Development and validation of novel and quantitative MRI methods for cancer evaluation

Harbir Singh Sidhu

Ph.D. Thesis,

Submitted for the degree of Doctor of Philosophy

September 2021

Declaration

I, Harbir Singh Sidhu, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Harbir S Sidhu

Abstract

Quantitative imaging biomarkers (QIB) offer the opportunity to further the evaluation of cancer at presentation as well as predict response to anti-cancer therapies before and early during treatment with the ultimate goal of truly personalised medical care and the mitigation of futile, often detrimental, therapy. Few QIBs are successfully translated into clinical practice and there is increasing recognition that rigorous methodologies and standardisation of research pipelines and techniques are required to move a theoretically useful biomarker into the clinic.

To this end, I have aimed to give an overview of what I believe to be some of key elements within the research field beginning with the concept of imaging biomarkers, introducing concepts in development and validation, before providing a summary of the current and future utility of a range of quantitative MR imaging biomarkers techniques within the oncological imaging field.

The original, prospective, research moves from the technical and analytical validation of a novel QIB use (T1 mapping in cancer), first in vivo qualification of this biomarker in cancer patient response assessment and prediction (sarcoma and breast cancer as well as prostate cancer separately), and then moving on to application of more established QIBs in cancer evaluation (R2*/BOLD imaging in head and neck cancer) as well as how existing MR data can be post-processed to improved cancer evaluation (further metrics derived from diffusion weighted imaging in head and neck cancer and textural analysis of existing clinical MR images utility in prostate cancer detection).

Impact Statement

Beyond dissemination through presentation at international conferences and meetings as well as peer reviewed publication; this body of research could realise beneficial effects in several ways both inside and outside of academia.

Within academia, the validation of a clinically feasible T1 mapping MR methodology which can be performed within a reasonable time frame on a whole body basis together with single centre qualification in a range of cancer cohorts lends an illustrative example of biomarker development as well as a novel use for this biomarker in an oncological setting which should be followed up in larger scale, multicentre collaborative studies to confirm repeatability and reproducibility on other platforms and across different scanners and institutions. Novel findings relating to cancerous lymph nodes in head and neck cancer oxygenation are not previously described and require further work to confirm the observation and if replicated its biophysical basis for potential exploitation in treatment planning for these patients. Finally, work demonstrating improvements in prostate cancer detection by application of textural analysis on existing MR datasets could feasibly be applied directly to conventionally acquired MRI datasets without need for recall or modification.

Outside of academia, I hope to see aspects of this body of work translating to clinical practice both within the NHS and further afield. Whole body MR evaluation has been adopted into NICE guidance for various cancer cohorts for both detection and response evaluation but the added value of QIBs embedded within these studies requires more rigorous standardisation and collaboration; I hope this work will contribute a small amount to this process. The importance of non-academic radiologists and clinicians being aware of this work and the public more generally should also facilitate this process.

Acknowledgments

First and foremost I would like to thank my primary supervisor, Prof Shonit Punwani for his amazing supervision, support (both professionally and personally), and patience! My secondary supervisor Prof Stuart Taylor again has been an exemplary supervisor, mentor and colleague whom I would also like to extend my gratitude. I would also like to acknowledge this work would not have been possible without the general support of our research group encompassing the entire Centre for Medical Imaging team as well as research bookings staff. I have also to express gratitude to the clinical radiology department at UCLH from my consultant colleagues to all the other allied healthcare professionals. Crucially, I would also like to thank all the patients both involved in these studies or just considered being involved at what I imagine to be the most difficult time in many of their

lives- many of the patients whose data is presented here are no longer with us, I have nothing but gratitude, admiration and respect for their selflessness, time and involvement. My family's support has been enduring through not only this PhD but all along my academic career and I would likely not be writing this if not for this.

