assessment of seven key measurements, with calculations and definitions outlined in Table 4.1 and tolerance limits in Table 4.2. Measures for tests A-E were calculated using the inline derived ADC maps, whilst tests F and G required the DW images. Further details on the methods used to complete this testing can be found in the Profile documentation [20].

4.2.4 Software Validation and Spatial Dependence

The QIBA Profile recommended investigating the analysis software used for testing Profile conformance. To do this, computer generated DICOM DRO datasets with *b*-values of 0, 500, 800, and 2000 s/mm^2 were imported into the offline DWI-fit Python script [102]. The offline fitting method was like that described for the inline ADC map derivation.

Offline ADC maps of the DROs were produced to estimate the %bias and standard deviation (SD) over a range of phantom relevant SNRs (50 to 100) and ADC values (0.1 - 1.1 μ m^2/ms) [20,140]. The same fit was used on the first repetition of each monthly DW image from the axial phantom scans (datasets = 12), using the central water vial ROI from tests A-E. R^2 was used as a measure of goodness-of-fit in both cases, and inline versus offline ADC values were compared to assess %bias.

The offline script was also designed to identify ROIs in the DW images that had experienced signal saturation. Signal saturation, also known as data clipping, occurs when the signal received is outside of the system's detectable range. In addition, the same SNR code as used in test F was used on the four repetition DROs available [102].

Table 4.1: Overview of tests completed to assess conformance to the Profile [20]. Note: Excluding test D, Profile testing was only required to be performed on the central water vial (at isocenter) using axial acquisitions. Further, short-term (ST), refers to the intra-day measurements acquired, whilst long-term (LT) refers to the intra-month measurements acquired.

Test	Relevant Equations	Definitions	
A	$\%\text{bias} = \frac{\mu - DC_T}{DC_T} \times 100\%$	DC_T	= True Diffusion Coefficient
		μ	= Mean of Measurements
B/C	$RC_{ST/LT} = 2.77 \times SD$	RC	= Repeatability coefficient
		CV	= Coefficient of variation
		SD	= Standard deviation
	$CV_{ST/LT} = 100\% \times \frac{SD}{/mu}$	$_{ST}$	= Over 4 x short-term measurements
		$_{LT}$	= Over 12 x long-term measurements
D	$R^2 = 1 - \frac{RSS}{TSS}$	R^2	= Coefficient of determination
		RSS	= Sum of squares of residuals
	AND	TSS	= Total sum of squares
		Y	= Measured ADC (all vials/months)
	$Y = \beta_0 + \beta_1 \times DC_T$	β_0	= Intercept
		β_1	= Slope
E	$CV_P = 100\% \times \frac{SD_{pix}}{\mu_{ROI}}$	ROI	= Region of interest (isocenter vial)
		SD_{Pix}	= Over ADC values within the ROI
F	$SNR = \frac{\mu_{ROI}[SignalImage]}{\mu_{ROI}[NoiseImage]}$	SNR	= Signal to noise ratio
		Signal	= Average of pixels values for each ROI over the 4 x ST repetitions
		Noise	= Average of pixel SD values for each ROI over the 4 x ST repetitions
G	$Dep_b = 100\% \times \lvert\frac{ADC_{b_0,b_{n+1}} - ADC_{b_0,b_n}}{ADC_{b_0,b_n}}\rvert$	Dep_b	= b-value dependence
		$ADC_{b_0,n_{n+1}}$	= ADC generated using b_0= 0 s/mm^2 and b_{n+1}, where $b_{n_1} > b_n$
		b_{1-3}	= 500, 900 or 2000s/mm^2

Following personal communication with QIBA, an estimate of spatial dependence (Dep_S) was completed using the diffusion phantom. Specifically, the %bias deviations along the lengths of the central water vial (axially), outer-ring water vial (axially), and central water vial (coronally), were respectively used to assess superior to inferior (SI), right to left (RL) and anterior to posterior (AP) spatial dependencies at approximately 4 cm from isocenter.

A Profile tolerance of ± 4 % was stipulated for each individual direction [20].

4.3 Results

The phantom was imaged 12 times over a one-year period, with an average four-week interval between imaging sessions. Average pre- and post-scan temperatures were -0.1 ± 0.1 °C and 0.0 ± 0.2 °C, respectively. For any monthly imaging session, the maximum (absolute) temperature changes pre- and post-scanning was 0.4 °C.

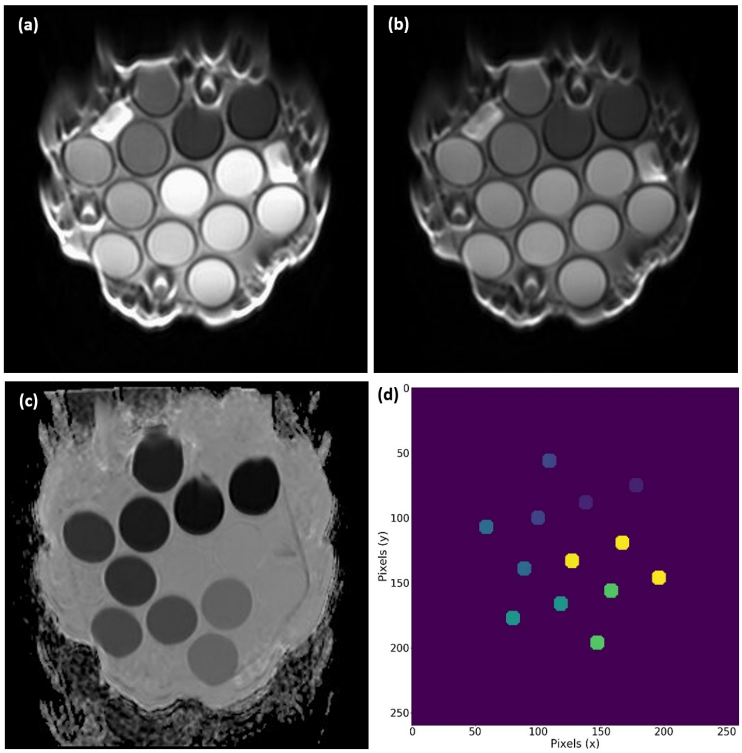

Figure 4.2: Month 7, repetition 1, axial images of one of the phantom's central slices, including the (a) isotropic diffusion weighted images with *b*-value = 0 s/mm^2 and (b) *b*-value = 500 s/mm^2. (c) Also, the inline derived ADC-map with (d) corresponding ADC regions of interest for all 13 vials is shown.

Typical DW images and inline ADC maps (and respective ROIs) are shown in Figure 4.2. Susceptibility-induced distortions in the ADC maps were primarily observed in outer ring 40 % and 30 % PVP vials for axial and sagittal acquisitions, respectively (see Figure 4.3). Consequently, central VOI's were selected to mitigate the observed distortions. A summary of the Profile test tolerances and acquired results for the central water vial are listed in Table 4.2.

Table 4.2: Accuracy, short-term repeatability, long-term repeatability, linearity, random error, SNR and *b*-value dependence (tests A to G) tolerance limits and mean value ± SD (where applicable), obtained from the 12 monthly measurements of the central water vial (as per Profile requirements). [a]Certain month's data have been excluded from the presented results due to retrospective findings of signal saturation occurring within the datasets. For SNR, this included excluding axial results acquired for months 1, 2, 6 and 9, and coronal and sagittal results for months 1 and 5, and 1 and 6, respectively. For Dep_b, results for month 1 were excluded in calculations for all directions.

Test	Performance Metric	Profile Tolerance	Axial Result	Coronal Result	Sagittal Result
A	$\lvert\%bias\rvert$	<3.60	+ 0.05 ± 0.01	+ 0.83 ± 0.00	+ 0.29 ± 0.01
B	$RC_{ST}(\mu m^2/ms)$	≤ 0.015	0.003 ± 0.001	0.005 ± 0.002	0.003 ± 0.001
	$CV_{ST}(\%)$	≤ 0.5	0.1 ± 0.0	0.1 ± 0.1	0.1 ± 0.0
C	$RC_{LT}(\mu m^2/ms)$	≤ 0.065	0.028	0.011	0.027
	$CV_{LT}(\%)$	≤ 2.2	0.9	0.3	0.9
D	R^2	> 0.9	1.0	1.0	1.0
	$Slope(\beta_1)$	$0.95 \leq \beta_1 \leq 1.05$	1.00	1.02	1.02
E	$CV_p(\%)$	< 2	0.38 ± 0.10	0.43 ± 0.04	0.38 ± 0.04
F	SNR^a	$\leq 50 \pm 5$	332 ± 146	269 ± 93	356 ± 68
G	$Dep_b(\%)^a$	< 2	0.4 ± 0.3	0.3 ± 0.2	1.4 ± 2.1

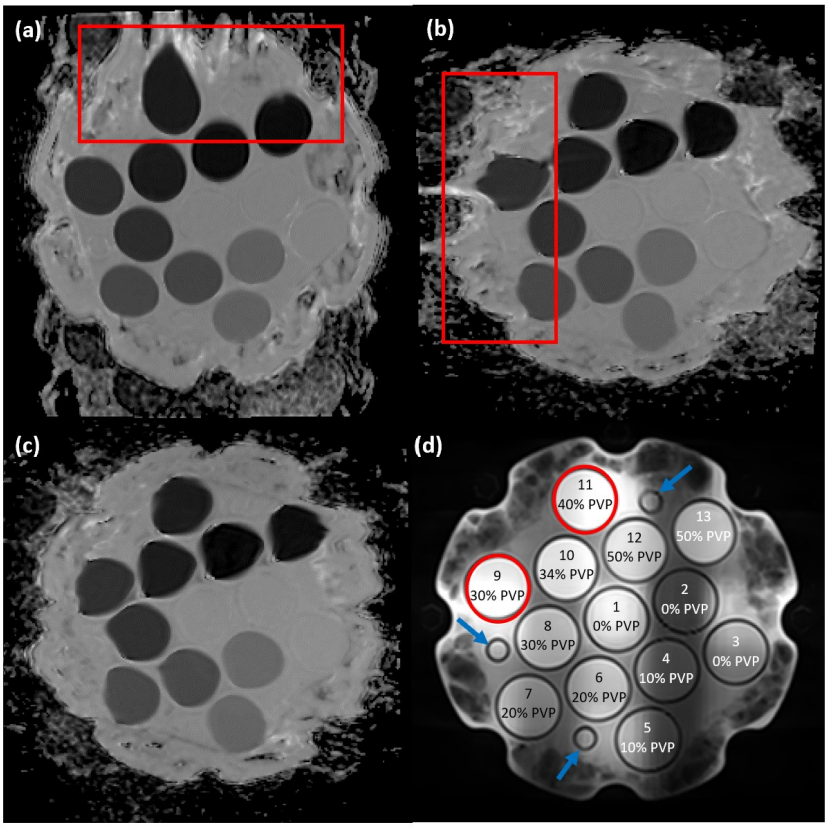

Figure 4.3: Month 7, repetition 1, central slices of the inline derived ADC maps for (a) axial, (b) sagittal and (c) coronal acquisitions. The phantom in (a) and (b) is orientated such that vials located within the red boxes areas are at the phantom's most anterior point. There were also vials in close proximity to the three 5 mL fiducial markers (indicated by the blue arrows in the localiser image (d)). These factors combined cause for the vials circled in red to be at highest risk of the susceptibility-induced distortions.

4.3.1 ADC Accuracy (A), Short (B) and Long-term (C) Repeatability, Linearity (D) and Precision (E)

Figure 4.4 highlights that Profile tolerance limits for short and long-term repeatability (excluding the axial CV_{LT} of the outer-ring 10 % PVP vial) and precision, were met for all vials (1-8) with concentrations 0 – 30 % PVP (ADC range: 0.4 – 1.1 $\mu m^2/ms$). The average ADC calculated for all directions/vials can be found in Table 4.3 and Figure 4.5.

In general, vials with lower diffusivities (higher concentrations of PVP) had inferior performance metrics, and even the sign of the %bias measurement varied for different imaging directions. The inner-ring 50% PVP vial (Figure 4.4) had an axial bias up to + 13.61

% and CV_{LT} = 5.8 %, coronal bias of -11.64 % and CV_{LT} = 9.28 %, and sagittal bias of -21.82 % and CV_{LT} = 7.4 %.

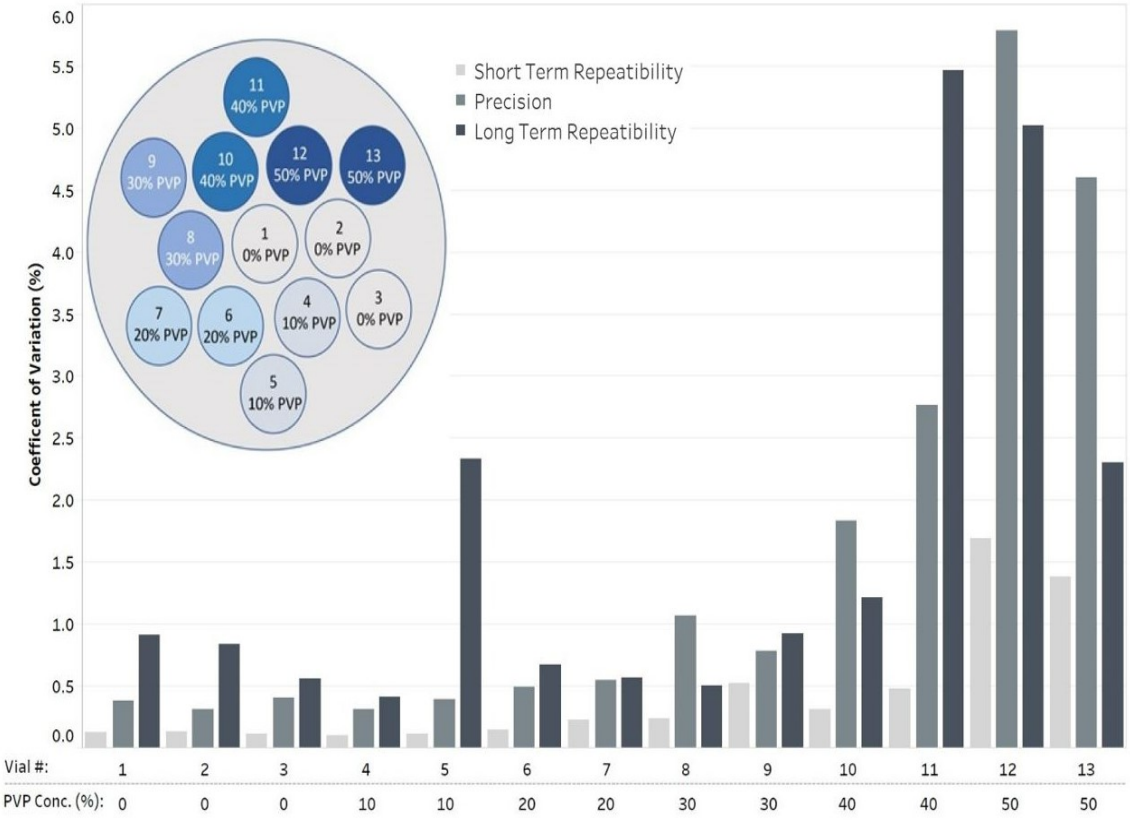

Figure 4.4: Short-term repeatability (CV_{ST}), precision (CV_P) and long-term repeatability (CV_{LT}) coefficients of variation derived for each vial for axial acquisitions, calculated as an average over the 12-month study. A vial arrangement diagram has been included in the top left for positional reference for all 13 vials. PVP concentrations (conc.) are shown (by mass fraction (%)) for inner- and outer-ring vials.

An overview of monthly %bias results for the central water vial are presented in Figure 4.6 for all repetitions and each imaging direction. It can be observed that the %bias was well within Profile tolerance range ($\pm$ 3.60%). From this figure, it is also evident that all within-session repetitions generated similar magnitude ADC values, whereas monthly repetitions fluctuated (around 0 % bias). Specifically, no monotonic trends in ADC variability with time were found over the four within-session repetitions, nor were any changes in artefacts observed.

Table 4.3: Mean and standard deviation (SD) of the Apparent Diffusion Coefficient (ADC) value ($\mu m^2/ms$) measured over the 12 monthly acquisitions at 0 °C, and for each orthogonal imaging direction. [a] NIST reference ADC values have also been presented: measured for the same batch of PVP samples as those embedded in the phantom's vials, at 0 °C.

Vial #:		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
ADC	PVP (%):	0	0	0	10	10	20	20	30	30	40	40	50	50
NIST[a]	Value	1.109	1.109	1.109	0.817	0.817	0.579	0.579	0.380	0.380	0.220	0.220	0.110	0.110
	SD	0.025	0.025	0.025	0.019	0.019	0.015	0.015	0.011	0.011	0.007	0.007	0.005	0.005
Coronal	Value	1.118	1.126	1.154	0.839	0.810	0.591	0.589	0.400	0.396	0.229	0.220	0.097	0.128
	SD	0.004	0.004	0.006	0.002	0.002	0.001	0.003	0.003	0.005	0.002	0.003	0.009	0.007
Sagittal	Value	1.112	1.125	1.129	0.836	0.818	0.600	0.605	0.404	0.397	0.228	0.214	0.086	0.094
	SD	0.010	0.008	0.003	0.002	0.003	0.001	0.003	0.003	0.004	0.003	0.003	0.007	0.010
Axial	Value	1.110	1.118	1.132	0.834	0.837	0.595	0.600	0.394	0.406	0.236	0.245	0.125	0.114
	SD	0.010	0.009	0.007	0.004	0.020	0.004	0.003	0.002	0.003	0.003	0.013	0.006	0.003

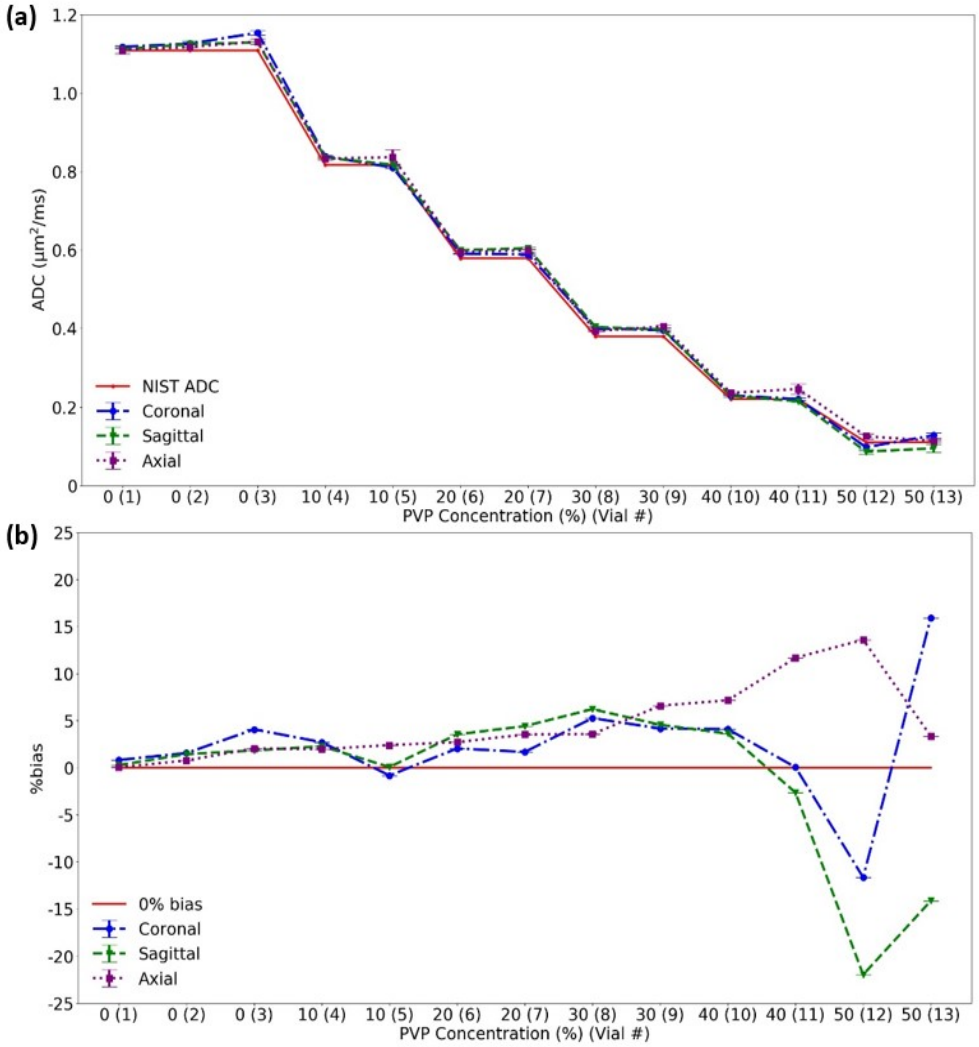

Figure 4.5: Graphical depiction of Table 4.3 data; (a) Presenting the mean and standard deviation (SD) of the Apparent Diffusion Coefficient (ADC) value (μm^2/ms) measured over the 12 monthly acquisitions at 0 °C for each orthogonal imaging direction; and (b) Presenting the same values as a %difference from the NIST ADC value. Note that vial 1 represents the central water vial with 0% PVP concentration.

Using all 13 vials average ADC over the duration of the 12 months study, a strong, positive, and linear correlation was found (Figure 4.7) between measured and reference ADC values for all directional acquisitions ($R^2 > 0.99$). Similarly, the slopes (β_1) in Figure 4.7 were all within the Profile tolerance range. It can be observed that all inner- and outer-ring vials of the phantom performed similarly, excluding 40 % and 50 % PVP vials (lowest ADC), which also had the largest SDs.

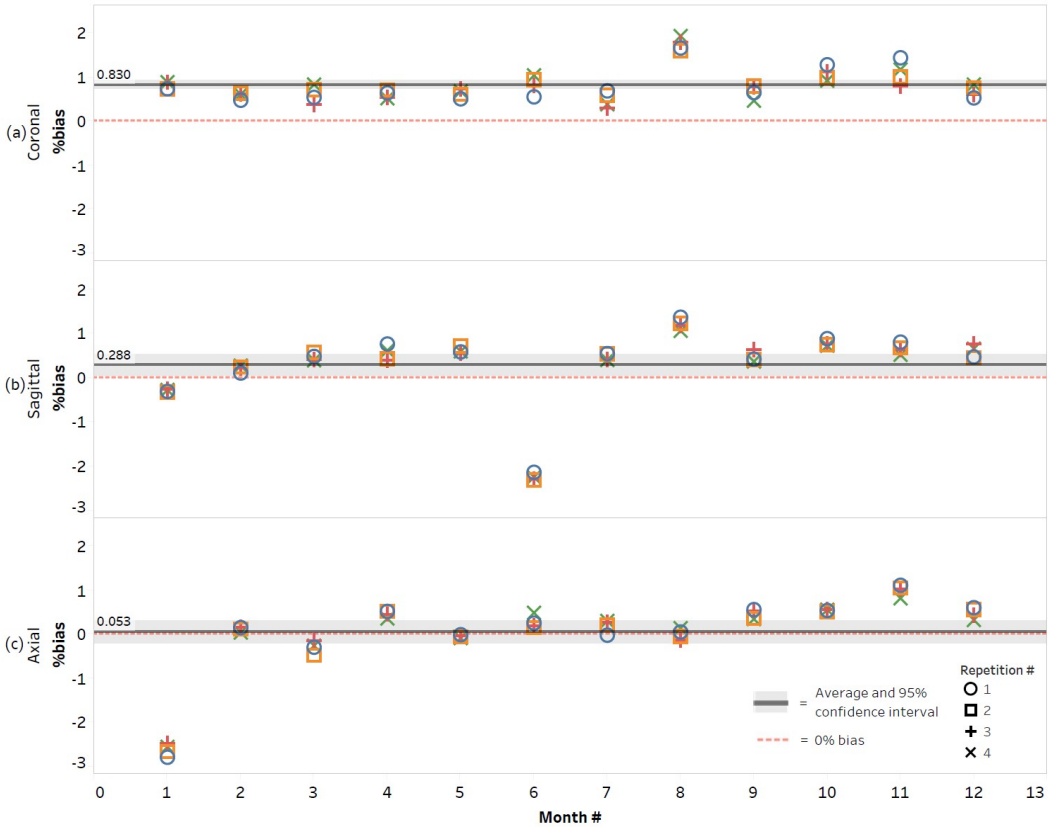

Figure 4.6: Bland-Altman plots for the central water vial coronal (a), sagittal (b) and axial (c) acquisitions, show the difference (%bias) between measured and reference ADC values over the 12-month study. The average %bias (and 95 % confidence intervals) is displayed and includes +0.830 % (+ 0.724 - +0.936), +0.288 % (+0.042 - +0.534) and +0.053 % (-0.204 – +0.309) for coronal, sagittal and axial acquisitions respectively.

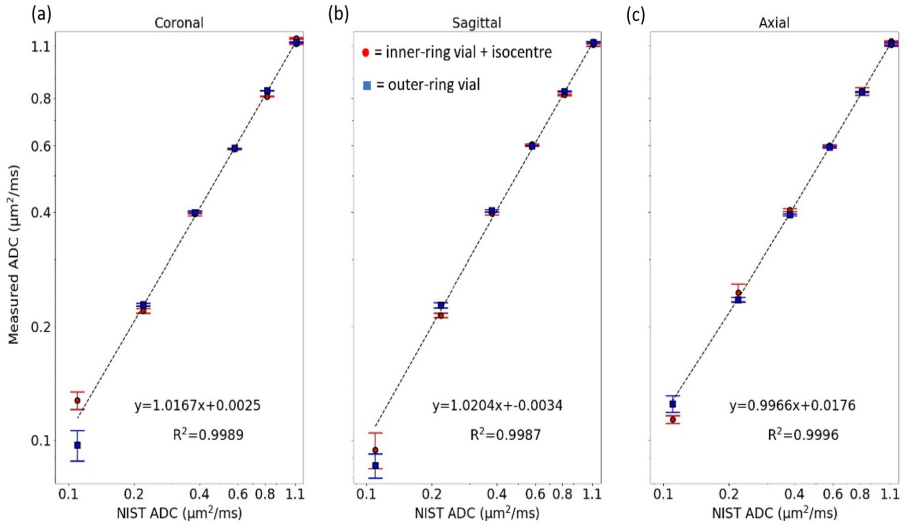

Figure 4.7: Correlation plots for coronal (a), sagittal (b) and axial (c) acquisitions with measured ADC values compared to the NIST reference ADC values for each vial. Note: All axes have employed a logarithmic scale and error bars are given as the standard deviation found between all 12 months ADC values. Also, the negative y-intercept of T1-time for the T1-IR sequence is unrealistic and caused by the large outliers recorded for certain vials.

4.3.2 SNR (F) and Software Dependence

On the axial *b*-value = 0 s/mm^2 magnitude images, pixel signal intensities in the ROI of the central water vial saturated for months 1, 2, 6 and 9. This resulted in failed SNR calculations (high signal, and minimal noise), and thus the SNR presented in Table 4.2 is an average over only 8 months of repetitions. Similarly, the coronal and sagittal acquisitions experienced saturation in the central water vial for months 1 and 5 and months 1 and 6, respectively. Thus, the SNR was only calculated over 10 months in both cases.

In assessing the image analysis software, equivalence (within SD) of inline versus offline axial central water vial ADC values were found: 1.110 ± 0.010 μm^2/ms and 1.112 ± 0.010 μm^2/ms, respectively. A strong correlation ($R^2 > 0.993$) between the two methods' measured ADC values were found when including all vials, with the inline method slightly underestimating the ADC on average by 0.2 %.

The offline derived ADC value fluctuations is shown in Figure 4.8, highlighting signal saturation minimally affected the ADC (for months 2, 6 and 9). However, larger SDs were found for saturated fits, especially for month 1.

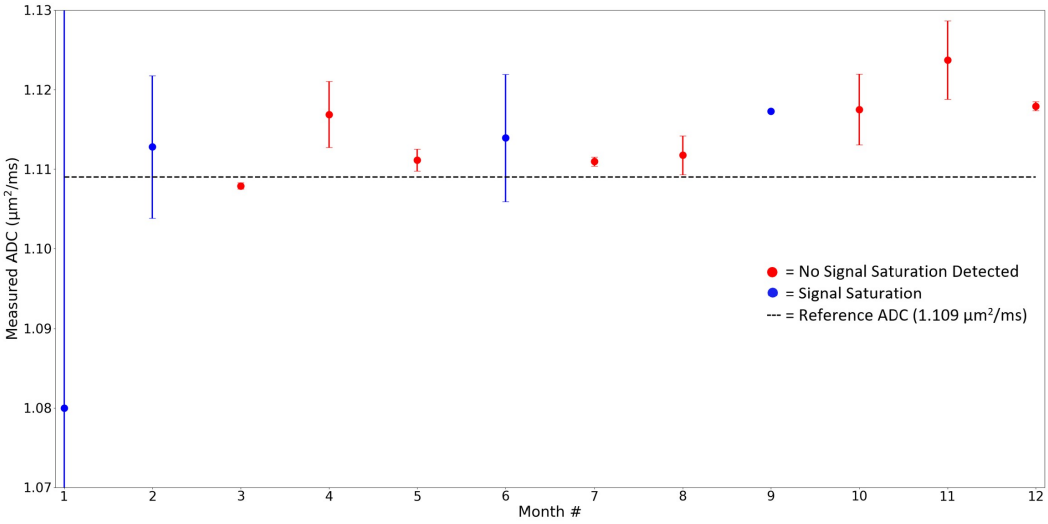

Figure 4.8: Observed monthly deviations for the central water vial ADC value over the 12-month study. Only the first repetition of axial acquisitions has been included, with ADC maps derived offline. The ADC is thus presented as the ADC value ± SD (calculated using mean square error). Not: The SD for month 1 extended to ± 0.578 μm^2/ms.

For the DRO study, a goodness-of-fit of $R^2 = 0.995$ was found with %bias ± SD remaining within the $\leq$ 3.60 % tolerance for most ADC/SNR combinations in phantom relevant ranges (Figure 4.9). For the DRO ROI of ADC = 1.1 μm^2/ms and SNR=100, a slight overestimation of SNR was found, 104.9 ± 4.5 (with 95% confidence interval (CI)), using the SNR analysis method implemented for test F.

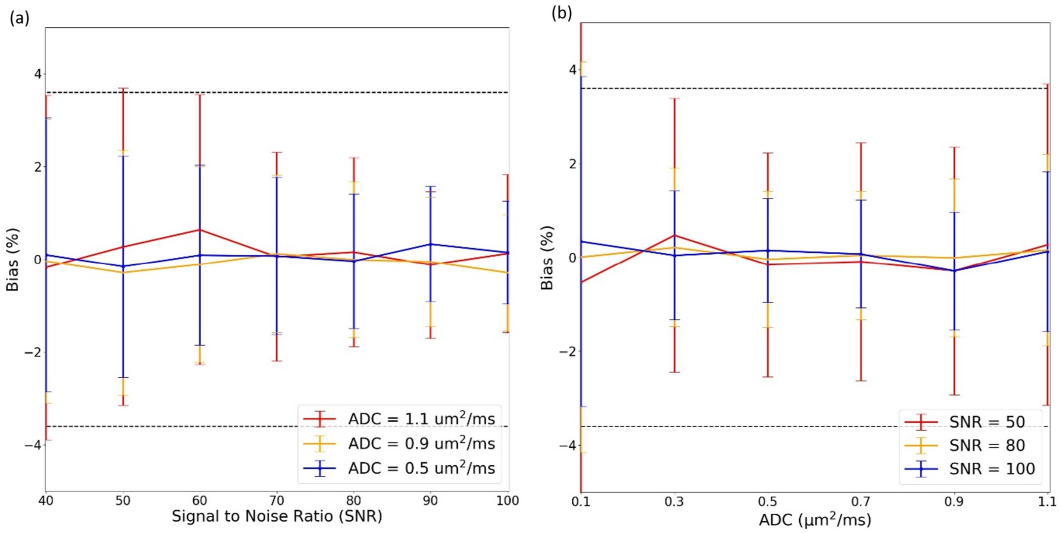

Figure 4.9: Examples of %bias for DWI-digital reference object (DRO) ADC maps derived offline. %bias and SD (error-bars) are presented with respect to input (a) ADC, and (b) signal-to-noise ratios (SNRs), over phantom relevant SNR and ADC DRO regions of interest, respectively. The Profile %bias tolerance of ± 3.6% has also been included.

4.3.3 b-value (G) and Spatial Dependence

It should be noted that month 1 results were excluded from this test due to considerable signal saturation effects causing axial *b*-value dependencies to be up to a maximum of 11.8 % for pairs $ADC_{0,500}$ and $ADC_{0,900}$. When considering only the remaining 11 months of measurements, axial *b*-value dependence was largest when calculated for the $ADC_{0,500}$ and $ADC_{0,2000}$ pair: 0.6 ± 0.4 % (mean $\pm$ SD). This was followed by respective dependencies of 0.4 ± 0.2 % for pairs $ADC_{0,500}$ and $ADC_{0,900}$, and 0.3 ± 0.1 % for pairs $ADC_{0,900}$ and $ADC_{0,2000}$.

Further considering the later 11 months for both axial and coronal plane measurements, all pair dependencies stayed below 1.3 % and were within a 0.4 % difference from the monthly average of all pairs. Sagittal *b*-value dependence, however, performed outside of Profile tolerance limits. This was largely due to month 6's sagittal measurements (affected by saturation), causing high dependencies of up to 12.0% for pairs $ADC_{0,500}$ and $ADC_{0,2000}$.

For spatial dependence, AP bias on average over all months was + 0.71 ± 0.85 %, followed by SI at + 0.43 ± 1.60% and RL of + 0.34 ± 1.45 %. All %bias stayed well within the $\pm$ 4% tolerance on average; however, SI varied substantially for months 1 and 2.

4.4 Discussion

In this chapter, the ADC derived on a 3T dedicated radiotherapy MRI scanner was found to be accurate, repeatable, and reproducible using systematic image acquisitions over one year. By using a standardized set of testing procedures such as the QIBA Profile, the ADC measured by the scanner in prospective single or QIBA-certified multi-site patient-based studies can be said to be reliable, with negligible contributions to ADC due to systematic errors [20, 133].

The average axial %bias (+ 0.05%) and repeatability (CV_{ST} = 0.1 %) for the central water vial (at isocenter) were measured to be well within conformance limits and were comparable to measurements reported in the literature: %bias $< \pm$ 4.3 % and $CV_{ST} <$ 3.2 % [11, 41, 107, 133]. Long-term system stability of deriving the ADC (axial CV_{LT} = 0.9 %) was also found in this study.

All acquisition parameters remained constant throughout the study, and the PVP solutions embedded in the phantom are known to be chemically stable [134]. Previous studies using this phantom found similar results, with the CV_{LT} to be within ranges of $<$ 2.2 % when acquiring 2 scans within 6 months on the same scanner [11, 124], and CV = 2.1 % when comparing between multiple MRI scanners [109].

Overall, this study showed there was minimal imaging directional dependence on ADC performance; a factor not investigated in previous studies using the same phantom. The worst accuracy/repeatability was observed for vials with lower diffusivities (40 % and 50 % PVP), in agreement with the literature [11, 37, 63, 107, 109]. Further in accordance with the literature [107, 109], there were no significant differences ($p > 0.05$) between inner- and outer-ring vial ADC values measured.

The large deviations observed for the lower diffusivity vials could be due to several factors including: eddy currents [20]; increased likelihood of susceptibility-induced distortions occurring near the higher concentrated PVP vials; gradient non-linearities, which can impact the ADC measured at farther distances from isocenter [11]; or from insufficient contrast to noise ratio (CNR) or SNR to correctly assess the ADC in the highly concentrated vials [107]. It should be noted that the 40% and 50% PVP vials at 0 °C have ADC values below physiological range [63].

The signal saturation observed in this study has not been reported in patient-based ADC imaging within the department, nor in previous literature investigating ADC variability.

This effect is not easy to detect: in the offline computed ADC maps, saturation was found to primarily affect the SDs derived from the DWI-ADC fit, whilst minimally impacting the actual ADC value. Saturation was mainly investigated due to the prior saturation findings discussed in Chapter 3. Further investigations, out of the scope of this thesis, would be required to find the cause and factors affecting the signal saturation.

Recall from Section 2.3.1 that in DWI-MRI image acquisitions, the *b*-value plays an important role. This parameter controls the amount of diffusion weighting applied in the sequence, resulting in greater diffusion-weighted contrast at higher *b*-values [2, 41]. The range of *b*-values selected in a sequence considers several factors. These include the diffusion characteristics of the tissue, the available SNR for optimizing diffusion contrast, and the need to suppress the effects of perfusion (blood flow), particularly at lower *b*-values [21, 41]. *b*-value combination recommendations for some anatomies, including for brain, liver, prostate, and breast, is provided in the Profile [20].

The *b*-value dependency in this study for all *b*-value pairs were within the 2 % Profile tolerance (even when including the saturated data in month 1). Pair $ADC_{0,500}$ and $ADC_{0,2000}$ had the largest *b*-value dependence. However, there was no significant difference ($p > 0.05$) between this pair and other combinations, including when all four *b*-values were used for the offline fit. Since the reference sample tested was distilled water (known to demonstrate a mono-exponential behaviour), this finding was expected and in agreement with the literature [11, 133, 135].

Adequate ADC fits and SNR calculations were found in this study by using offline analysis methods to analyse the DRO datasets. With only slight differences between inline and offline derived ADC values (correlation $R^2 > 0.993$), confidence was assured in utilizing the inline generated ADC-maps for majority of analysis as per departmental request; to use the same analysis method as implemented for patient images acquired on the same scanner. The underestimation of inline derived ADC values compared to offline has been

noted to occur in past studies and is likely vendor-specific [64]. Offline methods would need to be used for future multi-site studies to ensure the occurrence of a standardized analysis pathway [64].

The potential of temperature changes affecting the measured ADC, which can be up to 2.4 %/ °C [7, 21, 133], was removed using an ice-bath. However, preparing the ice-bath requires considerable time and can increase the occurrence of susceptibility-induced distortions (which are known to commonly occur in EPI-DWI acquisitions) [63, 134]. Quantifying the impact of these distortions on the ADC measurement would require further investigation [141]. With the recent release of room-temperature diffusion phantom reference ADC values, the need for ice-baths should be minimized in future studies (see Chapter 5) [108].

For spatial dependency, the average %bias over the 12-month study measured in each direction from isocenter was well within $\pm$ 4 % tolerance. It should be noted that the spatial offsets examined for this test were less than the offsets recommended by the Profile ($\pm$10 cm from isocenter). Consequentially, complete characterisation of the spatial dependence could not be achieved. Recommendations for future investigations would include the use of a large homogeneous phantom for this assessment. Although this testing is less important for small fields of view, diffusion studies completed on relatively uniform anatomies like brain have been shown to be significantly impacted by effects such as gradient non-linearities, and thus should be monitored [142].

Additionally, patient ADC values can also be biased by inadequate system SNR [20, 133]. In this study, all directional SNRs were found to be sufficient and considerably higher than that found in past studies, which failed to meet conformance (however, such studies used 1.5 T MRI-based systems) [11, 133]. It should be recognised that the results obtained in this study were for the assessment of baseline scanner performance. Given that phantoms lack tissue complexity, results presented in this study such as the SNR are likely superior

when compared with patient-based imaging [18]. For assessing clinical conformance to anatomy-specific Profile claims, in vivo test-retest assessments should be completed (e.g., for brain and prostate) [20].

There were other limitations in this study, including that imaging did again not occur on the days directly surrounding the two scanner upgrades involving the replacement of the Transmit-Box. Although no clear relationship between ADC value fluctuations and the timing of the upgrades were found like in the work described in chapter 3, similar upgrades have been found to affect patient-based ADC values in the past and should be closely monitored [143]. Future investigations will involve completing similar baseline testing (although at less frequent intervals) in a multi-site trial to validate results found, including changing the imaging direction and reference ADC values.

Findings from this study have led to department recommendations to conduct ADC QA testing annually, and directly before and after commencing a multi-center trial. This QA is in addition to performing the testing at times surrounding any major scanner upgrades. Specifically, completing this testing in only one imaging direction (axial as per QIBA guidelines) and on a pure water sample was considered sufficient following baseline performance measurements.

The work presented in this chapter extends knowledge in understanding ADC long-term variability on clinical MRI scanners. To the best of the authors' knowledge, no prior study has reported in detail testing all aspects of the QIBA Diffusion Profile. Particularly, this study demonstrated Profile conformance over a wide range of physiological relevant ADC values and over three orthogonal imaging directions using a novel diffusion phantom.

These are important findings for future clinical applications whereby patient and consequently phantom QA imaging is required in alternate directions and over different ADC valued anatomies [21, 137–139]. Finding high reliability in the ADC values derived pro-

motes the use of ADC in clinical trials to monitor and assess long-term treatment responses: essential for progressing the clinical implementation of qMRI technology.

4.5 Conclusion

In this chapter, the technical performance of a 3T dedicated radiotherapy MRI scanner was quantified over a 12-month period. Specifically, QIBA Profile conformance specifications were met, including adequate axial imaging accuracy (bias = + 0.05 %), short-term repeatability (CV_{ST} = 0.1 %) and long-term repeatability (CV_{LT} = 0.9 %). Whilst phantom-based results can be effectively used to assess baseline scanner performance, test-retest patient-based studies would be required to examine clinical conformance to anatomy-specific Profile claims.

Recommendations to the department regarding future ADC QA included completing conformance testing annually. This involves only axial imaging on a highly purified water sample: independent of the anatomical sites planned for prospective imaging.

Chapter 5

Developing QA methods for Anatomy-Specific qMRI Sequences

5.1 Introduction

In the previous two chapters, the accuracy and long-term repeatability of certain QIBs were investigated. These investigations were essential in determining scanner baseline performance metrics, and ensuring that systematic fluctuations in the scanner were not affecting long-term patient derived QIB values. However recalling from section 2.1.2, ART plans can be adapted on a daily basis. Given the prospective goal of using QIBs to assist in ART, it is important to determine if scanner performance can fluctuate within this short time period.

Furthermore, one limitation discussed in the last two chapters was that the sequences employed were generic in design. In clinical practice, sequences are designed to image specific patient anatomy. Until recent years, radiation oncology departments had limited methods or guidelines available to accurately assess their qMRI sequences' performance (see section 2.4). Thus, there exists unknowns in the accuracy and repeatability of the QIB values derived in many departments patient-based datasets.

There have been several studies completed investigating the performance of different T1- and ADC-mapping sequences in terms of accuracy and repeatability [19, 45, 123, 126]. One study in particular by Bane et al. (2018) used the System phantom to investigate different department-specific T1-mapping imaging protocols [19]. Results found accuracy and repeatability to be highly dependant on the sequence type and acquisition parameters used.

Similarly for ADC-mapping, the robustness of EPI, RESOLVE and ZoomIt sequences (see Section 2.3.1) was rigorously tested by Liney et al. (2015) [45]. This work involved imaging a phantom at ambient room temperature (AT) daily over one month. Overall, RESOLVE was found to have the highest repeatability ($\%CV_{daily}/\%CV_{weekly}$ = 0.66/3.57%) [45]. There are a few alternative studies which also have investigated the daily and/or weekly ADC repeatability [133, 144–146].

Most of the literature in this space however use single sample phantoms which are often uncharacterised [45, 133, 144, 146], and thus accuracy cannot be assessed. However, there was a recent study by Subashi et al. (2022) where they performed weekly ADC and T1-time measurements over one month on an MRL using the well characterised NIST phantoms [145]. High weekly repeatability and accuracy were found with the median ADC and T1-time %CV (and %bias) being 1.3% (-0.8%) and 1.1%(-0.1%), respectively.

Other past studies using the System phantom investigating the daily T1-time repeatability have used specialised MRI techniques such as MRF [40, 110, 125]. For example, Kato et al. (2020) used MRF to find the 100-day repeatability CV to be $< 1\%$ for T1-array vials with reference times > 300 ms [40]. Similarly, for the same vials using a multi-dynamic multi-echo sequence (simultaneous T1, T2 and PD measurements), the 10-day repeatability CV was found to be $< 2.07\%$ [125].

The work presented in this chapter, in part, aims to extend on the literature and the work

completed within the previous chapters. Specifically, the accuracy and daily repeatability of anatomy-specific T1- and ADC-mapping sequences used on a 3T radiotherapy dedicated MRI scanner was investigated. Additionally, the feasibility of measuring the ADC at AT in the Diffusion phantom was examined. This was with the aim to develop a simplified anatomy-specific QA protocol for the department, and provide guidance for other sites desiring to setup this type of QA.