I would finally and most importantly I'd like to thank my partner Jen for her enduring love, support and patience with this whole process- hopefully worth it JTB, and this work is dedicated to you!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SECTION A: INTRODUCTION and BACKGROUND

Chapter 1 – Literature Review

1.1 Imaging Biomarkers

The transformative effect of the growing understanding of the molecular bases of disease has led to ever increasing genotypic and phenotypic molecular information on the management of disease, varyingly referred to as 'precision' or 'personalised' medicine [1]. Tailoring and monitoring therapy to an individual's molecular signature of disease necessitates measurement of a variety of anatomical, physiological, and biochemical properties of the body- characteristics referred to as 'biomarkers'. A biomarker is defined as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or a response to a therapeutic intervention" [2]. It is acknowledged however that the success of medical science and industry in the twentieth century in delivering innovative discoveries to patients has recently given way to a relative stagnation of advancement in the twenty-first century- occurring mainly as a result of challenges in the translation of bench side basic science discoveries to bedside diagnostics and therapeutics. [3]. This 'pipeline' problem is a result of the increasingly challenging, inefficient and costly path of medical product development and has prompted multiple government agencies, along with other public and private partners, to develop an array of initiatives and consortia with the ostensible aim of tackling this problem.

In parallel, the importance of medical imaging for clinical decision making has been steadily increasing over the last four decades. Medical imaging provides the ability to detect and localise many changes that are important to determine whether a disease is present or therapy is effective, by depicting alterations in anatomic, physiological, biochemical or molecular processes [4]. With appropriate calibration, these imaging technologies can provide quantitative information about some tissue properties from which the imaging signal has been derived. Thus, such imaging methods also constitute biomarker measurement processes and are conceptually similar to laboratory or physiological assays. Clinical endpoints are vital in assessing those biomarkers predicting or monitoring the clinical response to therapy and may be defined as "a characteristic or variable that reflects how a patient feels, functions, or survives" [5]. Traditional clinical endpoints can be difficult to standardise or quantify, mandate long term follow-up and incur great cost when applied to clinical trials, particularly when a long-term endpoint (such as mortality) is used. A surrogate endpoint is “a biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit based on other scientific evidence” [5], the identification and application of which are used to expedite the process of clinical analysis.

The term quantitative imaging has been defined as “the extraction of quantifiable features from medical images for the assessment of normal [findings] or the severity, degree of change, or status of a disease, injury, or chronic condition relative to normal [findings]” [6]. By combining these two concepts, a quantitative imaging biomarker (QIB) can be defined as an objectively measured characteristic derived from an in vivo image as an indicator of normal biological processes, pathogenic processes, or response to a therapeutic intervention. Potential gains from the growth of QIB include increased diagnostic accuracy, decreased variability and subjectivity of image analysis, increased automation of data reporting, more robust association of imaging findings with other biological and clinical parameters, and the opportunity for large-scale attempts to link phenotypic imaging patterns with genomic profiles.

1.2 Imaging Biomarker Development and Validation

For imaging biomarkers to play an important role in the future evolution of precision medicine, both technical performance and clinical performance need to be evaluated rigorously [1]. To determine the usefulness of QIB measurements, it is crucial that the framework in which they are acquired is described in detail, including context of use, acquisition parameters, and measurement methods. After that framework is described, then the variability and error according to those settings can be quantified. Knowledge of these factors will facilitate reliable comparison of measurements over time and across imaging platforms. The Radiological Society of North America (RSNA) has organised the Quantitative Imaging Biomarkers Alliance (QIBA) with the purpose of improving the value and practicality of QIBs by reducing variability across devices, patients, and time [6]. The QIBA Metrology Working Group has produced a useful common framework defining relevant terminology [7-9] summarised in Table 1.

Terms	Definition
QIB	A characteristic derived from one or more in vivo images and objectively measured according to a ratio or interval scale as an indicator of normal biological processes, pathogenic processes, or response to a therapeutic intervention.
Measurand	The quantity intended to be measured (VIM clause 2.3).
Analytical validation	Demonstration of the accuracy, precision, and feasibility of biomarker measurement.
Qualification	Demonstration that a biomarker is associated with a clinical endpoint
Utilization	Assessment of biomarker performance in the specific context of its proposed use.
Bias	An estimate of a systematic measurement error (VIM clause 2.18).
Linearity	The ability to provide measured quantity values that are directly proportional to the value of the measurand in the experimental unit (ISO standard 18113).
Precision	The closeness of agreement between measured quantity values obtained by means of replicate measurements of the same or similar experimental units with specified conditions (VIM clause 2.15). Repeatability and reproducibility are types of precision.
Reference Value	A value, generally accepted as having a suitably small measurement uncertainty, to be used as a basis for comparison with values of quantities of the same kind (e.g. the mean of a large number of replicate measurements) by using a reference method (VIM clause 5.18).
Repeatability	The measurement precision with conditions that remain unchanged between replicate measurements (repeatability conditions) (VIM clause 2.20).
Repeatability Conditions	The set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions, same physical location, and replicate measurements of the same or similar experimental units

	over a short period of time.
Reproducibility	The measurement precision with conditions that vary between replicate measurements (reproducibility conditions) (VIM clause 2.25).
Reproducibility Conditions	The set of conditions that includes (a) different locations, operators, and measuring systems and (b) replicate measurements of the same or similar objects.
Truth or true values	In metrology, truth is the real or actual value of a quantity associated with some object. Because each physical measurement has some uncertainty in terms of whether it agrees with the real quantity value, the true value cannot be known with certainty.
NOTE: ISO = International Organization for Standardization, VIM = International Vocabulary of Metrology.	

Table 1.1 Terminology for Describing the Technical Performance of QIBs (adapted from [1] and [20]).

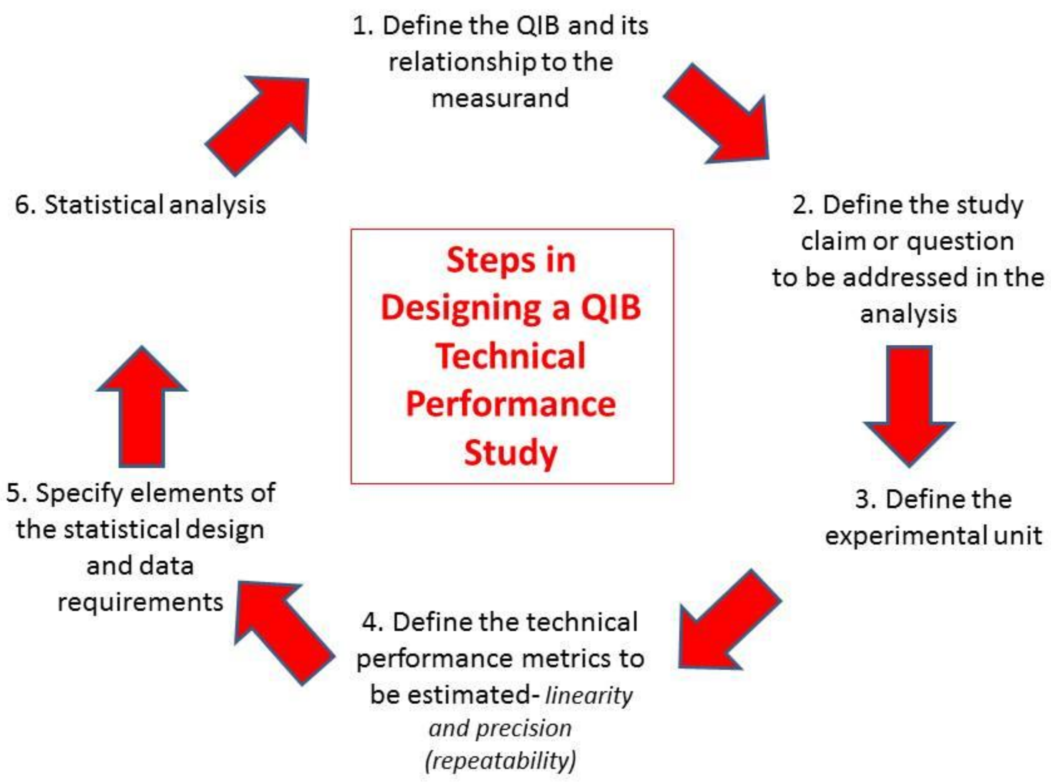


Figure 1-1 Steps in Designing a QIB Technical Performance Study (adapted from [1]).

More recently the European Society of Radiology (ESR) has issued a similar updated ‘Statement on the Validation of Imaging Biomarkers’ which also formalises the requirements of technical performance studies (‘analytical’ validation). Both societies reinforce as vital the demonstration that the detection and quantitative measurement of the QIB are accurate, reproducible and feasible. General considerations for the design of studies assessing the technical performance of a single QIB are outlined in Figure 1. This process includes generating data on limits of detection, limits of quantification, and reference normal values [11]. It also includes assessing both repeatability and reproducibility [2], with both specified by appropriate statistical parameters including the kappa (or weighted kappa), the intra-class correlation coefficient, co-efficient of variation (or standard deviation when mean is close to zero) or the confidence interval of the mean. Evaluations of technical performance and measurement error provide the foundation for establishing whether biomarkers should be reported as continuous variables or categorical and may also provide data to inform selection of rational cut-off values.