High frequency measurements can aid in estimating uncertainties in a departments' patient-based datasets, and also test a sequences' robustness prior to its implementation into clinical trials. This testing is especially important if the patient derived QIBs from the clinical trials are to be prospectively used to guide patient ART.

5.2 Methods

Both the NIST System and Diffusion phantoms were imaged daily (weekdays) at AT over one month using the same 3T MRI scanner described in chapters 3 and 4. Imaging was limited to 30-minutes per day due to a heavy clinical workload. An additional 15-minutes was permitted once weekly to rescan the Diffusion phantom at 0°C. In this study, all phantom setups, image acquisitions and analysis was completed by a single user (physicist with 3 years MRI experience).

5.2.1 Designing an Anatomy-Specific qMRI QA Protocol

The department generated a list of all T1- and ADC-mapping sequences used in their past and current clinical trials. Five T1-mapping (Table 5.1) and six ADC-mapping (Table 5.2) sequences were identified. This included imaging for head and neck (HN), prostate, rectum, cervical, and central nervous system (CNS) cancer patients.

All T1-mapping sequences were VIBE-based, with the addition of the Lung sequence being a STAR-VIBE to allow for free breathing (see section 2.3.4). As seen in Table

5.1, most of the sequences use the DFA method (with FA = 2° and 15°). Although the prostate-specific sequence had additional FA's to these, the timing limitations imposed for this study required a compromise of only using the two common FA's.

Table 5.1: Selected acquisition parameters utilized for all T1-mapping sequences used in the radiotherapy department. Note: HN = head and neck, CNS = central nervous system, FA = flip angle, TE = echo time, TR = repetition time, BW= bandwidth (Hertz/pixel) FOV = field of view, FE/PE/SE = frequency/phase/slice encoding respectively and Acq. Time = Acquisition time.

Anatomy:	**HN**	**Lung**	**Prostate**	**CNS**	**Rectum**
Sequence	VIBE	Star-VIBE	VIBE	VIBE	VIBE
SE Direction	Cor	Cor	Cor	Axial	Axial
FA (degrees)	2,15	2,15	2,5,10,15,20,30	2,15	2,15
TE (ms)	1.42	1.33	4.63	1.42	1.35
TR (ms)	4.09	7.64	4.63	4.0	4.09
BW (Hz/Px)	440	810	440	410	440
FOV (mm^2) (FE x PE)	220 x 220	340 x 340	250 x 203.3	200 x 200	220 x 220
# Slices	26	18	40	40	26
Slice Thickness (mm)	3.0	4.0	3.0	2.0	2.0
Acq. Time (∼min)	2.42	0.24	2.44	0.24	2.20

To also reduce the QA acquisition time, sequences that shared similar parameters were either removed from testing or were combined. For T1-mapping, only Lung, HN and CNS sequences were tested. For ADC-mapping, all RESOLVE sequences were combined, and similar for ZoomIt, as seen in Table 5.3. Additionally, the axial EPI ADC-mapping sequence assessed in Chapter 4 was also used. Additional information on each of the ADC-mapping sequences can be found in Section 2.3.1.

Table 5.2: Selected acquisition parameters utilized for all anatomic-specific ADC-mapping sequences used in the department. Note: SE = Slice-encoded direction, which then reflects the required orientation of the phantom for correct vial visualization of the ADC-maps. Also, the RESOLVE sequences have two associated echo times (TE); each of which correspond to the order of b-values listed above.

Anatomy:	**HN**	**Lung**	**Prostate**	**CNS**	**Rectum**	**Cervix**
Sequence	RESOLVE	ZoomIt	ZoomIt	RESOLVE	RESOLVE	RESOLVE
SE Direction	Axial	Axial	Axial	Axial	Axial	Axial
b-values (s/mm^2)	50,800	0,250,500,750	0,20,50,100,200,300,00,600,800	0,1000	50,800	50,800
TE (ms)	54,81	67	70	68,110	57,87	54,81
TR (ms)	5020	4000	2900	4000	4000	3730
BW (Hz/Px)	868	1792	1612	651	868	868
FOV (mm^2) (FE x PE)	260 x 260	300 x 166.8	190 x 106.4	220 x 220	220 x 220	220 x 220
# Slices	27	20	20	20	20	20
Slice Thickness (mm)	3.0	8.0	4.0	4.0	4.0	4.0
Acq. Time ($\sim$min)	4.18	3.18	5	4.1	5.5	3.56

Table 5.3: Selected acquisition parameters utilized in generalised ADC-mapping imaging protocol used in this study. The anatomy row indicated refers to the anatomy-specific regions that were combined from Table 5.2.

Anatomy:	**Standard (N/A)**	**Lung, Prostate**	**HN, CNS, Rectum, Cervix**
Sequence	EPI	ZoomIt	RESOLVE
SE Direction	Axial	Axial	Axial
b-values (s/mm^2)	0, 500, 900, 2000	0, 50, 250, 500, 800	50, 800
TE (ms)	106.0	80	66,108
TR (ms)	10,000	5900	4000
BW (Hz/Px)	1424	1786	651
FOV (mm^2) (FE x PE)	220 x 220	210 x 116	220 x 220
# Slices	25	25	20
Slice Thickness (mm)	4.0	4.0	4.0
Acquisition Time ($\sim$min)	2	6.19	5.18

5.2.2 Image Acquisition and Phantom Setup

The daily imaging protocol involved scanning both phantoms at AT. To minimize temperature fluctuations and ensure thermal equilibrium within the phantoms, they were stored in the MRI scanner room. The temperature was recorded pre- and post-scanning in each phantom using a NIST-traceable thermometer (as described in Chapter 3). The temperature for each day was recorded as the mean temperature measured between pre- and post-scanning of each phantom, and the average daily temperature ($\pm$ SD) was calculated using all daily temperatures recorded over the month.

Specifically, the Lung, HN and CNS T1-mapping sequences from Table 5.1 were used to image the T1-array of the System phantom. Also, the EPI, and generalised RESOLVE and ZoomIt ADC-mapping sequences in Table 5.3 were used to image the vials of the Diffusion phantom. For the ADC-mapping sequences, this study implemented the same shimming methods that were described in the prior chapter, specifically in Section 4.2.1.

To reduce setup time and also mitigate variability introduced by using different coils between sequences, an 18-channel Siemens body-coil was used for all acquisitions (Figure 5.1). This coil is commonly used for imaging the thorax, heart, abdomen and pelvis [147]. The external MRI lasers, of which are often used by an MR-simulator to mock a patients' radiotherapy position setup, were used to help align both the phantoms and body coil to isocenter.

For sequences with axial slice-encoding (SE) directions, the phantom orientations had to be rotated to properly image the phantoms' vials; similar to the methods described in Section 4.2.1. Without the presence of a head-coil in this study, stabilising foam was introduced to prevent phantom movement (see Figure 5.1). The coronal acquisitions for T1-mapping (HN and Lung) were instead able to use the flat bottom surface of the System phantom to achieve stability.

Figure 5.1: Phantom setup used for assessing the performance of the anatomy-specific T1- and ADC-mapping sequences. The figure above (left) shows the 18-channel body-coil mounted on top of the Diffusion phantom (axial position), using the aid of MR-safe patient feet supports. Further shown (top-right) is the stabilising foam that the phantoms were positioned onto in order to avoid any movement when the couch and phantom entered the bore. This foam was included in the commercial phantom packaging and had a circular cut in the centre (bottom right).

5.2.3 Image Analysis

The T1-mapping DFA images and inline derived ADC-maps were exported from the scanner console, and imported into an in-house developed Python script (see Appendix A). This script placed circular ROIs at manually identified central vial locations on the centre slice of the DFA images, and on the 3 central slices of the ADC maps. Each selected slice was manually inspected for large distortions and mitigated if necessary. The size of the ROIs utilized was manually specified, and was dependent on the sequence FOV and voxel size. The ROI diameter ranged between 0.8-1.0 cm for the T1-mapping sequences, and 1.0-2.25 cm for ADC-mapping sequences.

T1-time fitting methods for each sequence matched those described in Section 3.2.3, with the key difference of only using 2 FA's for the fit. This removed the ability to use the SD of the fit to estimate the error of each daily T1-time vial measurement. Instead, the T1-time was calculated on a pixel-by-pixel basis, and the error was estimated by calculating the SD between individual pixels T1-times within each ROI. ADC analysis instead used the average (and SD) of pixel values within the ROI, as directly obtained from the inline

ADC-maps.

The statistical analysis for T1-mapping was also completed using methods described in section 3.2.3. In particular, the long-term repeatability (%CV$_{LT}$ = %CV) was calculated using the difference between the daily T1-times measured. This is where long-term repeatability is defined as comparing QIBs between multiple days [20]. Also, the accuracy was calculated using the %bias between measured and reference vial T1-times [117].

For statistical analysis of ADC at AT, the value obtained for each vial could not be directly compared between daily measurements given the large temperature dependency known to exist [7, 21, 133]. Instead, the Diffusion phantom vial values, measured at a wide range of MRI bore temperatures (within the range of 16 - 22°C), was used to determine daily ADC reference values for each vial/temperature using a linear regression [108]. Each daily ADC measurement was then converted to a corresponding %bias, which was compared between days to calculate the %CV.

In this study, only physiologically relevant vials from each phantom were to be considered of interest. This includes vials 1-9 for the T1-array of the System Phantom [71, 106, 117, 118], and vials 1-11 in the Diffusion phantom at AT (approx. 19°C) [16, 50, 54, 60, 148–152]. Vial 2 for the System phantom and vial 1 for the Diffusion phantom were of particular interest in this study. This was due to their T1-time and ADC reference values closely reflecting the physiological values of the specific anatomies listed in Tables 5.1 and 5.2, respectively.

5.2.4 Temperature Dependency

One day per week following the AT acquisitions (each Tuesday), the ice-bath preparation methods described in Section 4.2.1 were performed on the Diffusion phantom. Once thermal equilibrium (0°C) was achieved, the phantom was re-scanned using the procedures described above for AT ADC. The ice-bath was then emptied from the phantom and filled

with tap water and returned to the MRI scanner room.

The statistical analysis methods employed for 0°C ADC measurements followed those described in Section 4.2.3. This included calculating the %bias, which was based on reference 0°C ADC values provided in the phantom manual, and also the weekly %CV between measurements. The weekly %CV was also calculated for the AT ADC measurements using the %bias recorded on the corresponding day of the week on which 0°C measurements were taken (each Tuesday). The weekly performance statistics, namely the %bias and %CV, measured for vial 1 at AT and 0°C were compared.

5.3 Results

A total of 23 daily T1- and ADC-mapping acquisitions were performed at AT over the month-long study. Additionally, 5 (weekly) ADC-mapping acquisitions were performed at 0°C. The phantom setup and image acquisition time took approximately 10 minutes for T1-mapping per day, and 15 minutes for ADC-mapping.

CNS T1-mapping required an additional physical phantom rotation and thus had the longest associated setup time. Additional time (approximately 5 minutes) was spent setting up the imaging protocol on the MRI console, removing the morning QA device from the couch, measuring the phantoms temperatures and also packing up the setup.

5.3.1 Temperature Dependencies

The average temperature measured in the System phantom was 19.39 $\pm$ 0.40°C. Temperatures remained relatively stable, excluding day 16 (temperature = 21.20°C) when the phantom was left outside the scanner room overnight.

Correlation coefficients (ρ) were calculated between the recorded daily temperature and measured T1-time for each sequence. There was no significant correlation when incor-

porating all vials in physiological range ($\rho < 0.01$). Considering vial 2 independently however, weak correlations (and coefficients of determination - R^2) of ρ = 0.6 (0.3), ρ = 0.4 (0.2) and ρ = 0.4 (0.1) were found for Lung, HN and CNS sequences respectively.

The AT recorded in the Diffusion phantom for each daily measurement remained within 0.5°C of the average temperature of 19.00 $\pm$ 0.16 °C. Interestingly, the phantoms average temperature on the day following ice-bath preparation (each Wednesday) was 18.73 $\pm$ 0.16°C, which was at least 0.3°C lower than the average temperature of any other day of the week. However for the weekly Diffusion phantom ice-bath measurements, the temperature was very stable at 0.00 $\pm$ 0.02°C.

There was no significant ρ between temperature and ADC when incorporating all vials ($\rho < 0.01$). Considering vial 1 individually however, the ρ (and R^2) was 0.3 (0.1), 0.8 (0.6) and 0.9 (0.9) for EPI, RESOLVE and ZoomIt sequences, respectively. This correlation reduced with increasing %PVP concentration, and was removed when the ADC value was converted to a %bias.

5.3.2 0°C vs AT ADC

On average, there was an approximate 3-hour gap between the AT and 0°C acquisitions performed on the same day. Typical ADC maps for each sequence and temperature are shown in Figure 5.2. Also, the approximate NIST reference values used to calculate the %bias for each vial can be found in Table 5.4.

Table 5.4: NIST reference ADC values ($\mu m^2/ms$) for the same batch of PVP as those embedded in the phantom's vials at: 0°C [108], and at 19°C (average daily AT temperature). Results underlined imply that the ADC reference values were outside of the expected physiological range, and were thus not considered to be of interest in this study. *It should be noted that the AT reference ADC values varied depending on the daily recorded temperature.

	Vials:	**1-3**	**4-5**	**6-7**	**8-9**	**10-11**	**12-13**
Temperature	**PVP (%):**	**0**	**10**	**20**	**30**	**40**	**50**
0°C	**NIST ADC Reference Value**	1.109	0.817	0.579	0.380	0.220	0.110
19°C	**NIST ADC Reference Value***	1.949	1.507	1.148	0.796	0.489	0.257

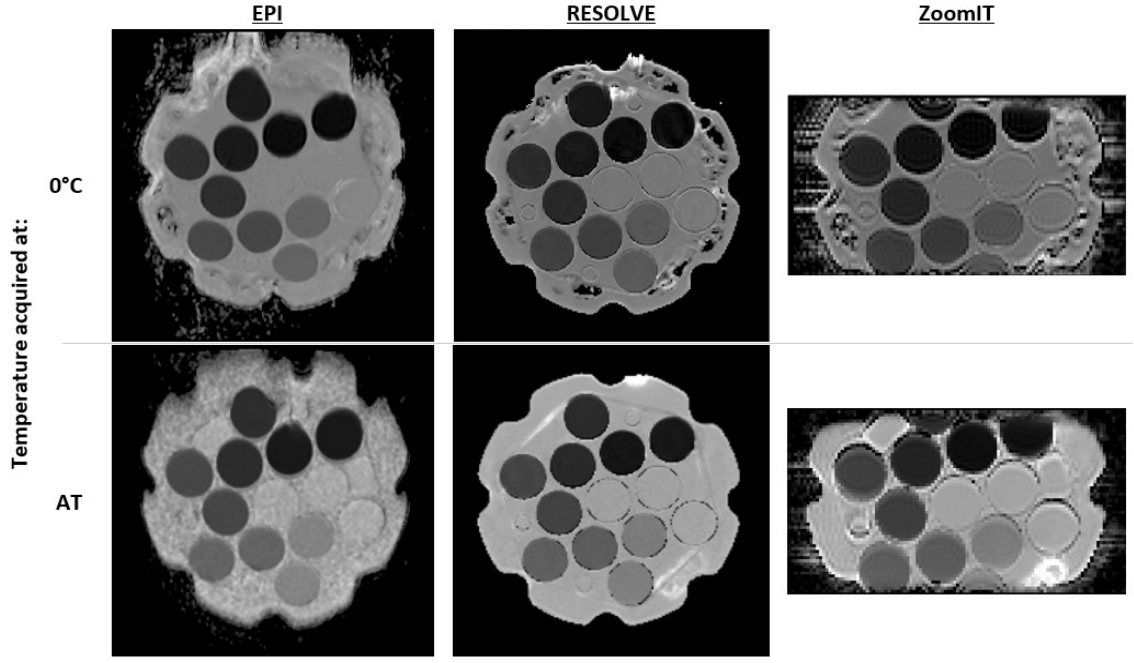

Figure 5.2: Typical inline ADC-maps of the Diffusion phantom for each ADC sequence and temperature tested. Specifically, these ADC maps were captured on the first day of the 0°C acquisitions (top), which corresponded to day 2 of the AT acquisitions (bottom). Note: Vials 5 and 11 were outside of the ZoomIt FOV, and were excluded from any statistical analysis for this sequence type. Corresponding vial numbers for this phantom in the same axial position can be observed in Figure 2.10.

Variations observed in the weekly %bias recorded for vial 1 can be seen in Figure 5.3 for both 0°C and AT acquisitions. Here, it is evident that 0°C and AT ADC was weakly correlated for EPI (ρ=-0.6), highly correlated for RESOLVE (ρ=0.99), whilst ZoomIt had no correlation (ρ=0.02). Further, the absolute maximum difference in vial 1 %bias found

between the weekly ADC measurements for EPI, RESOLVE and ZoomIt sequences respectively was 1.21%, 0.66%, and 0.56% at 0°C, and 3.44%, 2.62%,0.29% at AT.

The median weekly %bias and %CV measured for vial 1 over all sequences and temperatures is reported in Table 5.5. Also included in this table is the overall differences between median %bias and %CV between 0°C and AT measurements. It was found that the magnitude of differences in %bias between 0°C and AT measurements, and also the difference between weekly measurements at the same temperature, was less than the Profile %bias tolerance (|%bias <3.6%|) [20].

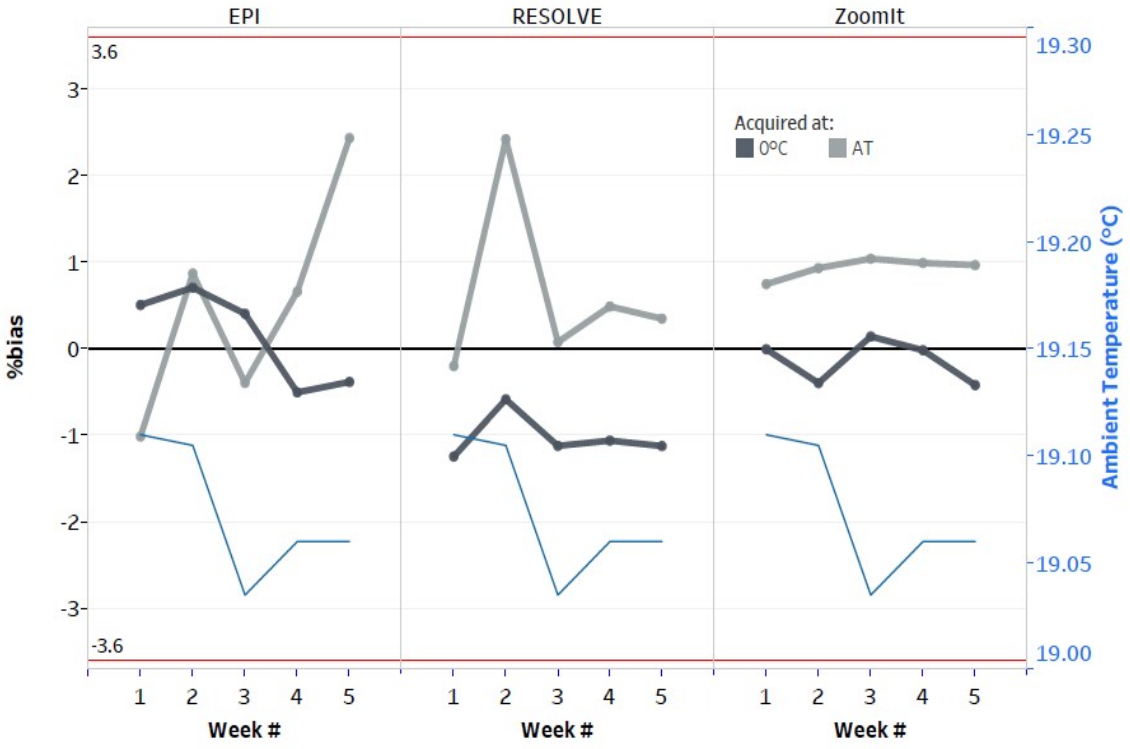

Figure 5.3: Variations in the ADC %bias for the 0% PVP vial (at isocenter) in the Diffusion phantom when at 0°C and Ambient temperature (AT). Each data point represents measurements acquired on the same day of each week (each Tuesday). Overlaid is the AT (blue) measured in the Diffusion phantom for each respective day, and also the Diffusion Profile bias tolerances of ±3.6% (red) [20]. It should be noted that the %bias for AT measurements have accounted for temperature deviations recorded for each day.

Table 5.5: Summary of the weekly median anatomic-specific ADC sequences accuracy (%bias) and repeatability (%CV), as measured at 0°C and ambient temperature (AT). Values are stated as the median %bias of the 5 days of measurements and their respective inter-quartile ranges (IQR). Notice there is a reduction in the IQRs for RESOLVE and ZoomIt sequences compared to EPI. All results were within Diffusion Profile tolerances of bias $< \pm 3.6\%$, and CV $< 2.2\%$.

	%	EPI	RESOLVE	ZoomIt
ADC (0°C)	**bias**	+0.41 (0.89)	-1.12 (0.06)	-0.01 (0.39)
	CV	0.49	0.23	0.23
ADC (AT)	**bias**	+0.66 (1.26)	+0.35 (0.41)	+0.97 (0.06)
	CV	1.18	0.92	0.1
Difference (AT - 0°C)	**bias**	0.25	1.47	0.98
	CV	0.69	0.69	-0.13

Further, the CV was much less than the Profile tolerance of 2.2% [20] for all sequences and temperatures. Given all measures were well within Profile tolerances, AT measurements were considered sufficient to measure the sequence-specific ADC accuracy and repeatability for QA purposes. The remainder of this study will thus solely focus on the daily AT results.

5.3.3 Accuracy

The average T1-times measured over the monthly study for each sequence is presented in Table 5.6. A sample of typical FA images produced are included in Figure 5.4. The daily variability in vial 2 T1-time as observed for these sequences is shown in Figure 5.5, and the average %bias for each vial within physiological range is presented in Figure 5.6.

Table 5.6: Reference T1-time characterised by NIST (at 20°C) for each vial and that experimentally measured using the departments anatomic-specific sequences for Lung, HN and CNS. NIST values have been obtained from the phantom manual [117], whilst experimental values have been presented as an average of the daily measurements (with respective SD's). Only vials in the expected physiological range (1-9) of the System phantom have been presented.