Clinical validation (‘qualification’) requires demonstration that the QIB is closely coupled or linked to the presence of the target disease or condition. This critical step in biomarker evaluation provides the basis for biomarker adoption in clinical and research applications, as well as for consideration of biomarker data by regulatory authorities as evidence of drug and device efficacy [12]. For a prognostic biomarker (intended to forecast disease course in the absence of treatment) a correspondence must be shown between the biomarker and the outcome of interest. For a predictive biomarker (intended to forecast disease course in the presence of a specific treatment) the statistical challenges are greater [13, 14]. The most stringent predictive validation is required when a QIB is being assessed for use as a surrogate endpoint. Weir and Walley have stated the challenge in validating a biomarker as a surrogate endpoint: “It is insufficient in the validation of a biomarker as a surrogate endpoint to show that it correlates well with the clinical endpoint … What is required is that effect of treatment on the biomarker correlates well with treatment effect on the final endpoint, so that a valid surrogate endpoint allows correct inference to be drawn regarding the effect of an intervention on the true clinical endpoint of interest” [15]. Only a small subset of biomarkers ever meets criteria for surrogacy. The World Health Organization

(WHO) [16] and Response Evaluation Criteria in Solid Tumors (RECIST) criteria [17] are the only quantitative imaging biomarkers currently accepted as surrogate endpoints in phase III trials of chemotherapy or radiotherapy of solid tumors. There is therefore requirement for a pragmatic approach to biomarker qualification based not only on statistics but also incorporating elements of biological plausibility and practical usefulness [18].

Finally, ‘utilization’ analysis involves the assessment of biomarker performance in the specific context of its proposed use and assesses whether the available evidence from validation and qualification provides sufficient support for the intended use of the biomarker [19]. Challenges for academic radiology in QIB research include prioritising and focusing among a wide set of important objectives, avoiding redundant efforts within a broad array of stakeholders, and maintaining objectivity with respect to basic tenets of standardisation and quality assurance while pursuing high-level technology evaluation [20].

1.3 Magnetic Resonance Imaging Biomarkers in Oncology

1.3.1 Technical Advances:

In parallel with other imaging modalities, ongoing technical advances have enabled magnetic resonance imaging (MRI) to produce exquisite three-dimensional images noninvasively, with high spatial resolution and soft tissue contrast within acceptable clinical time constraints. Stepwise improvements in the magnetic field strengths of clinical MRI platforms have occurred- from <1.0T to 1.0T, 1.0T to 1.5T, and most recently from 1.5T to 3.0T. The main theoretical advantage is superiority in signal-to-noise ratio (SNR) which increases linearly with respect to field strength [21]. Improvements in SNR can potentially be traded for reduced scan time, improved image resolution and/or greater anatomical coverage. Higher field strengths scanners do suffer from several disadvantages due to concomitant magnetic field inhomogeneities and dielectric effects [22] leading to increased motion related image artefact [23]; greater chemical shift artefact [22]; and higher specific absorption rate (SAR) [24]. Overall, a SNR gain of 1.6 to 1.8 times on average can be expected when transitioning from 1.5T to 3.0T imaging [25, 26].

Improving coil technology has also had a major impact on image SNR [27]. In general, the smaller the RF coil diameter the higher the SNR achieved by the coil. SNR advantages of a surface coil can be combined with extended spatial coverage by using an array of coils, based on the concept of the phased-array proposed by Roemer and colleagues [28]. The current generation of MRI scanners exploits this principle to use up to 128 channels further enhanced by use of channel independent coil architecture [29].

Magnetic field gradients are required to spatially localise MRI signals. Higher amplitude and faster switching gradient technology has improved image quality and reduced scanning time [30]. Faster gradients also enable specific pulse sequences such as interleaved spiral imaging, echo planar imaging (EPI) and segmented k-space imaging, which are employed in protocols requiring large field of view coverage within reasonable scan times; body diffusion weighted imaging has predominantly been enabled by EPI acquisition techniques [31].

Parallel imaging further reduces scan duration by using the sensitivity of phased-array coils to facilitate spatial localization, allowing images to be generated using a smaller number of phase encoding steps [32, 33]. This method obviates a fundamental restriction of Fourier imaging (the main contributor to scanning time) whereby only one point at a time can be sampled in k-space [34]. Frequency domain reconstruction techniques such as generalized auto-calibrating partially parallel acquisitions (GRAPPA) and image domain sensitivity encoding (SENSE) are examples of the algorithms being used in parallel imaging on clinical MRI scanners [35, 36]. Whilst these advances have led to faster acquisition of more detailed anatomic images, there is evidence that the robustness of quantitative metrics may be detrimentally affected by an increasing lack of standardisation of imaging parameters, pulse sequences and proprietary post-processing of images across different imaging platforms and manufacturers.

1.3.2 Evaluation of Tumour Size:

When evaluating tumours and more particularly response of tumours to anticancer treatment, a reliable and standardised methodology is essential, both in clinical research and standard clinical care. Therefore, more than 40 years ago, the WHO criteria for tumour evaluation were developed, based on tumour mass assessment on CT or MRI [16]. According to these criteria, the size of the tumour was assessed by two perpendicular diameters. A complete response (CR) was defined as total absence of disease, partial response (PR) as a reduction of at least 50% of the product of these two diameters, progressive disease (PD) as an increase of 25% and stable disease as every response between PR and PD. In 2000, new ‘RECIST’ evaluation criteria were developed, to make tumour evaluation easier [17]. RECIST is based on the sum of one-dimensional measurements of the greatest diameter of the tumour and/ or metastases. CR is defined by the complete absence of disease; PR is defined as more than 30% decrease of the sum of these largest diameters, PD as an increase of more than 20% of the sum of the largest diameters and stable disease as all other outcomes. Several validation studies have been performed for the classic cytotoxic tumour treatments [37-40]. RECIST 1.1 [17] responded to several limitations of the original system, perhaps significantly including metabolic assessment of tumours with FDG-PET-CT.

The main challenge for size based assessment is whether these criteria are sufficient for the evaluation of response to many targeted therapies. The effects of the new therapeutic modalities, such as angiogenesis inhibitors and anti-vascular therapies, are more complex. Necrosis and cavitation without a change in size are frequently observed. Thus, the effect of targeted therapy is often underestimated by using size based RECIST evaluation. For example, single-agent treatment with sorafenib [41] and bevacizumab [42] in metastatic renal cell cancer failed to achieve significant objective response rates according to the RECIST criteria, but did result in a significant increase in progression-free survival (PFS).

In addition, all size-based measurements can be subject to inter-observer error. Furthermore, the optimum measurement technique, necessary magnitude of size change, and appropriate response evaluation criteria are subject to debate. Current criteria require around 65% reduction in total tumour volume for partial response and 40–73% increase in volume for progressive disease (calculated from one-dimensional and two-dimensional measurements). The magnitude of these changes is greater than the 95% confidence intervals (CI) for repeatability on modern scanners, which are about 20% or less by volume (i.e. less than 10% change in one dimension) [43]. Therefore, investigators have questioned whether current criteria for partial response and progressive disease should be less stringent to detect small, but significant, changes in tumour size. An alternative approach in early-phase and late-phase clinical trials is utilising change in size as a continuous variable [44] and reducing the sample size needed in phase II trials.

1.3.3 Cellularity and Diffusion Weighted Imaging (DWI):

Rationale:

Increased and unregulated rates of cell proliferation are a hallmark of aggressive tumours. DWI is sensitive to the microscopic motion of water molecules, and allows for non-invasive characterisation of biological tissues on the basis of their water-diffusion properties [45, 46]. In its simplest form, DWI sequences incorporate two additional intense magnetic-field gradients symmetrically placed either side of a 180° RF pulse within a spin-echo imaging sequence. These additional gradients allow for the measurement of the microscopic motion of water molecules, with the resultant intensity of the diffusion weighted imaging signal dependent on their translational mobility within the direction of the applied gradient. The b value, expressed in s/mm^2, indicates the degree of diffusion weighting. The acquisition of images at multiple b-values allows for calculation of the apparent diffusion coefficient (ADC) in mm^2/s, usually presented as a quantitative parametric map on which a region of interest (ROI) can be drawn. From numerous studies, including work in packed cells [47], it is apparent that measurements using ADC values vary inversely with cell density, thus DWI has been used to infer tumor cellularity [48, 49]. Theoretically, ADC is an attractive marker of response because it should be a reproducible physical constant, independent of magnetic field, scanner, and operator; obviates the need for contrast medium; and is relatively time efficient. There are limitations however regarding reproducibility of ADC at higher field strengths and region of interest (ROI) placement on ADC maps (as a result of poor resolution and resulting partial volume effects) and artefacts resulting from use of fast readout schemes (e.g. echo planar imaging).