Vial Number:		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
T1 (ms)	**Value**	1884.0	1330.2	987.3	690.1	485.0	341.6	240.9	175.0	121.1
NIST	**SD**	30.3	20.4	14.2	10.1	7.1	5.0	3.5	2.5	1.8
T1	**Value**	1508.8	1163.3	913.9	599.4	390.0	249.66	170.8	123.1	81.1
Lung (ms)	**SD**	33.6	31.1	21.3	38.7	11.7	168.3	31.7	27.2	3.4
T1	**Value**	1842.3	1448.5	1137.9	767.7	487.0	310.8	219.0	165.6	112.2
HN (ms)	**SD**	42.8	24.8	16.9	11.3	8.3	6.5	4.3	3.2	2.3
T1	**Value**	2949.5	2135.8	1167.3	695.7	502.1	382.5	255.9	169.5	127.3
CNS (ms)	**SD**	81.7	193.9	97.1	31.0	22.0	15.2	12.0	6.9	11.9

From Figure 5.6, it can be seen that the magnitude of %bias varied as a function of reference vial T1-time. Further, it is apparent that Lung underestimated, and CNS largely overestimated, the T1-time relative to the reference T1-time. HN T1-mapping however was mostly stable over the wide-range samples tested within the phantom. Quantitatively, Table 5.7 presents the median %bias (and inter quartile ranges - IQR) calculated over all vials (1-9) for each sequence, and also the statistics for vial 2 individually.

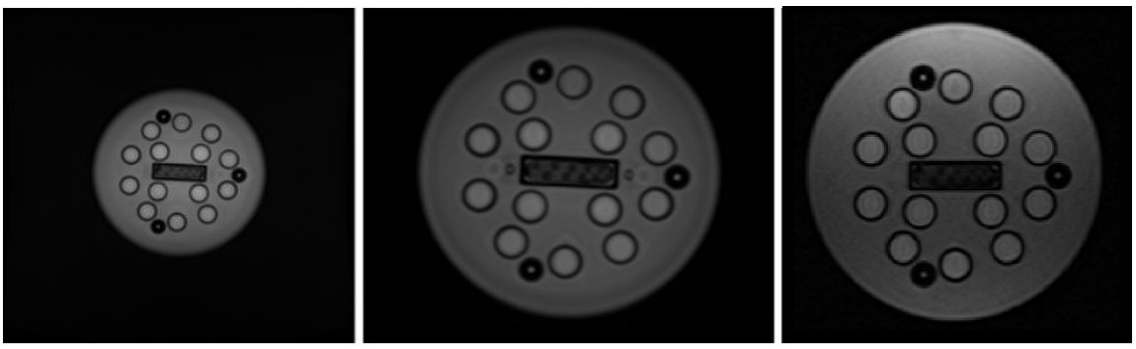

Figure 5.4: Images of the central slice of the T1-array of the System phantom, using a FA of 2°C. Images were produced using the Lung, HN and CNS T1-mapping sequences (ordered left to right). Notice the differences in FOV implemented for Lung relative to HN and CNS. This resulted in smaller regions of interest (ROIs) able to be used in Lung image analysis.

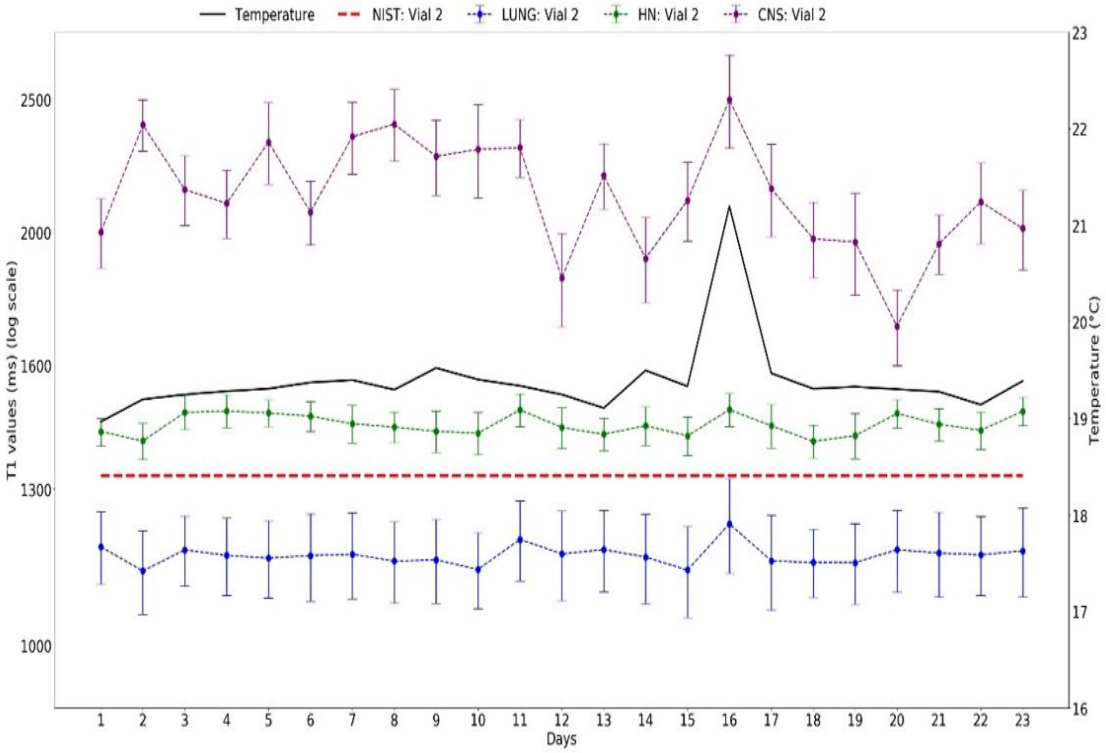

Figure 5.5: Monthly fluctuations observed in T1-time measurements for vial 2, with overlaid average temperature readings. Error bars were generated from the standard deviation of T1-time within the each vial for each day. The temperature and subsequent T1-time deviation observed to occur on day 16 was a result of the phantom being left outside the scanner room during the night prior to this measurement.

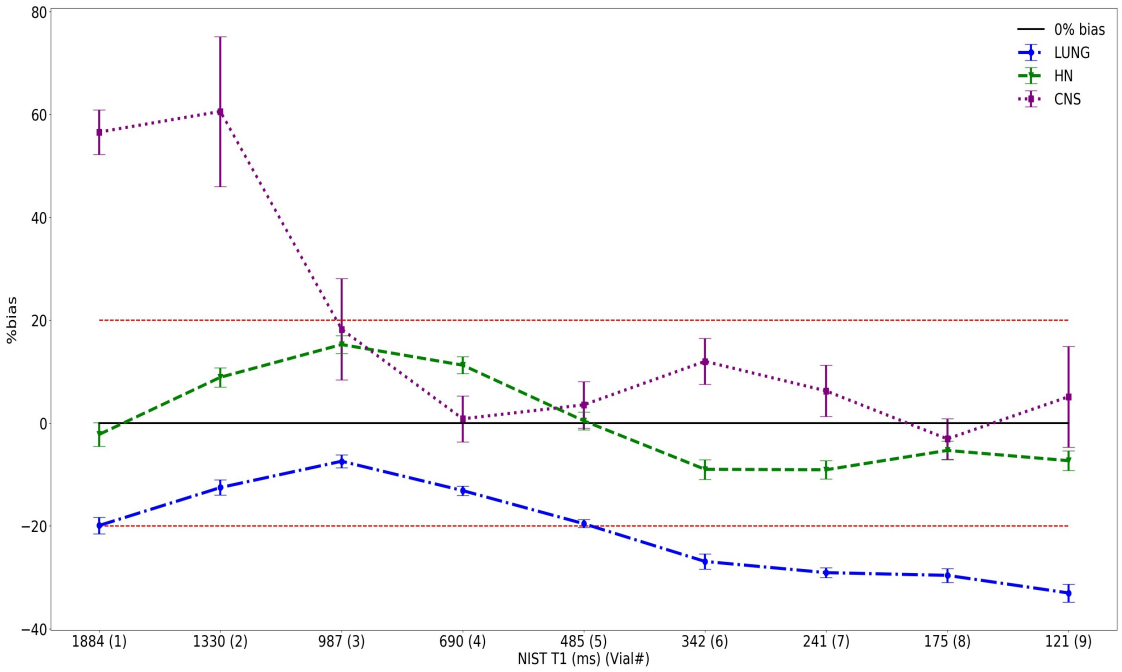

Figure 5.6: Average fluctuations observed in T1-time %bias measurements for all vials within physiological range. Error bars were generated from the standard deviation of each vial (calculated from daily measurements), and ± 20% tolerance bias lines have been added.

Similarly, the ADC %bias results for all vials of interest can be observed in Figure 5.7. In general, the %bias experienced an upward trend for vials with higher concentrations of PVP (lower diffusivities). This particularly can be observed for vials 6-11, with PVP concentrations ranging from 20-40%. Overall, all sequences overestimated the ADC over the 23 daily measurements, with RESOLVE having the best accuracy, followed by ZoomIt and then EPI. Quantitatively, the median %bias (and IQR) for EPI, RESOLVE and ZoomIt over all vials, and also for vial 1 individually, is summarised in Table 5.7.

Table 5.7: Summary of the anatomic-specific sequences accuracy (%bias) and repeatability (%CV) measured for the 6 sequences investigated at 3T. All vials considered within respective physiological ranges for T1-time and ADC were reported, along with vial 2 individually for T1-time and vial 1 for ADC. Values are stated as the median of all months/vials' acquisitions and their respective inter-quartile ranges (IQR). *Note: ZoomIt results exclude vials 5 and 11.

		%	Lung	HN	CNS
T1	**Vials:1-9**	**bias**	-19.44 (16.1)	-2.44 (16.54)	+7.8 (27)
		CV	1.69 (0.64)	1.92 (0.32)	4.46 (4.25)
	Vial: 2	**bias**	-12.59 (1.54)	+8.66 (3.51)	+58.62 (23.49)
		CV	1.69	1.72	9.08
		%	**EPI**	**RESOLVE**	**ZoomIt**
ADC	**Vials:1-11***	**bias**	+2.22 (2.41)	+1.52 (2.34)	+1.632 (1.73)
		CV	0.59 (0.19)	0.43 (0.17)	0.62 (0.42)
	Vial: 1	**bias**	+0.25 (1.13)	+0.19 (0.31)	+0.95 (0.13)
		CV	0.83	0.32	0.17

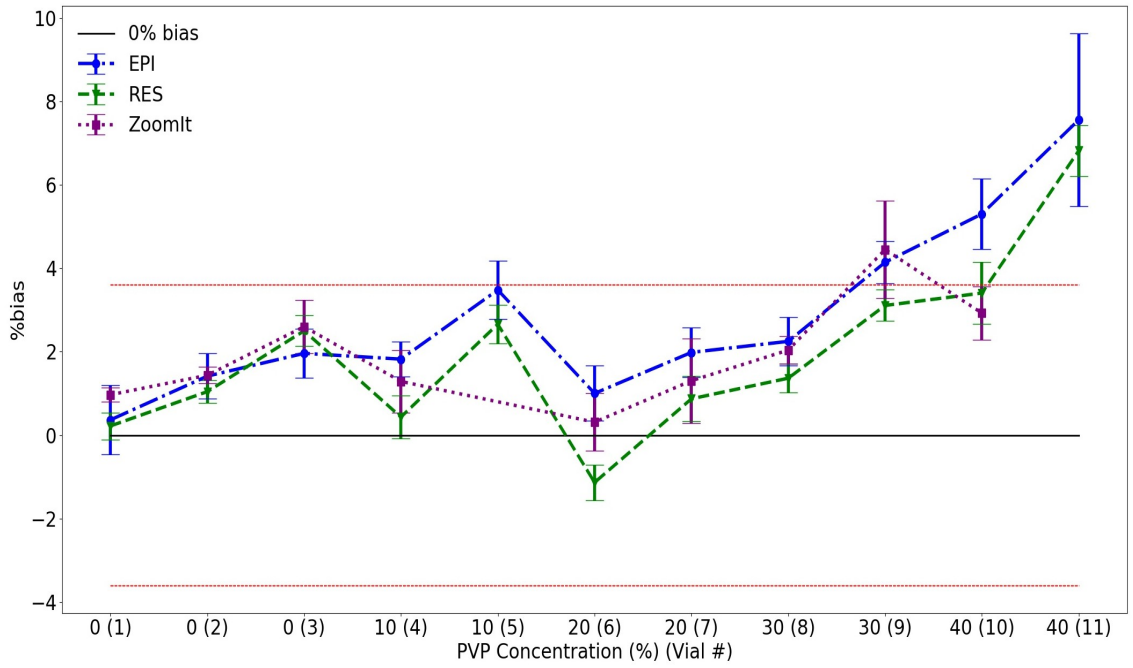

Figure 5.7: Average fluctuations observed in AT ADC %bias measurements for vials within physiological range. Error bars were generated from the standard deviation of each vial (calculated from daily measurements). Note: Vials 5 and 11 were not able to be included in ZoomIt results due to the reduced FOV in the sequence. Profile tolerance %bias limits of ±3.6% has also been added.

5.3.4 Linearity

Strong linearity was found for both T1-time and AT ADC measurements. Specifically, the coefficient of determination (R^2) was > 0.977 for all sequences which can be observed in the correlation plots in Figure 5.8 and 5.9. CNS and Lung sequences had the largest deviations at the highest and lowest reference T1-time vials, respectively.

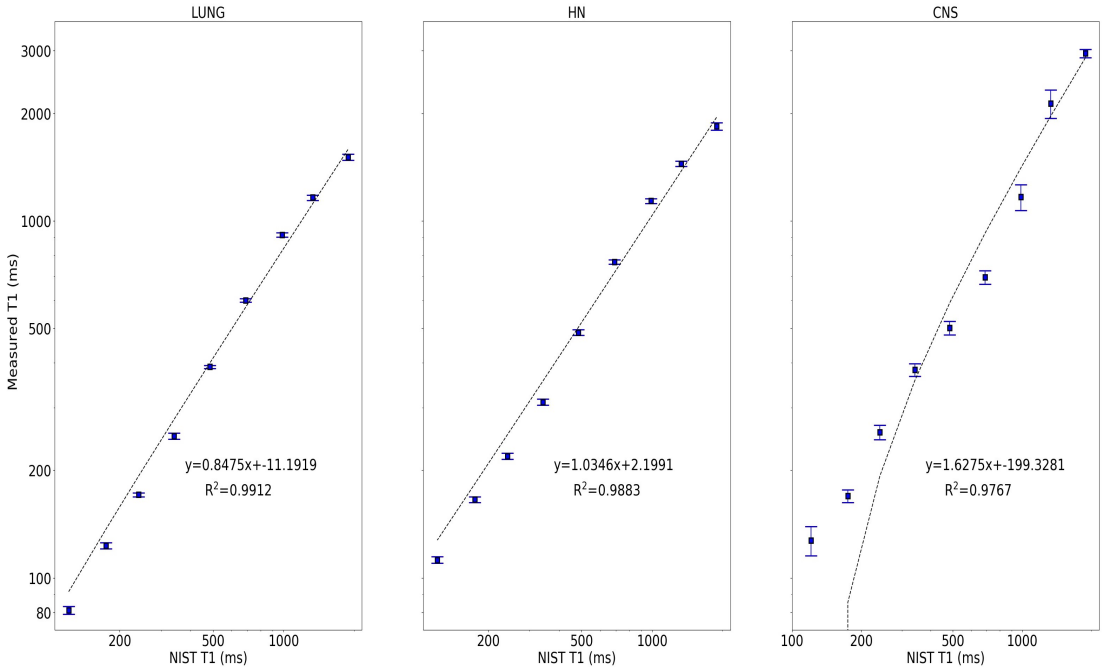

Figure 5.8: A strong linear correlation was found between physiologically relevant and measured T1-times for all anatomic-specific sequences. The highest coefficients of determination (R^2) were found for Lung, HN and then CNS sequences, respectively. Note: All axes have employed a logarithmic scale and the negative y-intercept of T1-time for Lung and CNS is unrealistic; caused by the large outliers recorded for certain vials.

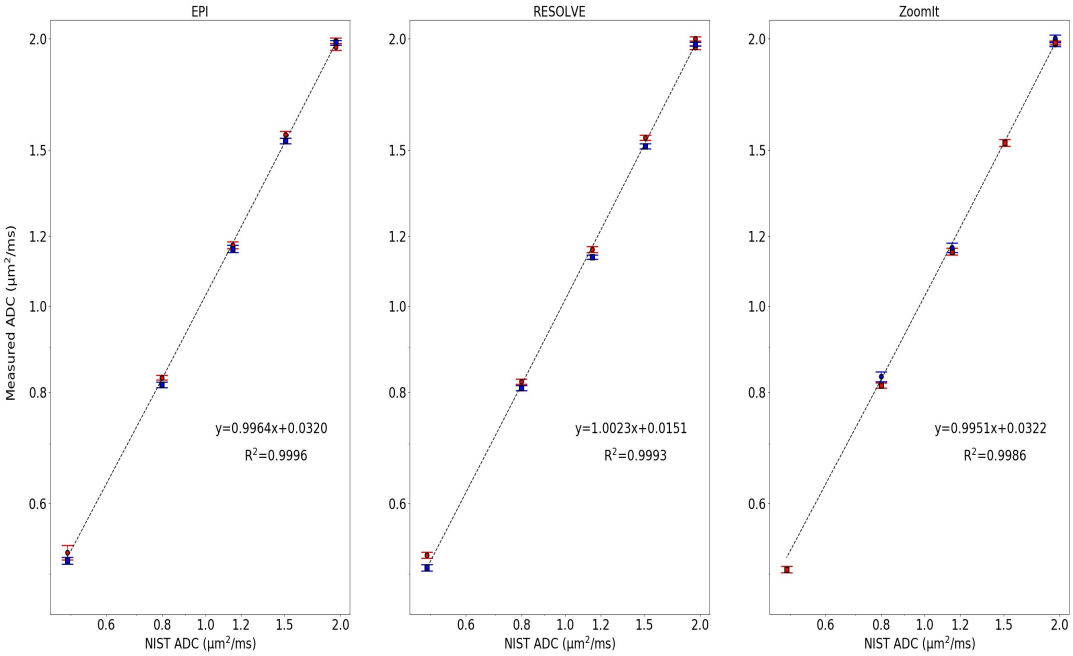

Figure 5.9: A strong linear correlation was again found between reference and measured AT ADC for all anatomic-specific sequences. A logarithmic scale in the axis was employed, and the NIST ADC reference values utilized were taken from Table 5.4. It can be observed that vials 5 and 11 were excluded from this plot for ZoomIt.

5.3.5 Repeatability

The %CV of the T1-time was relatively stable over all vials within physiological range for Lung and HN sequences (CV $<$ 2.6%), whilst CNS was quite variable (see Figure 5.10). Thus, CNS recorded the worst repeatability over all vials and also when considering vial 2 individually, as seen in Table 5.7.

The CV calculated over the 23 days for each ADC-mapping sequence was well within the long-term Profile tolerance limit of 2.2% (Figure 5.11) [20]. Table 5.7 highlights that this tolerance was met when considering the average of all physiologically relevant vials, and also when considering vial 1 individually.

Overall, RESOLVE had the best repeatability over the wider vial range, followed by EPI and then ZoomIt (even with vials 5 and 11 removed for ZoomIt). For a more standardized comparison, ZoomIt performed best when only considering vial 1 at isocenter; aligning

with results found in Section 5.3.2.

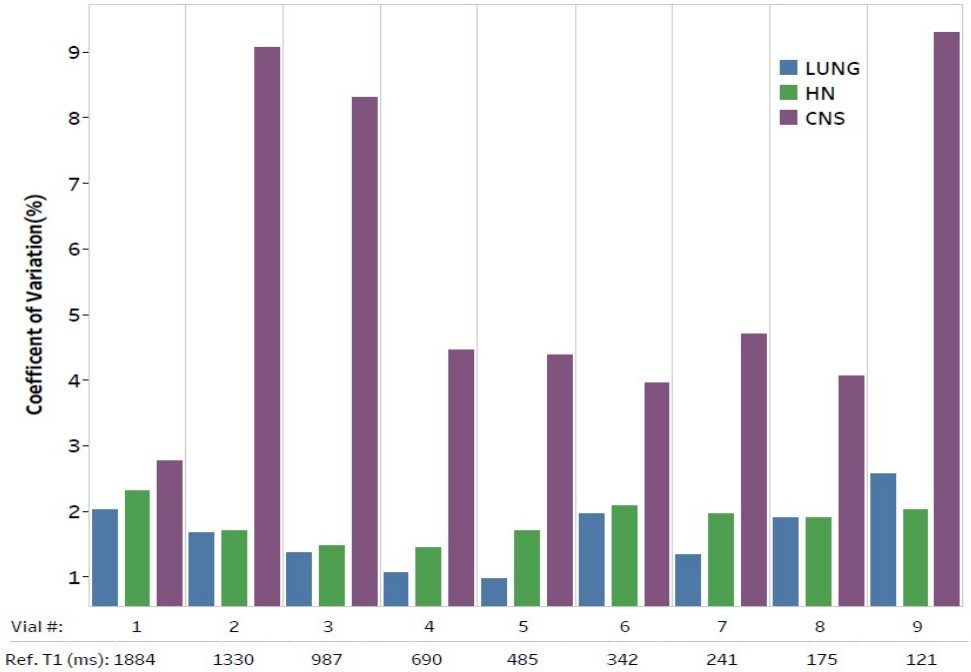

Figure 5.10: Coefficient of Variation (%CV) calculated for each vial in the physiological range of the System phantom from all daily acquisitions.