Oncological Imaging:

DWI has been used quantitatively to characterise soft tissue masses to determine malignancy. In one study, ADC values of chronic expanding haematomas were found to be significantly higher than malignant soft tissue tumors, providing a good discriminating feature between benign and malignant soft tissue tumors [50]. Further studies have examined the use of ADC values in discriminating malignant and benign lesions in a variety of tumour settings- for example breast [51], hepatic [52] and head and neck lesions [53].

However it is in the assessment of response to therapy that DWI has shown exponential growth in terms of publications. Early responses to various interventions have correlated with increasing ADC values in many studies. Preclinical studies suggest the observed increase in ADC values following therapy appears to be a generalized phenomenon and can stem from a number of cell death mechanisms including apoptosis [54], lytic necrosis, autophagy, and/or mitotic catastrophe [55]. Representative clinical reports illustrating increased ADC in correlation with ultimate tumor response following early intervention include studies involving human brain tumors [56], brain tumor animal models [57], prostate metastases to bone [58], human breast tumors [59], mouse models of breast cancer [60, 61], primary rectal tumors [62], colorectal cancer hepatic metastasis [63, 64] and mouse models of colorectal cancer [65]. The majority of these studies have been performed in combination with either cytoxic drugs and/or radiation.

Prediction of tumour response even before the onset of therapy would have considerable clinical benefit and several studies have shown the potential of DWI to prospectively predict the outcome of therapy in different tumours. Studies involving histological analysis have shown that pre-therapy ADC values are generally low in solid or cellular tumours compared with necrotic or cystic tumours. Higher pre-treatment ADC values generally tend to correlate with poorer response to therapy and prognosis [62–64]. For example, in a recent study characterizing hepatic metastases with DWI, 38 responding and 49 non-responding lesions were evaluated in patients with confirmed metastases originating from gastrointestinal cancers [64]. In these patients, the mean pre-therapy ADC for non-responding lesions was found to be significantly higher than that of non-responding lesions.

1.3.4 Vascularity and Dynamic Contrast Enhanced (DCE) MRI:

Rationale:

When a malignant tumour reaches approximately 1 mm^3 in volume, it can no longer rely on the passive diffusion of metabolites from tissue blood vessels to continue to proliferate. In such circumstances, tumours release various proteins which migrate to existing vasculature and cause vessels to change. Paramount among these proteins are vascular endothelial-derived growth factors (VEGFs), which cause the endothelial cells that form existing vessels to grow, proliferate and migrate up the VEGF concentration gradient [66]. Tumour vessels produced by this angiogenesis are characteristically leaky, fragile and incompletely formed. These differences may be exploited in order to characterize the neovasculature and, indirectly, the state of the tumour. DCE involves the serial acquisition of MR images of a tissue or tumour of interest before, during and after an intravenous injection of a paramagnetic contrast agent, which alters the MR signals from tissues by reducing the inherent nuclear magnetic resonance (NMR) relaxation time (T1; the time of recovery of the tissue's longitudinal magnetisation back towards equilibrium). Analysis of DCE images allows the generation of signal intensity versus time graphs, measurement of maximum enhancement, peak enhancement, and rate of peak enhancement, as well as enhancement gradient or signal enhancement ratio. This time course can also be analysed with an appropriate mathematical pharmacokinetic model allowing physiological parameters to be extracted that relate to, tissue perfusion, microvascular vessel wall permeability and extracellular volume fraction- including the volume transfer constant, K_{trans}, and the rate transfer constant, k_e [67]. Much recent work has focused on use of the parameter K_{trans}, which represents the product of the permeability and the vascular epithelial surface area from which the agent leaks into tissue. Although effort has been made to standardise terminology [68], the exact biological basis and numerical value of every variable differs for a given combination of blood flow/vascular permeability surface area and applied tracer kinetic model. Therefore, unlike standardised plasma or histology biomarkers, in practice K_{trans} refers to a family of related but different biomarkers depending on its precise generation and should be considered when applying DCE to multicentre phase II or phase III trials.

Oncological Imaging:

Some cancers demonstrate typical enhancement kinetics which can be used to distinguish malignant masses from benign lesions or normal tissue, which enhance and washout slowly [69]. Thus, malignant lesions tend to have high rate constants or transfer constants of exchange reflected as bright colours on parametric images, whereas benign or normal tissues show minimal or no change in colour. For example, DCE can be used to help distinguish tumours from benign processes in breast cancer [70-73]. Knopp et al [74] reported a correlation between vascular permeability as measured by DCE and staining of breast tumors for VEGF. Additionally, DCE MRI has demonstrated some utility prostate tumour localization and staging additive to T2 weighted images [75, 76].

DCE has been investigated in various studies as an early indicator of tumour response to therapy in cancers of the breast, cervix, bowel, liver, lung, and head and neck [77-81]. However, it is difficult to compare these findings due to differences in techniques and the population studied. Similarly, these studies used diverse endpoints to correlate with DCE parameters, such as known prognostic factors like histological grade and lymph-node involvement, or eventual survival characteristics. The timing of DCE in these studies is also variable, with parameters measured before, during, and at the end of therapy.

Several studies have assessed the predictive role of pre-therapy DCE parameters, such as relative signal intensity or peak enhancement [79], while some have reported that changes in enhancement and signal intensity in the very early stages of therapy have been shown to be associated with improved local tumour control [83]. George and colleagues [79] used DCE to quantify tumour permeability (as k_{trans}) in rectal tumours and showed that patients with a higher pre-treatment permeability responded better to chemoradiotherapy than those with a lower permeability. Emerging evidence indicates that multiparametric analysis of DCE-MRI data offers greater insight into the mechanism of drug action than studies measuring a single parameter, such as K_{trans} [80].

Post-therapy DCE might also be helpful in detecting residual disease, and for predicting early recurrence and identifying those who might benefit from salvage treatment. This predictive value of DCE MRI measurements after the completion of therapy has been described in several studies [81, 82]. At the end of a treatment regimen, high enhancement values might be attributed to the presence of persistent tumour, and are associated with increased risk of recurrence and poor survival. DCE currently remains a promising biomarker for assessing tumour angiogenesis and the effect of anti-angiogenic therapy [84].

1.3.4 Biochemical Composition and Magnetic Resonance Spectroscopy (MRS):

Rationale:

MRS is able to provide chemical information about tissue metabolites. Whereas conventional MRI illustrates gross anatomy by exciting the ^1H nuclei attached to tissue water; MRS resolves ^1H bound to different metabolites through differences in their electron shielding causing a spread of resonant frequencies. A number of nuclei can be assessed, but ^1H is most commonly used, followed by phosphorus (^{31}P). Results are displayed on a spectrum, which shows a series of peaks corresponding to different metabolites known to have specific chemical shifts. A signal voxel can be defined from which spectroscopic information can be gathered. Alternatively more complex imaging can provide spectroscopic information at multiple voxels (e.g. chemical shift imaging).

Oncological Imaging:

Increased levels of choline (considered a marker for cell proliferation) in association with decreased levels of creatine (considered a marker for energetic processes) and decreased levels of N-acetyl aspartate (considered a neuronal marker) have been found in the evaluation of brain neoplasms. In combination, it is possible to differentiate, for example, low-grade from high-grade gliomas [85, 86]. In the evaluation of breast lesions, there is an association with choline peak which may be detected in malignant lesions whilst choline levels are low or negligible in benign lesions or in healthy breast tissues. Prostate MRS obtains metabolic data based on the relative concentration of endogenous metabolites such as choline, creatine and citrate. The absolute values of citrate, creatine and choline levels contribute in the identification of areas suspicious for malignancy and the choline-creatine/citrate ratio correlates with the Gleason score allowing for a non-invasive selection of areas for prostate biopsy and evaluation of the tumour grade [87].

Whilst most MRS studies have focused on tumour diagnosis, several reports have highlighted its potential role in assessing response to therapy, including both preclinical [88] and clinical studies [89-91]. For example, Manton and colleagues [89] used MRS to predict response in 34 women receiving neoadjuvant chemotherapy for breast cancer. Imaging was done before the start of treatment and then again after two cycles of therapy. Early changes in the water:fat peak ratios after just two cycles of chemotherapy accurately predicted final tumour volume response in 69% of women whilst maintaining 100% specificity and PPV.