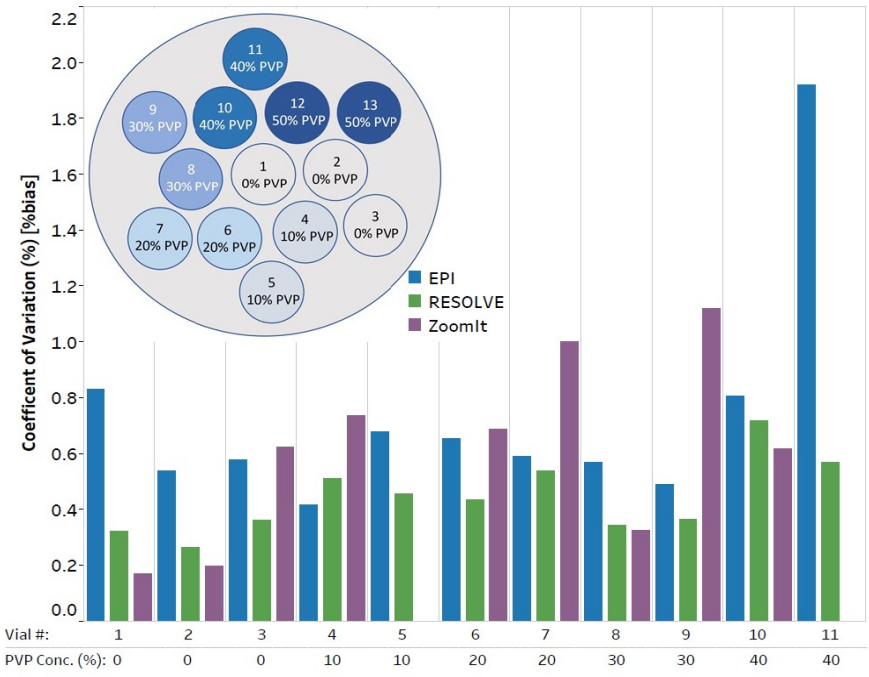

Figure 5.11: ADC coefficient of variation (%CV) calculated for each physiologically relevant vial at AT in the Diffusion phantom from all daily acquisitions. Note: Vials 5 and 11 were again excluded from analysis due to the ZoomIt FOV. Further, a much smaller CV scale has been employed in this plot relative to the T1-mapping results in Figure 5.10.

5.4 Discussion

QIBs such as T1-time and ADC have potential to be used either directly or indirectly in radiation oncology for disease detection and/or monitoring and assessing patient treatment responses [3, 16, 21, 24, 25]. However, unknowns in the QIBs values due to fluctuations in scanner performance make it difficult to conclude if changes in a patients QIB is actually due to a physiological response [153].

Using QA methods such as those described in this chapter allow departments to quantify this uncertainty for the anatomy-specific sequences used in their clinical trials. This QA protocol was designed to be easily and efficiently implemented into routine clinical QA. In this study, the protocol was tested on a radiotherapy dedicated scanner to assess the daily repeatability and accuracy of T1-time and ADC (at AT).

5.4.1 Temperature Dependencies

Temperature dependencies were an important aspect to this study. For AT ADC, the significant vial 1 ADC-temperature correlations (of up to ρ =0.9 and $R^2 = 0.9$) were removed by converting the ADC value to a %bias using temperature-corrected reference values. This correction method would have accounted for the temperature deviations found to occur on the day (around 18-hours) after ice-bath preparation, ultimately mitigating any potential effects on ADC %CV calculations. However, finding this particular deviation suggested that the time required for the phantom to return from 0°C to a thermal equilibrium state within the bore room may have been underestimated. This should be kept in mind for future investigations.

For T1-mapping, only a weak correlation was found when independently considering vial 2 ($\rho < 0.6$ and $R^2 < 0.3$). This was likely impacted by the major temperature variation recorded in the System phantom on day 16 (Figure 5.5). This temperature measurement in particular was an outlier ($>$ 2*SD above the mean phantom temperature) and could

have affected the statistics derived. One improvement to this study would be to better maintain external System phantom conditions (i.e., consistent storage location).

Alternatively, the measured T1-times could be converted to a temperature-dependant %bias using the new phantom reference T1-times provided over a range of temperatures [154]; similar to the post-processing analysis performed for AT ADC. This would allow T1-time repeatability measurements to be independent of large external temperature fluctuations. However, comparing daily measurements to only one set of NIST reference T1-times (i.e., characterized at 3T and 20°C) was implemented in this study to maintain consistency with the methods used in the literature [19, 40, 112, 125] and in Chapter 3.

Similar data modelling methods have been used in the past to calculate the ADC at different temperatures for various PVP solutions [155–157]. However, there are no past studies the author is aware of that have reported on comparing these measures to standardized ADC at 0°C in terms of accuracy and repeatability. This is especially the case when using the Diffusion phantom.

Recently, novel image processing software packages have become available which automate temperature-corrections for T1- and ADC-mapping; requiring only a recorded user input temperature [158, 159]. Additionally to improve measurement and analysis efficiencies, it is possible to integrate an MRI visible thermometer into a QA phantom to enable temperatures to be determined during image post-processing [158, 160]. This temperature value could also be automatically extracted using a designed image analysis software [158].

5.4.2 0°C vs AT ADC

The aim of this part of the study was to compare the performance of the same sequences when acquired at two different temperatures: 0°C and at AT. To do this, a novel approach was implemented where the Diffusion phantom was scanned on the same day of each

week over one month at both temperatures, and accuracy and repeatability statistics were compared.

Overall, the results outlined in Table 5.5 show comparable performance metrics between the 0°C and AT measurements, and all were within the Diffusion Profile conformance limits. Interestingly, the differences between weekly ADC measurements for the same temperature and sequence (ranging 0.29% to 3.44%), were of similar magnitudes to those observed between monthly EPI measurements in Chapter 4 (Figure 4.6).

Another key finding from Table 5.5 was that 0°C measurements were more reproducible for EPI and RESOLVE sequences, whilst ZoomIt was slightly more reproducible at AT. Considering all sequences performed within tolerance ranges, especially for ZoomIt, this finding was not considered of interest. Investigations outside the scope of this thesis would be required to determine the cause of this effect.

In comparison, most ADC reliability studies that exist in the literature use the recommendations by Padhani et al. (2009) and implement an ice-water bath to remove any thermal variability between ADC measurements [11, 21, 41, 63, 64, 109, 136, 161]. However, with the newly characterized Diffusion phantom vial values, which are known over a wide range of temperatures, it is now possible to measure the ADC at AT and temperature correct the ADC values during post-processing. This method can be advantageous for several reasons.

Firstly, at AT the Diffusion phantoms' reference vial values are higher and thus more closely reflect ADC's found in-vivo (Table 5.4) [16, 50, 54, 60, 148–152]. Secondly, clinical implementation of this type of qMRI QA in a department becomes simpler and more efficient. For example, reducing phantom preparation time for staff and minimizing the required resources which can't always be guaranteed like ice and a refrigerator [155]. Thirdly, removing the ice-bath minimises air-bubbles within the phantom and thus re-

duces the probability of susceptibility-induced distortions occurring (as seen in Figure 5.2).

One limitation to this part of the study was that only vial 1 results were compared between AT and 0°C measurements. This vial was selected due to water composition being expected in most anatomies planned for imaging. Further, its position at isocenter made it less likely to be affected by the gradient non-linearities and susceptibility distortions (which were discussed in Section 4.4). One improvement to this study would be to include all vials in the analysis. Additionally, if more time was available on the MRI scanner, the CV_{ST} and SNR could be compared by acquiring four repetitions of each sequence at each temperature (as per described in Section 4.2.3).

The findings from this part of the study, along with the advantages of implementing the AT method clinically, led to recommending the this method to departments for prospective ADC QA for clinical trials. If QA results were to become out of tolerance range, additional recommendations were given to repeat the QA using an ice-bath.

5.4.3 T1-Mapping: Accuracy and Daily Repeatability

Recall from Section 2.3.4 that obtaining patient-specific and accurate T1-maps are essential to generate reliable PK-parameters using DCE-MRI. In this study, the HN and Lung T1-mapping sequences were found to perform adequately when considering both the measurements acquired over all vials in physiologic range, and when considering vial 2 independently: median bias $< 19.44\%$ and CV $< 2\%$. In comparison, the CNS sequence continually overestimated the T1-time and the measurements had poor repeatability (see Table 5.7 and Figure 5.10). For example, the bias and CV in vial 2 was found to be +58.62% and 9.08%, respectively .

The performance of the T1-mapping sequences for vial 2 was of particular interest. This was due to the reference T1-time in the vial closely reflecting physiological values found

in the literature for the anatomies of interest. That is, longer reference T1-times (vials 1-4) specifically covered literature reported in-vivo T1-times for lung [162, 163], CNS/HN (e.g., muscle and cartridge, white and grey matter) [71, 164], and prostate tissue [71].

Finding large inaccuracies for longer T1-time ($>$ 1330 ms) vials matches other phantom-based findings in the literature [19, 40, 125, 145]. This could be attributed to insufficient recovery of the vials longitudinal magnetization when sequences employ very short TRs [19], and should be a factor considered for future sequence development. Substantial inaccuracies were also found for the Lung sequence for shorter T1-time vials (magnitude $>$ 26% - See Figure 5.6). Similar results have been reported in the literature [110, 112] and match the findings in Chapter 3.

According to the literature, sequences which generate T1-inaccuracies with magnitudes $>$ $\pm$15-20% can cause detrimental errors to the resultant PK-parameters, and are not recommended for clinical applications [19, 68, 118]. CNS results found for measurement %bias to exceeded this limit, and generated large %CVs and an unrealistic y-axis crossing for the linearity investigation (Figure 5.8).

For these reasons, the authors made recommendations to the department to find an alternative imaging sequence for CNS T1-mapping. This is especially important for imaging the tissues within the CNS region which have high T1-times (e.g, cerebral spinal fluid, grey and white matter) [71]. Other factors known to affect T1-mapping, such as B1-inhomogeneities, were not investigated as it was out of the scope of this thesis [19, 67, 126].

5.4.4 ADC-Mapping: Accuracy and Daily Repeatability

The average ADC accuracy and repeatability calculated for all sequences and vials was considered sufficient, as per the Diffusion Profile [20]. This included finding the bias and CV to be $<$ 2.22% and $<$ 0.62% over all vials in physiological range respectively, and $<$

0.95% and $<$ 0.83% for vial 1 individually. These performance metrics are comparable to the results found in Chapter 4 (see Figure 4.4), and also to the results of past phantom-based investigations [11, 41, 45, 107, 124, 133].

Additionally, matching the findings from Chapter 4, an upward trend in ADC %bias was observed for vials with higher PVP concentrations (lower reference ADC values). This could be attributed to the same factors noted in Section 4.4, such as an increased likelihood of susceptibility-induced distortions or insufficient SNR for these highly concentrated vials [107]. This observed trend of an increase in %bias should be considered when evaluating the ADC of anatomies with expected lower ADC values.

In this study, RESOLVE and ZoomIt sequences were found to have superior performance to EPI (Table 5.7). This is a common finding to past phantom and in-vivo investigations [45, 165, 166], and was expected given their enhanced sequence design features. The accuracy of ADC values measured using different DWI sequences can be influenced by a number of factors, as discussed in Section 2.3.2.

Specifically, single-shot EPI (SS-EPI) sequences, such as those used in this study, are known to be venerable to susceptibility-induced distortions and signal loss [45, 63, 145, 167]. Susceptibility distortions are also known to increase with higher field strength, and thus were expected to occur in this study using a 3T MRI scanner [63]. These were mainly observed at the vial-water interface of the highly concentrated PVP vials in both the work presented in this chapter and in Chapter 4.

Whilst it would have been possible to further optimize this SS-EPI sequence to reduce distortions and enhance performance metrics, such optimization was not completed. This was to maintain acquisition parameters with those utilized in Chapter 4 and the QIBA Profile guidelines [20].

In contrast, RESOLVE uses a multi-shot EPI approach, which helps reduce the acquisition time by acquiring only a portion of the k-space data per shot. This technique aids to reduce susceptibility distortions and blurring effects (see Section 2.3.1); offering superior performance over the standard EPI sequence, as the results of this study highlight [46–48].

As noted in Section 2.3.1, ZoomIt is a sequence that accelerates the acquisition time by only 'zooming into' specific regions in the image and enhances B1-homogeneity through the use of a parallel transmission method. This approach involves the simultaneous transmission of RF pulses across multiple RF channels [6,45,166,168]. However, this method has been observed to cause variations in measured ADC values for certain anatomies and b-values [168]. This could explain the observed higher repeatability for vial 1 at AT, but lower accuracy (bias), when compared to EPI and RESOLVE (Table 5.7). While each of these factors for the different sequences can contribute to inaccuracies in the ADC values measured in this study, a further investigation would be needed to determine the individual impact of each factor on ADC quantification.

Differing to prior investigations, the ADC was measured at AT with the %bias calculated via temperature correcting the Diffusion phantom reference ADC values. This allowed for measuring a wide-range of ADC samples which more closely reflect those expected in-vivo; important when imaging diverse anatomies. One limitation to this part of the study was that only inline ADC-maps were used for analysis. This method is not recommended for multi-site investigations due to potential analysis software dependencies [64]. An improvement would be to use an offline-fitting approach like that employed in Chapter 4. However, inline maps were used as per departmental request to follow ADC analysis protocols used in their single-site clinical trial practices.

Timing is also an important consideration when providing recommendations to departments for sequence utilization. For example, RESOLVE and ZoomIt took over 2.5 times the acquisition time required for EPI (Table 5.3) [45]. This would result in longer ac-

quisition times for patients and also for the QA. However, the largely improved performance metrics for these sequences outweighed this factor, and thus all anatomy-specific sequences were recommended for continued use in clinical trials given QA protocols are followed.

5.4.5 General Findings and Limitations

Timing considerations were also of concern when designing the QA protocol used in this study (i.e., 30 minutes MRI scanner time per day). This led to compromises in phantom setup (e.g., using an 18-channel body coil for all sequences), and merging acquisition parameters between different sequences. Such changes from the clinical protocol could have impacted the results found.

One parameter compromise involved reducing the FOV of ZoomIt in the PE (superior to inferior) direction which removed visualisation of vials 5 and 11 (Figure 5.4). Compromising on the FOV is common when using ZoomIt [45, 124]. It was also considered acceptable for intended Lung and prostate imaging which is generally more focused on measuring wider (left to right) FOVs. Excluding the higher-concentrated and outer-ring PVP vials however, could have impacted the comparison between different sequences' overall accuracy, repeatability and linearity.

Signal saturation, an effect shown to affect measured T1-times (majorly) and ADC (minimally) in the previous two chapters, was not observed to occur in this study. This could be attributed to the T1-mapping sequences using lower magnitude FAs (see Table 5.1), or that the RESOLVE ADC-mapping sequence used higher b-values (> 0 s/mm^2 - see Table 5.3). However, this could not explain why saturation was not observed to occur for the EPI and ZoomIt ADC-mapping sequences.

To further investigate this occurrence, the individual DW EPI-DWI images acquired in this study and for Chapter 4 were compared. Findings included a large reduction in signal

intensity (approximately by one-third) of the images acquired for this study. Although minimally affecting the ADC, this would reduce the probability of saturation occurring in the b-value = 0 s/mm^2 EPI and ZoomIt DW images. This was likely due to a change in coil selection, however further investigation would be required to confirm this hypothesis.

In this study, there was no scanner hardware or software upgrades over the month-long duration. Also, potentially large sources of variability were removed by utilizing a single user for each phantom set-up, and by also using a consistent and semi-automated image analysis method. Improvements to this QA method for implementation into a large-scale clinical trial setting would be to use a fully automated analysis software, and to design and use a phantom-positioning apparatus; fitted to the particular MR couch to minimise phantom position variations.

The QA protocol developed for this study was designed to be simple, efficient and transferable to other departments. This QA would largely compliment the recommended baseline testing specified in the previous chapters. It is also especially important for departments carrying out single-site clinical trials involving QIBs; enabling a verification method of their qMRI sequences accuracy and repeatability.

This type of qMRI QA should be performed for each sequence planned for use: both at the start and end of any clinical trial, and alongside the annual qMRI QA of the scanner. If any system upgrades were to occur over the duration of the trial, then the QA should also be repeated directly afterwards to ensure no major changes to the scanner performance.

5.5 Conclusion

Anatomy-specific sequence qMRI QA is essential if the QIBs derived from these sequences are to be used for subsequent treatment response monitoring or outcome predictions. Thus, the QA procedures developed in this study were designed to be easily implemented into routine clinical QA in a department.

Using AT ADC measurements for QA purposes was recommended due to finding both the AT and 0°C measurements to be comparable and within Diffusion Profile tolerances (median vial 1 $|\text{bias}| < 1.12\%$ and $\text{CV} < 1.18\%$). This was along side the logistical benefits of scanning the phantom at AT in a clinical department, including a more simple and efficient QA setup. EPI, RESOLVE and ZoomIt sequences were all recommended for continued use in the department with the AT median %bias and %CV measured over all vials in physiological range being $< 2.22\%$ and $< 0.62\%$, respectively.

Similarly, Lung and HN T1-mapping sequences were recommended for continued use in the department with findings of adequate sequence performance in terms of their accuracy ($|\text{bias}| < 19.44\%$) and repeatability ($\text{CV} < 2\%$). However, the CNS T1-mapping sequence was found to have substantial performance variability, and thus resulted in recommendations to the department of investigating alternative T1-mapping sequences for this anatomy.

It was recommended that T1-time and AT ADC QA should be performed annually, both before and after a clinical trial, and around the time of any system upgrades occurring throughout the trial. Recommendations for further improving qMRI QA efficiencies involved designing a phantom-positioning apparatus and using a fully-automated analysis software for image processing.

Chapter 6

Discussion and Conclusions

6.1 General Discussion

In recent years, qMRI techniques have shown promise to assist and enhance the radiotherapy workflow. This stems from the techniques' unique ability to measure physiological changes in tissues using non-invasive and non-radiative methods [2, 3, 11–16]. However, such techniques have current limitations imposed which restricts their ability to be used in the clinic or in clinical trials. This includes the unknown accuracy and repeatability of qMRI derived QIBs on clinical MRI scanners [7, 37]. This results from many of the studies in the literature often not completing (or failing to report on implementing) any type of routine qMRI QA to assess these important reliability metrics, prior to collecting evidence for the clinical utility of these techniques.

Historically, the lack of conducting qMRI QA or technical performance testing could be attributed to the minimal resources that were available for departments to use and complete this type of testing [3, 7, 19, 37]. This would have been enhanced by the dearth of studies completed which demonstrate variability in qMRI MRI scanner performance and thus the need for such QA. However, in recent years, developments have occurred in this space. This includes the manufacturing of dedicated qMRI phantoms [100, 104, 106–108] and qMRI testing Profiles [20, 101].

This thesis serves to aid clinical departments to assess the accuracy and repeatability of qMRI QIBs through reporting on the development of both accurate and efficient QA procedures. Following implementation, it would allow a department to have increased assurance on the reliability of the QIBs generated from their MRI scanners. This is essential for collecting QIB-based patient datasets on an MRI scanner and serves as evidence for the clinical implementation of a QIB [18]. This includes using measured changes in a tissue's QIB values to guide the modification of a treatment plan in ART.

This thesis also highlighted the need for this type of QA through measuring qMRI performance fluctuations in a 3T radiotherapy dedicated MRI scanner. Specifically, the qMRI QIBs investigated were the T1- and T2-relaxation times and the ADC. Evidence was gathered through studies that monitored the long-term performance of the scanner using monthly interval imaging over one year; simulating the potential imaging frequency required for long-term patient response monitoring following treatment. Further, short-term monitoring was conducted using daily interval imaging over one month; simulating inter-fraction patient imaging over the course of radiotherapy treatment.

6.1.1 Research Aims

This thesis presents a body of work focused on assessing the scanner-specific uncertainty of qMRI for the implementation of QIBs in the clinic. The details below summarise the key results of this work as a whole, stated in terms of the aims and set research questions.

Aim 1: Quantify the longitudinal intra-scanner accuracy and repeatability of T1 and T2 relaxation times.

For qMRI QIBs to have both wide-spread clinical applications and be comparable between MRI scanners in multi-centre trials, the values being derived must be accurate, reproducible, and independent of scanner performance [7, 37]. Chapter 3 reported on a method that was used to investigate and determine the baseline accuracy and repeata-

bility of T1- and T2-relaxation times on a dedicated radiotherapy MRI scanner. This included conducting monthly imaging acquisitions using a System phantom and standardized acquisition sequences, over one year. T1-time was measured using both IR and VFA sequences, whilst T2-time was measured using a T2-SE sequence. Accuracy and repeatability were then mainly assessed on vials with relaxation times expected to be in physiological ranges at 3T.

Research Question:

How accurate and reproducible are the T1- and T2-relaxation times extracted from a clinical MRI scanner, when measured monthly, over a one-year period?

The T1- and T2-relaxation times were sufficiently accurate and reproducible when measured monthly over one-year on a clinical MRI scanner. This was surmised by determining a baseline %bias (accuracy) of +2.0, +6.5, and +8.5%, respectively, and long-term CV (repeatability) of 2.1, 4.2, and 1.9%, respectively, when using T1-IR, T1-VFA and T2-SE sequences.

Shorter reference relaxation time vials in the phantom were also found to have increased measurement instability. This was likely due to the acquisition parameters implemented in the sequences, and output effects from signal saturation. No systematic variations in accuracy over time were observed, which suggests less frequent QA measurements would be required on the scanner following the establishment of baseline measurements.

One limitation of this study included that the sequences tested were not of the type typically used for patient imaging due to their long acquisition times (see Sections 2.3.4 and Table 3.1). This was addressed in the work presented in Chapter 5, through the use of anatomic-specific T1-mapping sequences which are used in a depart-

ment's current clinical trials.

Aim 2: Assess the overall technical performance of a scanner measuring the ADC.

The work presented in Chapter 4 followed similar methodologies to Chapter 3. This included performing monthly measurements on the same 3T MRI scanner over one year using standard imaging sequences. In the case of this work, a Diffusion phantom was imaged at 0°C and used to assess the overall inter-scanner ADC technical performance, including its accuracy and repeatability. These performance metrics were compared against key tolerance values set by the QIBA Diffusion Profile. Establishing conformance to this Profile helps ensure that ADC is being reliably generated on a scanner for use in prospective multi-site trials.

Although not required for Profile conformance testing, supplementary investigations into factors that could impact the technical performance of ADC in the clinic were undertaken. This included alternating the slice-encoded imaging direction and changing the tested samples' reference ADC values. These factors are important when considering imaging different anatomical sites that require multi-directional imaging, and contain anatomies that encompass a wide range of ADC values [20, 137–139]. Such an investigation is essential to progress the clinical implementation of ADC: ensuring that changes observed to occur in a tissue's ADC value are due to a physiological response and not measurement variability.

Research Questions:

How accurate, repeatable, and reproducible is the ADC value that is extracted from an MRI scanner, according to the QIBA Diffusion Profile?

According to the QIBA Diffusion Profile, the ADC values extracted from a 3T ra-

diotherapy dedicated MRI scanner were found to be accurate, repeatable, and reproducible. Specifically, the results reported in Chapter 4 indicate that the (axial) ADC generated in vial 1 by the MRI scanner was accurate (bias = + 0.05 %), repeatable (CV_{ST} = 0.1 %) and reproducible (CV_{LT} = 0.9 %), which was within the Diffusion Profile tolerance ranges [20]. Other performance metrics such as linearity (R^2 = 1.0) and b-value dependence (Dep_b = 0.4%) were within tolerance ranges [20]. Additionally, software dependency was briefly assessed in this study. Findings were inclusive of a strong correlation ($R^2 > 0.99$) between inline and offline methods for generating the ADC.

Although conformance testing was only required to be carried out using the ADC measured in vial 1 of the Diffusion phantom, monthly imaging incorporated all the phantom's vials to measure the ADC over a wide range of physiologically relevant ADC samples. Profile tolerance limits for several performance metrics, including long and short-term repeatability, were met for almost all phantom vials with the reference ADC range of 0.4–1.1 μm^2/ms. In general, lower diffusivity vials recorded the largest variability in performance.

Whilst conformance to the Diffusion Profile was found for this scanner, an inherent limitation of this study included that no in vivo test-retest assessments were performed. Such testing would supplement this study, and would need to be performed before achieving the anatomic-specific Profile claims for different treatment sites (e.g., prostate) [18, 20]. A further limitation of this study includes the detection of geometric distortions, particularly around vials with high PVP concentrations. Future investigations assessing the impact of these distortions on ADC measurements would complement this study's findings.

What impact does alternating the slice-encoding imaging direction have on ADC technical performance?

Using the vial 1 ADC values, it was found that alternating the slice-encoding imaging direction had minimal dependence on ADC technical performance. This conclusion was able to be found by repeating imaging in the coronal and sagittal plane during each monthly data acquisition and comparing ADC values to the axial results. This involved changing the slice-encoding imaging direction and physically rotating the phantom to match the selected imaging direction. Specific findings included determining accurate (%bias $\leq$ +0.83), repeatable (CV_{ST} = 0.1 %) and reproducible ($CV_{LT} \leq 0.9$ %) results for both imaging planes. Lower performance and image distortions were found to occur for vials at the furthest distance from isocenter, and for vials at phantom locations likely to experience air gaps.

Aim 3: Monitor and quantify the daily fluctuations of measured T1-relaxation time and ADC values, when using anatomy-specific qMRI sequences on a clinical MRI scanner.

The QA methods described in Chapters 3 and 4 used standardized sequences [20, 115] to calculate longitudinal scanner performance baselines and provide a basis for future multi-centre trials. The work presented in Chapter 5 extended upon these studies by investigating performance fluctuations on the clinical MRI scanner when using anatomic-specific sequences already utilized within the department. This meant adaptations to the phantom setups had to occur, including changing the coil type and phantom orientations.

Further, imaging was acquired daily to capture any fluctuations in performance that could have occurred between the monthly measurements of the prior chapters. This imaging frequency more closely reflected the inter-fraction imaging intervals that a patient would realistically receive if undertaking qMRI imaging each day throughout treatment. Such findings are especially important for prospective clinical trials which aim to use the measured changes in derived QIB values to guide future ART decisions; ensuring

inter-fraction changes in the patient's tissue QIB values are larger than any systematic fluctuations that are inherent to the scanner [7, 18].

Research Questions:

Is it possible to simplify ADC qMRI QA by measuring ADC samples at AT?

In Chapter 5, it was determined that it was possible to simplify this type of QA and measure ADC at AT instead of at 0°C. This conclusion was derived from additional scanning of the Diffusion phantom once weekly, where the ADC was re-measured when the samples within the phantom reached 0°C (as per the phantom preparation methodology described in Chapter 4). The ADC results recorded for vial 1 at different temperatures were found to be comparable and within Diffusion Profile tolerance ranges [20], indicating the ability to simplify ADC QA: median vial 1 $|\text{bias}| \leq 1.12\%$ and $\text{CV} \leq 1.18\%$.

This finding would be beneficial to a clinical department given the added logistical benefits of scanning the phantom at AT. This includes a more simple and efficient QA setup, and reducing the likelihood of susceptibility-induced image distortions [155]. A temperature correction method was required to be employed for the AT samples. However, this was easily achieved using semi-automated analysis software.

What intra-scanner variability exists in terms of accuracy and repeatability of T1-time and ADC when using anatomy-specific sequences daily?

Using the department's anatomic-specific sequences on the 3T MRI scanner, minimal intra-scanner variability in accuracy and repeatability was found when measuring the ADC daily (at AT). However, larger variability existed when the T1-time

was measured. Chapter 5 demonstrated that when using these anatomic sequences, the AT ADC median |bias| and %CV over all vials in the physiological range was $\leq 2.22\%$ and 0.62%, respectively. These results were well within Conformance tolerance limits and were thus considered sufficient [20].

However, the T1-mapping sequences had varying results. This included Lung and HN T1-mapping sequences having sufficient accuracy ($|\text{bias}| < 20\%$) and repeatability ($\text{CV} < 2\%$). The T1-mapping sequence for CNS imaging, however, was recommended to be investigated for potential sequence parameter improvements. This was due to the large inaccuracies ($> 20\%$) and low-repeatability observed to occur in several of the phantoms vials [68, 118]. E.g., Vial 2 bias = +58.62% and CV = 9.08%.

One limitation of this part of the study was that temperature dependencies in the System phantom were not accounted for when comparing daily recorded T1-times. Although daily temperature fluctuations were low in general, any potential consequential effects of these fluctuations on repeatability could have been mitigated if such dependencies were accounted for.

6.1.2 Overall qMRI QA Recommendations:

Several key recommendations for qMRI QA have been noted throughout this thesis, based on the findings within each chapter. Some of these recommendations are listed below:

- Post-processing analysis: Users should implement a partial or fully automated post-processing image analysis methodology. This includes ROI detection and placement, offline generation of QIB maps, ROI-based calculations, and statistical analysis.
- Setup apparatus: To minimise phantom positioning variability and any repercussion effects on QIB measurements, users are encouraged to construct and utilize a phantom-positioning apparatus (e.g., using 3D-printing). In theory, this would

be fitted to the department's MRI scanner couch and is used for all QA purposes. Keeping a record of coil setup should be completed to ensure the same distance between phantom/patient surface and coil is maintained between imaging repetitions. Implementing these steps should aid in minimising QA setup time.

- qMRI QA baseline testing: Prior to an MRI scanner being used in a clinical trial that plans to collect patient-based QIB data, qMRI QA baseline testing should be performed. This can be completed using similar long-term monitoring methods as per described in this thesis.
- qMRI QA frequency: Following baseline testing and determining adequate scanner performance, QIB QA is recommended to be completed annually, both before and after a clinical trial, and at times surrounding any major scanner upgrades. This includes QA for the same sequences planned to be used in the clinical trial.
- ADC QA temperature: Users can complete ADC QA at room temperature. This is assuming that the phantom temperature is: initially brought to thermal equilibrium with the MRI bore room; is monitored before and after image acquisition, and; can be corrected during the image analysis. If significant performance deviations are found to occur, users can implement standardized 0°C phantom conditions to validate findings.
- Designing multi-site QIB clinical trials: Baseline testing should be completed for each site entering the clinical trial, with additional QA performed at all relevant time points aforementioned during the clinical trial. This will assist in ensuring similar qMRI performance between MRI scanners, and that the selected sequences are performing adequately before using department resources and patient time to acquire them. Users should implement a centralised and consistent post-processing analysis method, which ideally uses offline-generated QIB maps. Users should also attempt to use the same imaging sequences between MRI scanners, including image acquisition parameters and phantom/patient setup to reduce systematic measurement variability.

6.2 Future Work

From the work presented in this thesis, several questions regarding qMRI QA were identified which require further investigation. Some of these questions include:

Is there a relationship between qMRI and clinical QA performance?

This thesis focused on the results of short- and long-term qMRI QA results, obtained using a 3T radiotherapy dedicated scanner. Parallel to completing this testing, staff within the department carried out routine clinical QA on the same MRI scanner (e.g., SNR and geometric distortion testing). It would be beneficial to investigate if the qMRI and routine MRI QA performance results were at all correlated over this time period. However, this type of study could be completed on any MRI scanner and could aid in identifying the cause of systematic changes in QIB values/performance.

One simple and efficient way to incorporate both clinical and qMRI QA testing would involve using a dual-purpose phantom, such as the ISMRM/NIST System phantom. Recall from section 2.4.3 that this phantom has features/arrays that enable clinical performance metrics to be measured, such as image resolution and geometric distortions [106]. This is additional to containing the arrays required for measuring T1- and T2-times. This dual-purpose phantom would enable measuring both performance metrics simultaneously.

More broadly speaking to address this question, novel or current international MRI QA guidelines, such as those developed by the AAPM Task Group 284 [98], need to be amended to include qMRI QA. Although this is outside the scope of this thesis, it is aimed that the results reported on within this thesis will aid in the development of such documentation.

How can different image analysis software impact qMRI measurements?

Although briefly discussed in Chapter 4, the impact of using different post-processing image analysis software or algorithms was not of focus in this thesis. In practice, inline-derived QIB maps (when possible) are a simple and efficient way for clinical departments to obtain MRI QIB estimates from an ROI without the use of any post-processing software. However, without testing the accuracy of each algorithm embedded within each vendors system, the accuracy of the QIB maps produced is unknown.

Further, for studies involving multiple sites, using inline maps can result in low reproducibility [64]. A comprehensive investigation into the best offline software/algorithm, and comparing its accuracy to inline-derived methods, would be of interest to be completed. Following on, quantifying the effects stemming from inaccurate QIBs, such as the T1-time input to DCE-MRI PK-parameter calculations, would be of particular interest when using different software. Beneficial aids for this type of work may include DROs and the use of more automated software to remove user variability.

What are the causes of signal saturation and how can the effect be minimized?

One factor impacting the results obtained in Chapters 3 and 4 was signal saturation (especially for T1-mapping). Given that observing saturation was only possible when generating the offline QIB maps, prior studies reporting on QIBs in phantoms or patients may have failed to detect this effect. For these reasons, it would be advantageous to develop automated detection methods for signal saturation in the MRI scanner and/or post-processing software. Further, it would be beneficial to dedicate an investigation to finding the causes of signal saturation in qMRI imaging

and find ways to minimize its occurrence [112]. Since this effect was not observed in the images acquired for Chapter 5, it is proposed that either coil or acquisition parameter selections could impact the magnitude of saturation observed.

Can achieving Profile conformance be an indicator of achieving anatomy-specific Profile claims?

As noted in Chapter 4, QIBA Diffusion Profile conformance results provided baseline scanner performance metrics for one MRI scanner. However, no in-vivo test-retest assessments were completed as per the requirement to achieve clinical conformance to anatomy-specific Diffusion Profile claims, such as for prostate imaging [18, 20]. The next phase of this work would involve completing such assessments using prospective volunteer and/or patient datasets. This would enable further validation of adequate scanner qMRI performance before commencing a multi-site trial based on imaging particular anatomy.

Could this work be translated to an MR-Linac?

Finally, building on both the foundations of this work and past studies in the literature [2, 7, 11, 18, 21, 37, 41, 64, 124], it is predicted that it would be simple to complete a similar level of qMRI QA testing on an MR-linac. This type of study could be approached using a single- or multi-site study, and ideally would incorporate volunteer and/or patient imaging to avoid phantom-based imaging limitations (e.g., SNR) [18].

Given that MR-linacs have the unique ability able to collect daily qMRI scans whilst delivering each fraction of radiotherapy treatment, these systems could easily collect large-scale qMRI datasets; aiding in building the evidence required for QIB clinical implementation [11, 18, 41]. Ensuring that these systems are reliable in

generating QIBs is essential to achieve the most accurate results, which preferably occurs before collecting large datasets.

6.3 Conclusion

The developed methodologies, presented results, and recommendations provided throughout this thesis were intended to aid departments to assess their MRI scanners' qMRI QIB accuracy and repeatability. As part of this thesis, three specific QIBs were investigated and the concepts and methodologies described have the potential to be applied more broadly. This includes developing QA methods for different qMRI QIBs and generating their performance baselines. Similarly, although all work presented was completed for one type of MRI scanner, the same principles can be applied to any department's MRI scanner system, including an MR-Linac.

Incorporating such assessments has the potential to increase department efficiencies, including completing QA and selecting optimal MRI acquisition parameters for specific anatomies. Further, these assessments can help improve the reliability of a department's patient-based qMRI QIB clinical trials data. These factors together have the potential to improve the calibre of existing scientific evidence which suggests probable benefits of implementing qMRI techniques into the radiotherapy workflow. As such, these techniques could ultimately be used to enhance a patient's treatment regimen. This includes plan adaptation decisions being guided by the patient's quantitative physiological changes; leading to improved treatment outcomes for the patient.

References

[1] Cancer Council, *Facts and Figures: Cancer Statistics in Australia*. URL: https://www.cancer.org.au/cancer-information/what-is-cancer/facts-and-figures. Accessed online: 13/09/2021, 2021.

[2] I. Dregely, D. Prezzi, C. Kelly-Morland, E. Roccia, R. Neji, and V. Goh, "Imaging biomarkers in oncology: Basics and application to MRI," *Journal of Magnetic Resonance Imaging*, vol. 48, pp. 13–26, jul 2018.

[3] X. Allen Li, *Adaptive Radiation Therapy*, vol. 1. CRC Press, Taylor and Francis Group, jun 2011.

[4] A. Hunt, V. N. Hansen, U. Oelfke, S. Nill, and S. Hafeez, "Adaptive Radiotherapy Enabled by MRI Guidance," *Clinical Oncology*, vol. 30, pp. 711–719, nov 2018.

[5] B. W. Raaymakers, I. M. Jürgenliemk-Schulz, G. H. Bol, M. Glitzner, A. N. Kotte, B. Van Asselen, J. C. De Boer, J. J. Bluemink, S. L. Hackett, M. A. Moerland, S. J. Woodings, J. W. Wolthaus, H. M. Van Zijp, M. E. Philippens, R. Tijssen, J. G. Kok, E. N. De Groot-Van Breugel, I. Kiekebosch, L. T. Meijers, C. N. Nomden, G. G. Sikkes, P. A. Doornaert, W. S. Eppinga, N. Kasperts, L. G. Kerkmeijer, J. H. Tersteeg, K. J. Brown, B. Pais, P. Woodhead, and J. J. Lagendijk, "First patients treated with a 1.5 T MRI-Linac: Clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment," *Physics in Medicine and Biology*, vol. 62, pp. L41–L50, nov 2017.

[6] J. M. Winfield, G. S. Payne, A. Weller, and N. M. DeSouza, "DCE-MRI, DW-MRI, and MRS in cancer: Challenges and advantages of implementing qualitative and

quantitative multi-parametric imaging in the clinic," *Topics in Magnetic Resonance Imaging*, vol. 25, pp. 245–254, nov 2016.

[7] A. Shukla-Dave, N. A. Obuchowski, T. L. Chenevert, S. Jambawalikar, L. H. Schwartz, D. Malyarenko, W. Huang, S. M. Noworolski, R. J. Young, M. S. Shiroishi, H. Kim, C. Coolens, H. Laue, C. Chung, M. Rosen, M. Boss, and E. F. Jackson, "Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials," *Journal of Magnetic Resonance Imaging*, vol. 49, pp. e101–e121, jun 2019.

[8] H. C. Thoeny and B. D. Ross, "Predicting and monitoring cancer treatment response with diffusion-weighted MRI," *Journal of Magnetic Resonance Imaging*, vol. 32, no. 1, pp. 2–16, 2010.

[9] F. Kuang, Z. Yan, J. Wang, and Z. Rao, "The value of diffusion-weighted MRI to evaluate the response to radiochemotherapy for cervical cancer," *Magnetic Resonance Imaging*, vol. 32, no. 4, pp. 342–349, 2014.

[10] L. G. Kessler, H. X. Barnhart, A. J. Buckler, K. R. Choudhury, M. V. Kondratovich, A. Toledano, A. R. Guimaraes, R. Filice, Z. Zhang, and D. C. Sullivan, "The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions," *Statistical Methods in Medical Research*, vol. 24, pp. 9–26, feb 2015.

[11] E. S. Kooreman, P. J. van Houdt, M. E. Nowee, V. W. van Pelt, R. H. Tijssen, E. S. Paulson, O. J. Gurney-Champion, J. Wang, F. Koetsveld, L. D. van Buuren, L. C. ter Beek, and U. A. van der Heide, "Feasibility and accuracy of quantitative imaging on a 1.5 T MR-linear accelerator," *Radiotherapy and Oncology*, vol. 133, pp. 156–162, apr 2019.

[12] F. M. Fennessy, R. R. Mckay, C. J. Beard, M. E. Taplin, and C. M. Tempany, "Dynamic contrast-enhanced magnetic resonance imaging in prostate cancer clinical

trials: Potential roles and possible pitfalls," *Translational Oncology*, vol. 7, no. 1, pp. 120–129, 2014.

[13] R. G. Abramson, L. R. Arlinghaus, A. N. Dula, C. C. Quarles, A. M. Stokes, J. A. Weis, J. G. Whisenant, E. Y. Chekmenev, I. Zhukov, J. M. Williams, and T. E. Yankeelov, "MR Imaging Biomarkers in Oncology Clinical Trials," *Magnetic Resonance Imaging Clinics of North America*, vol. 24, pp. 11–29, feb 2016.

[14] A. D. King, K. K. Chow, K. H. Yu, F. K. F. Mo, D. K. Yeung, J. Yuan, K. S. Bhatia, A. C. Vlantis, and A. T. Ahuja, "Head and neck squamous cell carcinoma: Diagnostic performance of diffusion-weighted MR imaging for the prediction of treatment response," *Radiology*, vol. 266, pp. 531–538, feb 2013.

[15] W. Huang, X. Li, Y. Chen, X. Li, M. C. Chang, M. J. Oborski, D. I. Malyarenko, M. Muzi, G. H. Jajamovich, A. Fedorov, A. Tudorica, S. N. Gupta, C. M. Laymon, K. I. Marro, H. A. Dyvorne, J. V. Miller, D. P. Barbodiak, T. L. Chenevert, T. E. Yankeelov, J. M. Mountz, P. E. Kinahan, R. Kikinis, B. Taouli, F. Fennessy, and J. Kalpathy-Cramer, "Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: A multicenter data analysis challenge," *Translational Oncology*, vol. 7, no. 1, pp. 153–166, 2014.

[16] W. D. Foltz, A. Wu, P. Chung, C. Catton, A. Bayley, M. Milosevic, R. Bristow, P. Warde, A. Simeonov, D. A. Jaffray, M. A. Haider, and C. Ménard, "Changes in apparent diffusion coefficient and T2 relaxation during radiotherapy for prostate cancer," *Journal of Magnetic Resonance Imaging*, vol. 37, pp. 909–916, apr 2013.

[17] A. Campbell, L. M. Davis, S. K. Wilkinson, and R. L. Hesketh, "Emerging functional imaging biomarkers of tumour responses to radiotherapy," *Cancers*, vol. 11, jan 2019.

[18] P. J. van Houdt, H. Saeed, D. Thorwarth, C. D. Fuller, W. A. Hall, B. A. McDonald, A. Shukla-Dave, E. S. Kooreman, M. E. Philippens, A. L. van Lier, R. Keesman, F. Mahmood, C. Coolens, T. Stanescu, J. Wang, N. Tyagi, A. Wetscherek, and U. A.

van der Heide, "Integration of quantitative imaging biomarkers in clinical trials for MR-guided radiotherapy: Conceptual guidance for multicentre studies from the MR-Linac Consortium Imaging Biomarker Working Group," *European Journal of Cancer*, vol. 153, pp. 64–71, aug 2021.

[19] O. Bane, S. J. Hectors, M. Wagner, L. L. Arlinghaus, M. P. Aryal, Y. Cao, T. L. Chenevert, F. Fennessy, W. Huang, N. M. Hylton, J. Kalpathy-Cramer, K. E. Keenan, D. I. Malyarenko, R. V. Mulkern, D. C. Newitt, S. E. Russek, K. F. Stupic, A. Tudorica, L. J. Wilmes, T. E. Yankeelov, Y. F. Yen, M. A. Boss, and B. Taouli, "Accuracy, repeatability, and interplatform reproducibility of T1 quantification methods used for DCE-MRI: Results from a multicenter phantom study," *Magnetic Resonance in Medicine*, vol. 79, pp. 2564–2575, may 2018.

[20] Quantitative Imaging Biomarkers Alliance, "QIBA Profile: Diffusion-Weighted Magnetic Resonance Imaging (DWI)," tech. rep., 2019.

[21] A. R. Padhani, G. Liu, D. Mu-Koh, T. L. Chenevert, H. C. Thoeny, T. Takahara, A. Dzik-Jurasz, B. D. Ross, M. Van Cauteren, D. Collins, D. A. Hammoud, G. J. Rustin, B. Taouli, and P. L. Choyke, "Diffusion-weighted magnetic resonance imaging as a cancer biomarker: Consensus and recommendations," in *Neoplasia*, vol. 11, pp. 102–125, Elsevier B.V., 2009.

[22] World Health Organization, *Cancer Today: Australia*. URL: https://gco.iarc.fr/. Accessed online: 26/06/2022., 2022.

[23] M. Barton, "Radiotherapy utilization in New South Wales from 1996 to 1998," *Australasian Radiology*, vol. 44, no. 3, pp. 308–314, 2000.

[24] T. L. Chenevert, P. E. Mckeever, and B. D. Ross, "Monitoring Early Response of Experimental Brain Tumors to Therapy Using Diffusion Magnetic Resonance Imaging," tech. rep., 1997.

[25] L. Liu, B. Yin, D. Y. Geng, Y. P. Lu, and W. J. Peng, "Changes of T2 relaxation time from neoadjuvant chemotherapy in breast cancer lesions," *Iranian Journal of Radiology*, vol. 13, jul 2016.

[26] World Health Organization, "WHO Handbook for Reporting Results of Cancer Treatment," *WHO offset publication*, vol. 48, 1979.

[27] A. Fransson, P. Andreo, and R. Pötter, "Strahlentherapie und Onkologie Aspects of MR Image Distortions in Radiotherapy Treatment Planning," *Strahlenther Onkol*, vol. 177, no. 2, pp. 59–73, 2001.

[28] P. Dawson, P. M. Deluca, K. Doi, E. Fantuzzi, R. A. Gahbauer, B. D. Michael, H. G. Paretzke, S. M. Seltzer, H. Tatsuzaki, G. F. Whitmore, and A. Allisy, "ICRU REPORT 83: Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT)," *Journal of the ICRU*, vol. 10, no. 1, 2010.

[29] A. Mundt and J. Roeske, *Image Guided Radiation Therapy*. People's Medical Publishing House, mar 2011.

[30] P. Keall, P. Poulsen, and J. T. Booth, "See, Think, and Act: Real-Time Adaptive Radiotherapy," *Seminars in Radiation Oncology*, vol. 29, pp. 228–235, jul 2019.

[31] P. Tofts, *Quantitative MRI of the brain : measuring changes caused by disease*. Wiley, 2003.

[32] F. Bloch, "Nuclear Induction," *Physical Review*, vol. 70, pp. 460–474, jul 1946.

[33] J. T. Bushberg, J. A. Seibert, E. M. Leidholdt, and J. M. Boone, *The essential physics of medical imaging*, vol. 3. Philadelphia: Lippincott Williams and Wilkins, 3 ed., 2012.

[34] H. Haynes, W. Holmes, and W. M. Holmes, "The Emergence of Magnetic Resonance Imaging (MRI) for 3D Analysis of Sediment Beds," tech. rep., 2013.

[35] D. W. McRobbie, E. A. Moore, and M. J. Graves, "MRI from picture to proton," tech. rep., Cambridge University Press, 2006.

[36] R. Bitar, G. Leung, R. Perng, S. Tadros, A. R. Moody, J. Sarrazin, C. McGregor, M. Christakis, S. Symons, A. Nelson, and T. P. Roberts, "MR pulse sequences: What every radiologist wants to know but is afraid to ask," *Radiographics*, vol. 26, pp. 513–537, mar 2006.

[37] K. E. Keenan, J. R. Biller, J. G. Delfino, M. A. Boss, M. D. Does, J. L. Evelhoch, M. A. Griswold, J. L. Gunter, R. S. Hinks, S. W. Hoffman, G. Kim, R. Lattanzi, X. Li, L. Marinelli, G. J. Metzger, P. Mukherjee, R. J. Nordstrom, A. P. Peskin, E. Perez, S. E. Russek, B. Sahiner, N. Serkova, A. Shukla-Dave, M. Steckner, K. F. Stupic, L. J. Wilmes, H. H. Wu, H. Zhang, E. F. Jackson, and D. C. Sullivan, "Recommendations towards standards for quantitative MRI (qMRI) and outstanding needs," *Journal of Magnetic Resonance Imaging*, vol. 49, pp. e26–e39, jun 2019.

[38] L. E. Olsson, M. Johansson, B. Zackrisson, and L. K. Blomqvist, "Basic concepts and applications of functional magnetic resonance imaging for radiotherapy of prostate cancer," *Physics and Imaging in Radiation Oncology*, vol. 9, pp. 50–57, jan 2019.

[39] H. Ding, C. V. Jimeno, H. Ye, T. Lindner, M. Grech-sollars, J. O'callaghan, C. Hiley, M. D. Chouhan, T. Niendorf, D. M. Koh, C. Prieto, and S. Adeleke, "Current applications and future development of magnetic resonance fingerprinting in diagnosis, characterization, and response monitoring in cancer," *Cancers*, vol. 13, no. 19, 2021.

[40] Y. Kato, K. Ichikawa, K. Okudaira, T. Taoka, H. Kawaguchi, K. Murata, K. Maruyama, G. Koerzdoerfer, J. Pfeuffer, M. Nittka, and S. Naganawa, "Comprehensive evaluation of B+1-corrected FISP-based magnetic resonance fingerprinting: Accuracy, repeatability and reproducibility of T1 and T2 relaxation times for ismrm/nist system phantom and volunteers," *Magnetic Resonance in Medical Sciences*, vol. 19, no. 3, pp. 168–175, 2020.

[41] E. S. Kooreman, P. J. van Houdt, R. Keesman, F. J. Pos, V. W. van Pelt, M. E. Nowee, A. Wetscherek, R. H. Tijssen, M. E. Philippens, D. Thorwarth, J. Wang, A. Shukla-Dave, W. A. Hall, E. S. Paulson, and U. A. van der Heide, "ADC measurements on the Unity MR-linac – A recommendation on behalf of the Elekta Unity MR-linac consortium," *Radiotherapy and Oncology*, vol. 153, pp. 106–113, dec 2020.

[42] L. K. Bittencourt, "Feasibility study of computed vs measured high b-value (1400 s/mm 2) diffusion-weighted MR images of the prostate," *World Journal of Radiology*, vol. 6, no. 6, p. 374, 2014.

[43] M. Bastiani and A. Roebroeck, "Unraveling the multiscale structural organization and connectivity of the human brain: The role of diffusion MRI," *Frontiers in Neuroanatomy*, vol. 9, no. June, 2015.

[44] M. Blasche, P. Riffel, and M. Lichy, "TimTX TrueShape and syngo ZOOMit Technical and Practical Aspects," *Siemens Product News*, pp. 74–84, 2012.

[45] G. P. Liney, L. Holloway, T. M. Al Harthi, M. Sidhom, D. Moses, E. Juresic, R. Rai, and D. J. Manton, "Quantitative evaluation of diffusion-weighted imaging techniques for the purposes of radiotherapy planning in the prostate," *British Journal of Radiology*, vol. 88, may 2015.

[46] J. Cohen-Adad, "High-Resolution DWI in Brain and Spinal Cord with syngo RESOLVE," *Clinical Neurology*, pp. 16–23, 2012.

[47] D. A. Porter and R. M. Heidemann, "High resolution diffusion-weighted imaging using readout-segmented echo-planar imaging, parallel imaging and a two-dimensional navigator-based reacquisition," *Magnetic Resonance in Medicine*, vol. 62, no. 2, pp. 468–475, 2009.

[48] J. Qiu, J. Liu, Z. Bi, X. Sun, X. Wang, J. Zhang, C. Liu, J. Zhu, and N. Qin, "Integrated slice-specific dynamic shimming diffusion weighted imaging (DWI)

for rectal Cancer detection and characterization," *Cancer Imaging*, vol. 21, dec 2021.

[49] B. Taouli, R. K. Thakur, L. Mannelli, J. S. Babb, S. Kim, E. M. Hecht, V. S. Lee, and G. M. Israel, "Renal lesions: Characterization with diffusion-weighted imaging versus contrast-enhanced MR imaging," *Radiology*, vol. 251, pp. 398–407, may 2009.

[50] G. K. Gouhar and E. S. H. Zidan, "Diffusion-weighted imaging of breast tumors: Differentiation of benign and malignant tumors," *Egyptian Journal of Radiology and Nuclear Medicine*, vol. 42, pp. 93–99, mar 2011.

[51] Y. Liu, R. Bai, H. Sun, H. Liu, X. Zhao, and Y. Li, "Diffusion-weighted imaging in predicting and monitoring the response of uterine cervical cancer to combined chemoradiation," *Clinical Radiology*, vol. 64, pp. 1067–1074, nov 2009.

[52] T. E. Yankeelov, M. Lepage, A. Chakravarthy, E. E. Broome, K. J. Niermann, M. C. Kelley, I. Meszoely, I. A. Mayer, C. R. Herman, K. McManus, R. R. Price, and J. C. Gore, "Integration of quantitative DCE-MRI and ADC mapping to monitor treatment response in human breast cancer: initial results," *Magnetic Resonance Imaging*, vol. 25, pp. 1–13, jan 2007.

[53] S. R. Chung, Y. J. Choi, C. H. Suh, J. H. Lee, and J. H. Baek, "Diffusion-weighted magnetic resonance imaging for predicting response to chemoradiation therapy for head and neck squamous cell carcinoma: A systematic review," *Korean Journal of Radiology*, vol. 20, pp. 649–661, apr 2019.

[54] M. Lambrecht, V. Vandecaveye, F. De Keyzer, S. Roels, F. Penninckx, E. Van Cutsem, C. Filip, and K. Haustermans, "Value of diffusion-weighted magnetic resonance imaging for prediction and early assessment of response to neoadjuvant radiochemotherapy in rectal cancer: Preliminary results," *International Journal of Radiation Oncology Biology Physics*, vol. 82, pp. 863–870, feb 2012.

[55] D. Musio, F. De Felice, A. L. Magnante, M. Ciolina, C. N. De Cecco, M. Rengo, A. Redler, A. Laghi, N. Raffetto, and V. Tombolini, "Diffusion-weighted magnetic resonance application in response prediction before, during, and after neoadjuvant radiochemotherapy in Primary Rectal Cancer Carcinoma," *BioMed Research International*, vol. 2013, 2013.

[56] M. Matoba, H. Tuji, Y. Shimode, I. Toyoda, Y. Kuginuki, K. Miwa, and H. Tonami, "Fractional change in apparent diffusion coefficient as an imaging biomarker for predicting treatment response in head and neck cancer treated with chemoradiotherapy," *American Journal of Neuroradiology*, vol. 35, pp. 379–385, feb 2014.

[57] A. Gladwish, M. Milosevic, A. Fyles, J. Xie, J. Halankar, U. Metser, H. Jiang, N. Becker, W. Levin, L. Manchul, W. Foltz, and K. Han, "Association of apparent diffusion coefficient with disease recurrence in patients with locally advanced cervical cancer treated with radical chemotherapy and radiation therapy," *Radiology*, vol. 279, pp. 158–166, apr 2016.

[58] S. M. Schreuder, R. Lensing, J. Stoker, and S. Bipat, "Monitoring treatment response in patients undergoing chemoradiotherapy for locally advanced uterine cervical cancer by additional diffusion-weighted imaging: A systematic review," *Journal of Magnetic Resonance Imaging*, vol. 42, pp. 572–594, sep 2015.

[59] C. Onal, G. Erbay, and O. C. Guler, "Treatment response evaluation using the mean apparent diffusion coefficient in cervical cancer patients treated with definitive chemoradiotherapy," *Journal of Magnetic Resonance Imaging*, vol. 44, pp. 1010–1019, oct 2016.

[60] S. Dashottar, T. Preeth Pany, and N. Lohia, "Role of apparent diffusion coefficient as a biomarker in the evaluation of cervical cancer," *Indian Journal of Radiology and Imaging*, vol. 29, pp. 25–32, jan 2019.

[61] R. Farjam, C. I. Tsien, F. Y. Feng, D. Gomez-Hassan, J. A. Hayman, T. S. Lawrence, and Y. Cao, "Investigation of the diffusion abnormality index as a new

imaging biomarker for early assessment of brain tumor response to radiation therapy," *Neuro-Oncology*, vol. 16, pp. 131–139, jan 2014.

[62] H. Yabuuchi, M. Hatakenaka, K. Takayama, Y. Matsuo, S. Sunami, T. Kamitani, M. Jinnouchi, S. Sakai, Y. Nakanishi, and H. Honda, "Non-small cell lung cancer: Detection of early response to chemotherapy by using contrast-enhanced dynamic and diffusion-weighted MR imaging," *Radiology*, vol. 261, pp. 598–604, nov 2011.

[63] N. Tyagi, M. Cloutier, K. Zakian, J. O. Deasy, M. Hunt, and A. Rimner, "Diffusion-weighted MRI of the lung at 3T evaluated using echo-planar-based and single-shot turbo spin-echo-based acquisition techniques for radiotherapy applications," *Journal of Applied Clinical Medical Physics*, vol. 20, pp. 284–292, jan 2019.

[64] D. C. Newitt, D. Malyarenko, T. L. Chenevert, C. C. Quarles, L. Bell, A. Fedorov, and F. Fennessy, "Multisite concordance of apparent diffusion coefficient measurements across the NCI Quantitative Imaging Network," *Journal of Medical Imaging*, vol. 5, p. 1, oct 2017.

[65] T. Yankeelov and J. Gore, "Dynamic Contrast Enhanced Magnetic Resonance Imaging in Oncology:Theory, Data Acquisition,Analysis, and Examples," *Current Medical Imaging Reviews*, vol. 3, pp. 91–107, may 2007.

[66] E. Henderson, B. K. Rutt, and T.-Y. Lee, "Temporal sampling requirements for the tracer kinetics modeling of breast disease," tech. rep., 1998.

[67] J. G. Whisenant, R. D. Dortch, W. Grissom, H. Kang, L. R. Arlinghaus, and T. E. Yankeelov, "Bloch–Siegert B1-Mapping Improves Accuracy and Precision of Longitudinal Relaxation Measurements in the Breast at 3 T," *Tomography*, vol. 2, pp. 250–259, dec 2016.

[68] F. M. Fennessy, A. Fedorov, S. N. Gupta, E. J. Schmidt, C. M. Tempany, and R. V. Mulkern, "Practical considerations in T1 mapping of prostate for dynamic contrast enhancement pharmacokinetic analyses," *Magnetic Resonance Imaging*, vol. 30, pp. 1224–1233, nov 2012.

[69] A. Sengupta, R. K. Gupta, and A. Singh, "Evaluation of B1 inhomogeneity effect on DCE-MRI data analysis of brain tumor patients at 3T," *Journal of Translational Medicine*, vol. 15, dec 2017.

[70] C. M. J. Bazelaire, G. D. Duhamel, N. M. Rofsky, and D. C. Alsop, "MR imaging relaxation times of abdominal and pelvic tissues measured in vivo at 3.0 T: preliminary results," *Radiology*, vol. 230, pp. 652–659, mar 2004.

[71] J. Z. Bojorquez, S. Bricq, C. Acquitter, F. Brunotte, P. M. Walker, and A. Lalande, "What are normal relaxation times of tissues at 3 T?," *Magnetic Resonance Imaging*, vol. 35, pp. 69–80, jan 2017.

[72] J. Yuan, S. K. K. Chow, D. K. W. Yeung, A. T. Ahuja, and A. D. King, "Quantitative evaluation of dual-flip-angle T1 mapping on DCE-MRI kinetic parameter estimation in head and neck.," *Quantitative imaging in medicine and surgery*, vol. 2, no. 4, pp. 245–24553, 2012.

[73] S. Russek, M. Boss, E. Jackson, D. Jennings, J. Evelhoch, J. Gunter, and G. Sorensen, "Characterization of NIST/ISMRM MRI System Phantom," in *Proc. Intl. Soc. Mag. Reson. Med. 20*, 2012.

[74] K. Christensen, D. Grant, E. Schulman, and C. Walling, "Optimal determination of relaxation times of fourier transform nuclear magnetic resonance. Determination of spin-lattice relaxation times in chemically polarized species," *Journal of Physics and Chemistry*, vol. 78, pp. 1971–1977, sep 1974.

[75] J. A. Brookes, T. W. Redpath, F. J. Gilbert, A. D. Murray, and R. T. Staff, "Accuracy of T1 measurement in dynamic contrast-enhanced breast MRI using two- and three-dimensional variable flip angle fast low-angle shot," *Journal of Magnetic Resonance Imaging*, vol. 9, pp. 163–171, feb 1999.

[76] K. Sung, B. L. Daniel, and B. A. Hargreaves, "Transmit B 1 + field inhomogeneity and T1 estimation errors in breast DCE-MRI at 3 tesla," *Journal of Magnetic Resonance Imaging*, vol. 38, pp. 454–459, aug 2013.

[77] C. Roberts, S. Hughes, J. H. Naish, K. Holliday, Y. Watson, S. Cheung, G. A. Buonaccorsi, H. Young, N. Clarke, and G. J. Parker, "Use of An Individually Measured Hematocrit in DCE-MRI studies," tech. rep., Proc. Intl. Soc. Mag. Reson. Med.19, 2011.

[78] M. C. Schabel and G. R. Morrell, "Uncertainty in T1 mapping using the variable flip angle method with two flip angles," *Physics in Medicine and Biology*, vol. 54, jan 2009.

[79] A. Lutti, C. Hutton, J. Finsterbusch, G. Helms, and N. Weiskopf, "Optimization and validation of methods for mapping of the radiofrequency transmit field at 3T," *Magnetic Resonance in Medicine*, vol. 64, no. 1, pp. 229–238, 2010.

[80] E. Bluemke, A. Bertrand, K. Y. Chu, N. Syed, A. G. Murchison, R. Cooke, T. Greenhalgh, B. Burns, M. Craig, N. Taylor, K. Shah, F. Gleeson, and D. Bulte, "Using Variable Flip Angle (VFA) and Modified Look-Locker Inversion Recovery (MOLLI) T1 mapping in clinical OE-MRI," *Magnetic Resonance Imaging*, vol. 89, no. November 2021, pp. 92–99, 2022.

[81] N. M. Rofsky, V. S. Lee, G. Laub, M. A. Pollack, G. A. Krinsky, D. Thomasson, M. M. Ambrosino, and W. Jeffrey C., "Abdominal MR Imaging with a Volumetric Interpolated Breath-hold Examination," *Radiology*, vol. 212, no. 3, 1999.

[82] eaMon Koh, edward RJ walton, and P. Watson, "Pictorial review viBe Mri: an alternative to ct in the imaging of sports-related osseous pathology?," tech. rep., 2018.

[83] N. M. Bruckmann, J. Kirchner, J. Morawitz, L. Umutlu, W. P. Fendler, K. Herrmann, A. K. Bittner, O. Hoffmann, T. Fehm, M. E. Lindemann, C. Buchbender, G. Antoch, and L. M. Sawicki, "Free-breathing 3D Stack of Stars GRE (StarVIBE) sequence for detecting pulmonary nodules in 18F-FDG PET/MRI," *EJNMMI Physics*, vol. 9, dec 2022.

[84] S. Kumar, R. Rai, A. Stemmer, S. Josan, L. HollowAy, S. Vinod, D. Moses, M. Franzcr, and G. Liney, "Feasibility of free breathing Lung MRI for Radiotherapy using non-Cartesian k-space acquisition schemes," tech. rep., 2017.

[85] Z. Cheng, Z. Wu, G. Shi, Z. Yi, M. Xie, W. Zeng, C. Song, C. Zheng, and J. Shen, "Discrimination between benign and malignant breast lesions using volumetric quantitative dynamic contrast-enhanced MR imaging," *European Radiology*, vol. 28, pp. 982–991, mar 2018.

[86] J. Y. Kim, S. H. Kim, Y. J. Kim, B. J. Kang, Y. Y. An, A. W. Lee, B. J. Song, Y. S. Park, and H. B. Lee, "Enhancement parameters on dynamic contrast enhanced breast MRI: Do they correlate with prognostic factors and subtypes of breast cancers?," *Magnetic Resonance Imaging*, vol. 33, no. 1, pp. 72–80, 2015.

[87] A. A. Thomas, J. Arevalo-Perez, T. Kaley, J. Lyo, K. K. Peck, W. Shi, Z. Zhang, and R. J. Young, "Dynamic contrast enhanced T1 MRI perfusion differentiates pseudoprogression from recurrent glioblastoma," *Journal of Neuro-Oncology*, vol. 125, pp. 183–190, oct 2015.

[88] A. D. King, S. K. K. Chow, K. H. Yu, F. K. F. Mo, D. K. Yeung, J. Yuan, B. K. H. Law, K. S. Bhatia, A. C. Vlantis, and A. T. Ahuja, "DCE-MRI for Pre-Treatment Prediction and Post-Treatment Assessment of Treatment Response in Sites of Squamous Cell Carcinoma in the Head and Neck," *PLoS ONE*, vol. 10, dec 2015.

[89] S. Gaddikeri, R. S. Gaddikeri, T. Tailor, and Y. Anzai, "Dynamic contrast-enhanced MR imaging in head and neck cancer: Techniques and clinical applications," *American Journal of Neuroradiology*, vol. 37, pp. 588–595, apr 2016.

[90] M. A. Zahra, L. T. Tan, A. N. Priest, M. J. Graves, M. Arends, R. A. Crawford, J. D. Brenton, D. J. Lomas, and E. Sala, "Semiquantitative and Quantitative Dynamic Contrast-Enhanced Magnetic Resonance Imaging Measurements Predict Radiation

Response in Cervix Cancer," *International Journal of Radiation Oncology Biology Physics*, vol. 74, pp. 766–773, jul 2009.

[91] K. A. Kumar, K. K. Peck, S. Karimi, E. Lis, A. I. Holodny, M. H. Bilsky, and Y. Yamada, "A Pilot Study Evaluating the Use of Dynamic Contrast-Enhanced Perfusion MRI to Predict Local Recurrence After Radiosurgery on Spinal Metastases," *Technology in Cancer Research and Treatment*, vol. 16, pp. 857–865, dec 2017.

[92] A. Fedorov, J. Fluckiger, G. D. Ayers, X. Li, S. N. Gupta, C. Tempany, R. Mulkern, T. E. Yankeelov, and F. M. Fennessy, "A comparison of two methods for estimating DCE-MRI parameters via individual and cohort based AIFs in prostate cancer: A step towards practical implementation," *Magnetic Resonance Imaging*, vol. 32, no. 4, pp. 321–329, 2014.

[93] X. Li, W. Huang, and W. D. Rooney, "Signal-to-noise ratio, contrast-to-noise ratio and pharmacokinetic modeling considerations in dynamic contrast-enhanced magnetic resonance imaging," *Magnetic Resonance Imaging*, vol. 30, pp. 1313–1322, nov 2012.

[94] M. J. Firbank, R. M. Harrison, E. D. Williams, and A. Coulthard, "Quality assurance for MRI: Practical experience," *British Journal of Radiology*, vol. 73, no. 868, pp. 376–383, 2000.

[95] K. Sandgren, "Development of a Quality Assurance Strategy for Magnetic Resonance Imaging in Radiotherapy," *Thesis*, 2015.

[96] National Electrical Manufacturers Association, "Standard for Determination of Signal-to-Noise Ratio (SNR) in Diagnostic Magnetic Resonance Imaging," tech. rep., 2008.

[97] National Electrical Manufacturers Association, "NEMA Standards Publication National Electrical Manufacturers Association Determination of Slice Thickness in Diagnostic Magnetic Resonance Imaging," tech. rep.

[98] E. S. Paulson, K. Mcgee, Y. Hu, J. Balter, and J. Bayouth, "Task group 284 report : magnetic resonance imaging simulation in radiotherapy : considerations for clinical implementation , optimization , and quality assurance,"

[99] D. C. Sullivan, N. A. Obuchowski, L. G. Kessler, D. L. Raunig, C. Gatsonis, E. P. Huang, M. Kondratovich, L. M. McShane, A. P. Reeves, D. P. Barboriak, A. R. Guimaraes, and R. L. Wahl, "Metrology standards for quantitative imaging biomarkers1," *Radiology*, vol. 277, no. 3, pp. 813–825, 2015.

[100] K. E. Keenan, M. Ainslie, A. J. Barker, M. A. Boss, K. M. Cecil, C. Charles, T. L. Chenevert, L. Clarke, J. L. Evelhoch, P. Finn, D. Gembris, J. L. Gunter, D. L. Hill, C. R. Jack, E. F. Jackson, G. Liu, S. E. Russek, S. D. Sharma, M. Steckner, K. F. Stupic, J. D. Trzasko, C. Yuan, and J. Zheng, "Quantitative magnetic resonance imaging phantoms: A review and the need for a system phantom," *Magnetic Resonance in Medicine*, vol. 79, pp. 48–61, jan 2018.

[101] Radiological Society of North America, "Profile: DCE MRI Quantification," tech. rep., 2012.

[102] Radiological Society of North America, *Quantitative imaging Data Warehouse*. URL: https://qidw.rsna.org/. Accessed online: 25/08/2021.

[103] C. C. Chen, Y. L. Wan, Y. Y. Wai, and H. L. Liu, "Quality assurance of clinical MRI scanners using ACR MRI phantom: Preliminary results," *Journal of Digital Imaging*, vol. 17, pp. 279–284, dec 2004.

[104] R. A. Lerski and J. D. De, "Performance assessment and quality control in MRI by Eurospin test objects and protocols," tech. rep., 1993.

[105] J. L. Gunter, M. A. Bernstein, B. J. Borowski, C. P. Ward, P. J. Britson, J. P. Felmlee, N. Schuff, M. Weiner, and C. R. Jack, "Measurement of MRI scanner performance with the ADNI phantom," *Medical Physics*, vol. 36, no. 6, pp. 2193–2205, 2009.

[106] K. F. Stupic, M. Ainslie, M. A. Boss, C. Charles, A. M. Dienstfrey, J. L. Evelhoch, P. Finn, Z. Gimbutas, J. L. Gunter, D. L. Hill, C. R. Jack, E. F. Jackson, T. Karaulanov, K. E. Keenan, G. Liu, M. N. Martin, P. V. Prasad, N. S. Rentz, C. Yuan, and S. E. Russek, "A standard system phantom for magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 86, pp. 1194–1211, sep 2021.

[107] M. A. Boss, T. L. Chenevert, J. C. Waterton, D. M. Morris, H. Ragheb, A. Jackson, N. Desouza, D. J. Collins, B. E. Van Beers, P. Garteiser, S. Doblas, S. E. Russek, K. E. Keenan, E. F. Jackson, and G. Zahlmann, "Temperature-controlled Isotropic Diffusion Phantom with Wide Range of Apparent Diffusion Coefficients for Multicenter Assessment of Scanner Repeatability and Reproducibility," tech. rep., ISMRM, Milan, Italy, 2014.

[108] CaliberMRI, *Diffusion Phantom for ADC qMRI Standardization: Revision G*. URL: https://qmri.com/diffusion-phantom-manual-spec-sheet/. Accessed online: 20/09/2021., aug 2021.

[109] E. M. Palacios, A. J. Martin, M. A. Boss, F. Ezekiel, Y. S. Chang, E. L. Yuh, M. J. Vassar, D. M. Schnyer, C. L. MacDonald, K. L. Crawford, A. Irimia, A. W. Toga, and P. Mukherjee, "Toward precision and reproducibility of diffusion tensor imaging: A multicenter diffusion phantom and traveling volunteer study," *American Journal of Neuroradiology*, vol. 38, pp. 537–545, mar 2017.

[110] Y. Jiang, D. Ma, K. E. Keenan, K. F. Stupic, V. Gulani, and M. A. Griswold, "Repeatability of magnetic resonance fingerprinting T1 and T2 estimates assessed using the ISMRM/NIST MRI system phantom," *Magnetic Resonance in Medicine*, vol. 78, pp. 1452–1457, oct 2017.

[111] C. F. Maier, S. G. Tan, H. Hariharan, and H. G. Potter, "T2 quantitation of articular cartilage at 1.5 T," *Journal of Magnetic Resonance Imaging*, vol. 17, pp. 358–364, mar 2003.

[112] K. E. Keenan, Z. Gimbutas, A. Dienstfrey, and K. F. Stupic, "Assessing effects of scanner upgrades for clinical studies," *Journal of Magnetic Resonance Imaging*, vol. 50, pp. 1948–1954, dec 2019.

[113] T. Ihalainen, O. Sipilä, and S. Savolainen, "MRI quality control: Six imagers studied using eleven unified quality parameters," *European Radiology*, vol. 14, pp. 1859–1865, oct 2004.

[114] J. Jang, L. H. Ngo, G. Captur, J. C. Moon, and R. Nezafat, "Measurement reproducibility of sliceinterleaved T1 and T2 mapping sequences over 20 months: A single center study," *PLoS ONE*, vol. 14, jul 2019.

[115] Qualibre-MD, "Qalibre-MD. NIST/ISMRM System Phantom. Manual: m14130-500-0-doc Revision L," 2016.

[116] M. A. Boss, A. M. Dienstfrey, Z. Gimbutas, K. E. Keenan, A. B. Kos, J. D. Splett, K. F. Stupic, and S. E. Russek, "Magnetic resonance imaging biomarker calibration service: proton spin relaxation times," tech. rep., National Institute of Standards and Technology, Gaithersburg, MD, may 2018.

[117] CaliberMRI, "System Phantom for T1, T2 and Proton Density qMRI Standardization, Manual: Revision O," 2020.

[118] S. E. Russek, K. F. Stupic, J. R. Biller, M. A. Boss, K. E. Keenan, and E. Mirowski, "Electromagnetics for Quantitative Magnetic Resonance Imaging," tech. rep., 2020.

[119] M. Newville, T. Stensitzki, D. Allen, M. Rawlik, A. Ingargiola, and A. Nelson, "Non-Linear Least-Squares Minimization and Curve-Fitting for Python," 2016.

[120] A. McCann, C. McGrath, and P. Wilson, "Initial Experiences with the ISMRM / NIST Quantitative MRI 'System Standard' Phantom," in *IPEM. Quantitative MRI: Clinical Applications and Quality Assurance*, nov 2018.

[121] D. Milford, N. Rosbach, M. Bendszus, and S. Heiland, "Mono-exponential fitting in T2-relaxometry: Relevance of offset and first echo," *PLoS ONE*, vol. 10, dec 2015.

[122] A. Shallan, R. Guijt, and M. Breadmore, "Capillary Electrophoresis: Basic Principles," *Encyclopedia of Forensic Sciences*, pp. 549–559, feb 2013.

[123] K. E. Keenan, M. Boss, E. F. Jackson, S.-J. Kown, D. Jennings, and S. Russek, "NIST/ISMRM MRI System Phantom T1 Measurements on Multiple MRI Systems," (Utah, USA), ISMRM, 2013.

[124] Y. Wang, S. Tadimalla, R. Rai, J. Goodwin, S. Foster, G. Liney, L. Holloway, and A. Haworth, "Quantitative MRI: Defining repeatability, reproducibility and accuracy for prostate cancer imaging biomarker development," *Magnetic Resonance Imaging*, vol. 77, pp. 169–179, apr 2021.

[125] A. Hagiwara, M. Hori, J. Cohen-Adad, M. Nakazawa, Y. Suzuki, A. Kasahara, M. Horita, T. Haruyama, C. Andica, T. Maekawa, K. Kamagata, K. K. Kumamaru, O. Abe, and S. Aoki, "Linearity, Bias, Intrascanner Repeatability, and Interscanner Reproducibility of Quantitative Multidynamic Multiecho Sequence for Rapid Simultaneous Relaxometry at 3 T: A Validation Study with a Standardized Phantom and Healthy Controls," *Investigative Radiology*, vol. 54, pp. 39–47, jan 2019.

[126] N. Stikov, M. Boudreau, I. R. Levesque, C. L. Tardif, J. K. Barral, and G. B. Pike, "On the accuracy of T1 mapping: Searching for common ground," *Magnetic Resonance in Medicine*, vol. 73, pp. 514–522, feb 2015.

[127] QIBA, "Quantitative Imaging Biomarkers Alliance (QIBA) Profile: DCE-MRI Quantification (DCEMRI-Q)," jul 2017.

[128] J. K. Barral, E. Gudmundson, N. Stikov, M. Etezadi-Amoli, P. Stoica, and D. G. Nishimura, "A robust methodology for in vivo T1 mapping," *Magnetic Resonance in Medicine*, vol. 64, no. 4, pp. 1057–1067, 2010.

[129] T. S. Ahearn, R. T. Staff, T. W. Redpath, and S. I. Semple, "The use of the Levenberg-Marquardt curve-fitting algorithm in pharmacokinetic modelling of DCE-MRI data," *Physics in Medicine and Biology*, vol. 50, no. 9, 2005.

[130] P. Bevington and D. Robinson, "Data reduction and error analysis for the physical sciences," *New York: McGraw-Hill*, p. 163, 1963.

[131] Y. Lee, M. F. Callaghan, J. Acosta-Cabronero, A. Lutti, and Z. Nagy, "Establishing intra- and inter-vendor reproducibility of T1 relaxation time measurements with 3T MRI," *Magnetic Resonance in Medicine*, vol. 81, pp. 454–465, jan 2019.

[132] O. Potvin, A. Khademi, I. Chouinard, F. Farokhian, L. Dieumegarde, I. Leppert, R. Hoge, M. N. Rajah, P. Bellec, and S. Duchesne, "Measurement Variability Following MRI System Upgrade," *Frontiers in Neurology*, vol. 10, jul 2019.

[133] B. Moreau, A. Iannessi, C. Hoog, and H. Beaumont, "How reliable are ADC measurements? A phantom and clinical study of cervical lymph nodes," *European Radiology*, vol. 28, pp. 3362–3371, aug 2018.

[134] N. P. Jerome, M.-V. Papoutsaki, M. R. Orton, H. G. Parkes, J. M. Winfield, M. A. Boss, M. O. Leach, N. M. Desouza, and D. J. Collins, "Development of a temperature-controlled phantom for magnetic resonance quality assurance of diffusion, dynamic, and relaxometry measurements," *Medical Physics*, pp. 2998–3007, 2016.

[135] D. Malyarenko, C. J. Galb An, F. J. Londy, C. R. Meyer, T. D. Johnson, A. Rehemtulla, B. D. Ross, and T. L. Chenevert, "Multi-system Repeatability and Reproducibility of Apparent Diffusion Coefficient Measurement Using an Ice-Water Phantom," *J. Magn. Reson. Imaging*, vol. 37, pp. 1238–1246, 2013.

[136] R. Paudyal, A. S. Konar, N. A. Obuchowski, V. Hatzoglou, T. L. Chenevert, D. I. Malyarenko, S. D. Swanson, E. Locastro, S. Jambawalikar, M. Z. Liu, L. H. Schwartz, R. M. Tuttle, N. Lee, and A. Shukla-Dave, "Repeatability of quanti-

tative diffusion-weighted imaging metrics in phantoms, head-and-neck and thyroid cancers: Preliminary findings," *Tomography*, vol. 5, pp. 15–25, mar 2019.

[137] N. Takeshige, T. Aoki, K. Sakata, S. Kajiwara, T. Negoto, S. Nagase, S. Tanoue, Y. Uchiyama, M. Hirohata, T. Abe, and M. Morioka, "Sagittal diffusion-weighted imaging in preventing the false-negative diagnosis of acute brainstem infarction: Confirmation of the benefit by anatomical characterization of false-negative lesions," *Surgical Neurology International*, vol. 10, no. 180, 2019.

[138] Felfeli P, Wenz H, Al-Zghloul M, Groden C, and Förster A, "Combination of standard axial and thin- section coronal diffusion- weighted imaging facilitates the diagnosis of brainstem infarction," 2017.

[139] M. A. Jacobs, K. J. Macura, A. Zaheer, E. S. Antonarakis, V. Stearns, A. C. Wolff, T. Feiweier, I. R. Kamel, R. L. Wahl, and L. Pan, "Multiparametric Whole-body MRI with Diffusion-weighted Imaging and ADC Mapping for the Identification of Visceral and Osseous Metastases From Solid Tumors," *Academic Radiology*, vol. 25, pp. 1405–1414, nov 2018.

[140] D. Malyarenko, A. Fedorov, L. Bell, M. Prah, S. Hectors, L. Arlinghaus, M. Muzi, M. Solaiyappan, M. Jacobs, M. Fung, A. Shukla-Dave, K. McManus, M. Boss, B. Taouli, T. Yankeelov, C. Quarles, K. Schmainda, T. Chenevert, and D. Newitt, "Toward uniform implementation of parametric map Digital Imaging and Communication in Medicine standard in multisite quantitative diffusion imaging studies," *Journal of Medical Imaging*, vol. 5, p. 1, oct 2017.

[141] D. Le Bihan, C. Poupon, A. Amadon, and F. Lethimonnier, "Artifacts and Pitfalls in Diffusion MRI," *J. Magn. Reson. Imaging*, vol. 24, pp. 478–488, 2006.

[142] H. Y. Mesri, S. David, M. A. Viergever, and A. Leemans, "The adverse effect of gradient nonlinearities on diffusion MRI: From voxels to group studies," *NeuroImage*, vol. 205, jan 2020.

[143] B. Gunda, R. Porcher, M. Duering, J.-P. Guichard, and J. Mawet, "ADC Histograms from Routine DWI for Longitudinal Studies in Cerebral Small Vessel Disease: A Field Study in CADASIL," *PLoS ONE*, vol. 9, no. 5, p. 97173, 2014.

[144] M. E. Miquel, A. D. Scott, N. D. Macdougall, R. Boubertakh, N. Bharwani, and A. G. Rockall, "In vitro and in vivo repeatability of abdominal diffusion weighted MRI," *British Journal of Radiology*, vol. 85, pp. 1507–1512, nov 2012.

[145] E. Subashi, A. Dresner, and N. Tyagi, "Longitudinal assessment of quality assurance measurements in a 1.5 T MR-linac: Part II—Magnetic resonance imaging," *Journal of Applied Clinical Medical Physics*, mar 2022.

[146] T. L. Chenevert, C. J. Galbán, M. K. Ivancevic, S. E. Rohrer, F. J. Londy, T. C. Kwee, C. R. Meyer, T. D. Johnson, A. Rehemtulla, and B. D. Ross, "Diffusion coefficient measurement using a temperature-controlled fluid for quality control in multicenter studies," *Journal of Magnetic Resonance Imaging*, vol. 34, pp. 983–987, oct 2011.

[147] Siemens Healthineers, *Body 18: Features and Benefits.* URL: https://www.siemens-healthineers.com/magnetic-resonance-imaging/options-and-upgrades/coils/body-18/. Accessed online: 10/01/2022., 2022.

[148] E. Çolak, S. Bayraktaroğlu, Ö. Akagündüz, R. Savaş, and M. Esassolak, "Correlation of ADC values measured using 3T diffusion-weighted MRI and SUVs from fluorodeoxyglucose PET/CT in head and neck squamous cell carcinomas," *Gulhane Medical Journal*, vol. 62, pp. 72–79, 2020.

[149] J. G. McPherson, A. C. Smith, D. A. Duben, K. L. McMahon, M. Wasielewski, T. B. Parrish, and J. M. Elliott, "Short- and long-term reproducibility of diffusion-weighted magnetic resonance imaging of lower extremity musculature in asymptomatic individuals and a comparison to individuals with spinal cord injury," *BMC Musculoskeletal Disorders*, vol. 19, dec 2018.

[150] J. Helenius, L. Soinne, J. Perkiö, O. Salonen, A. Kangasmäki, M. Kaste, R. A. D. Carano, H. J. Aronen, and T. Tatlisumak, "Diffusion-Weighted MR Imaging in Normal Human Brains in Various Age Groups," *American journal of neuroradiology*, vol. 23, no. 2, pp. 194–199, 2002.

[151] Y. Zhang, J.-Y. Chen, C.-M. Xie, Y.-X. Mo, X.-W. Liu, Y. Liu, and P.-H. Wu, "Diffusion-Weighted Magnetic Resonance Imaging for Prediction of Response of Advanced Cervical Cancer to Chemoradiation," tech. rep., 2011.

[152] V. N. Harry, S. I. Semple, F. J. Gilbert, and D. E. Parkin, "Diffusion-weighted magnetic resonance imaging in the early detection of response to chemoradiation in cervical cancer," *Gynecologic Oncology*, vol. 111, pp. 213–220, nov 2008.

[153] S. Chen, R. Liu, C. Ma, Y. Bian, J. Li, P. Yang, M. Wang, and J. Lu, "Repeatability of Apparent Diffusion Coefficient at 3.0 Tesla in Normal Pancreas," *Cureus*, jun 2021.

[154] CaliberMRI, *System Phantom: T1 Batch 2 Values*. URL: https://qmri.com/wp-content/uploads/2021/06/T1-Batch-2-Values-compressed.pdf.Accessed: 2022-05-23, 2022.

[155] A. McCan, P. Wison, and C. McGrath, "Initial Experiences with the NIST/RSNA/NCI Quantitative MRI Diffusion Phantom," in *IPEM- Quantitative MRI- Applications and Quality Assurance*, (Belfast), nov 2018.

[156] F. Wagner, F. B. Laun, T. A. Kuder, A. Mlynarska, F. Maier, J. Faust, K. Demberg, L. Lindemann, B. Rivkin, A. M. Nagel, M. E. Ladd, K. Maier-Hein, S. Bickelhaupt, and M. Bach, "Temperature and concentration calibration of aqueous polyvinylpyrrolidone (PVP) solutions for isotropic diffusion MRI phantoms," *PLoS ONE*, vol. 12, jun 2017.

[157] P. Pullens, P. Bladt, J. Sijbers, A. I. Maas, and P. M. Parizel, "Technical Note: A safe, cheap, and easy-to-use isotropic diffusion MRI phantom for clinical and multicenter studies," *Medical physics*, vol. 44, pp. 1063–1070, mar 2017.

[158] CaliberMRI, *qCal MR: Automate the known for confident MRI standardization*. URL: https://qmri.com/qmri-platform/qcal-software/. Accessed online: 23/05/2022., 2022.

[159] Zachary Chin, *Automated quality assurance of T1 and T2 quantitative MRI measurements for stability and reproducibility analysis in multi-centre quantitative biomarker studies*. Msc, RMIT University, nov 2020.

[160] K. E. Keenan, K. F. Stupic, S. E. Russek, and E. Mirowski, "MRI-visible liquid crystal thermometer," *Magnetic Resonance in Medicine*, vol. 84, pp. 1552–1563, sep 2020.

[161] M. Cercignani, N. G. Dowell, and P. S. Tofts, *Quantitative MRI of the Brain : Principles of Physical Measurement, Second Edition*. Taylor and Francis Group, 2 ed., jan 2018.

[162] M. B. Nichols and C. B. Paschal, "Measurement of longitudinal (T1) relaxation in the human lung at 3.0 Tesla with tissue-based and regional gradient analyses," *Journal of Magnetic Resonance Imaging*, vol. 27, pp. 224–228, jan 2008.

[163] O. Dietrich, J. Raya, M. Peller, J. G. Raya, U. Fasol, M. F. Reiser, and S. O. Schoenberg, "Oxygen-enhanced MRI of the lung at 3 Tesla: Feasibility and T1 relaxation times Functional Lung MRI-Perfusion and Ventilation View project SNR Quantification in MRI View project Oxygen-enhanced MRI of the lung at 3 Tesla: Feasibility and T 1 relaxation times," tech. rep., 2006.

[164] G. E. Gold, E. Han, J. Stainsby, G. Wright, J. Brittain, and C. Beaulieu, "Musculoskeletal MRI at 3.0 T: Relaxation Times and Image Contras," *American Roentgen Ray Society*, aug 2004.

[165] M. Xie, Z. Ren, D. Bian, D. Li, L. Yu, F. Zhu, R. Huang, Z. Zhang, S. Suye, and C. Fu, "High resolution diffusion-weighted imaging with readout segmentation of long variable echo-trains for determining myometrial invasion in endometrial

carcinoma," *Cancer imaging : the official publication of the International Cancer Imaging Society*, vol. 20, p. 66, sep 2020.

[166] Q. Q. Zhou, W. Zhang, Y. S. Yu, H. Y. Li, L. Wei, X. S. Li, Z. Z. He, and H. Zhang, "Comparative Study between ZOOMit and Conventional Intravoxel Incoherent Motion MRI for Assessing Parotid Gland Abnormalities in Patients with Early- or Mid-Stage Sjögren's Syndrome," *Korean Journal of Radiology*, vol. 23, no. 4, p. 455, 2022.

[167] K. W. Yeom, S. J. Holdsworth, A. T. Van, M. Iv, S. Skare, R. M. Lober, and R. Bammer, "Comparison of readout-segmented echo-planar imaging (EPI) and single-shot epi in clinical application of diffusion-weighted imaging of the pediatric brain," *American Journal of Roentgenology*, vol. 200, no. 5, pp. 437–443, 2013.

[168] L. Guo, C. Liu, W. Chen, Q. Chan, and G. Wang, "Dual-source parallel RF transmission for diffusion-weighted imaging of the abdomen using different b values: Image quality and apparent diffusion coefficient comparison with conventional single-source transmission," *Journal of Magnetic Resonance Imaging*, vol. 37, no. 4, pp. 875–885, 2013.

Appendix A

In-house developed Python scripts

The in-house developed Python scripts developed for the purpose of this thesis can be found within the Ingham Institute Physics GitHub repository:

https://github.com/ingham-physics/MR-Sim-Project/tree/main/Maddie_qMRI

Specifically, the code developed for the purpose of analysing the data presented in Chapter 3 can be found within the sub-folder named 'Sys_Phant'. Similarly, the code used for the analysis described in Chapter 4 is located within the sub-folder labelled 'Dif_Phant'. And lastly, the code used for the analysis described in Chapter 5, is included inside the sub-folder labelled 'Daily_QA'.