1.3.5 Hypoxia and T2*/Blood Oxygen Level Dependent (BOLD) MRI:

Rationale:

Hypoxia results in changes in patterns of gene expression that alter the malignant potential of tumours by leading to more aggressive survival traits. As a result, hypoxic cancers are difficult to treat with radiation [92] or cytotoxic chemotherapy. The presence of hypoxia within human tumours before starting treatment has been observed in a variety of tumour types including squamous cell carcinomas, gliomas, adenocarcinomas (breast & pancreas) and in sarcomas. Hypoxia-mediated aggressive behaviour of cancer cells and their resistance to therapy is orchestrated by the heterodimeric transcription factor, hypoxia inducible factor-1 alpha (HIF-1α), via a number of molecular events required for the adaptation of tumour cells to hypoxia [93]. Imaging may be a good way of non-invasively selecting cancer patients who would benefit from treatments that overcome, circumvent or take advantage of the presence of hypoxia.

BOLD (or intrinsic susceptibility-weighted) MRI relies on the paramagnetic property of deoxyhaemoglobin, which creates susceptibility variations in the magnetic field (or microscopic field gradients), which in turn increase the transverse relaxation rate R2* (1/T2*) of water in blood and the tissue surrounding blood vessels. An increase in the deoxyhaemoglobin concentration (that is, hypoxia) leads to a decrease in the signal intensity on the T2* image and a faster R2* [94]. Deoxyhaemoglobin therefore acts as an intrinsic BOLD contrast agent for imaging tissue hypoxia. Specific gradient-recalled echo (GRE) sequences are required to detect changes in R2*. The major drawback of the BOLD-based approach is that T2* is also effected by many other parameters, including macroscopic B0 field inhomogeneities, the transverse relaxation parameter T2, water diffusion, and the characteristics of the blood vessel network. By observing the spatial and temporal variations of the MR signal in response to an external stimulus, these static artefacts (field inhomogeneity and T2 variation) are minimized. Therefore, variations in R2* have been evaluated in tumour models using inhaled carbogen (95% O_2: 5% CO_2) to intensify the otherwise small changes in signal intensity: the CO_2 induces vasodilation and the O_2 tension is high with 95% O_2 so that subtracted images with and without carbogen reveal regions of hypoxia where signal change is greatest [95]. Unfortunately, the hyperventilation induced by breathing carbogen in humans is poorly tolerated so reliance has been largely on R2* measurements whilst breathing air and/or 100% oxygen.

Oncological Imaging:

A study in breast cancer patients has shown R2* values to be significantly lower in tumour than normal breast parenchyma prior to the commencement of chemotherapy [96], suggesting that breast tumours are less hypoxic than normal breast tissue, possibly because of their high vascularity. This contrasts with other published data in prostate cancer [97], where R2* is increased, indicating increased hypoxia in these tumours. The increased R2* in normal breast tissue has also been related to the fibro-collagenous ligaments of Cooper, which maintain normal breast structural integrity and contribute to higher R2* values. In responders following treatment, the R2* value has been shown to increase, likely as a result of decreased blood flow; however, in this one published study this parameter was not as useful when compared with other DCE-MRI parameters or even morphological parameters such as tumour size in indicating response [96].

1.3.6 Biophysical properties and T1 Relaxation Mapping MRI:

Rationale:

The T1 relaxation time, also known as the spin-lattice relaxation time, is a measure of how quickly the net magnetisation vector (NMV) recovers to its ground state in the direction of the static magnetic field (B0). The return of excited nuclei from the high energy state to the low energy or ground state is associated with loss of energy to the surrounding nuclei. Nuclear magnetic resonance was originally used to examine solids in the form of lattices, hence the name "spin-lattice" relaxation. At a time equal to the T1 value of a tissue, the signal will recover to 63% of its initial value after the RF pulse has been applied. T1 relaxation is fastest when the motion of the nucleus (rotations and translations or "tumbling rate") matches that of the Larmor frequency. As a result, T1 relaxation is dependent on the main magnetic field strength that specifies the Larmor frequency.

Quantitation of T1 may be used to characterise normal and pathological tissue and there is increasing interest in its use in producing robust DCE parameters. In vivo T1 quantitation presents several difficulties (most notably respiratory artefact and field of view requirement). Several techniques have been explored; the most commonly used being inversion recovery (IR; which remains the reference standard method), multiple flip angle (MFA), and Look-Locker (LL) techniques. While there are many techniques for T1 mapping [98], there is also a large variation of reported T1 values for the same tissues and field strengths, making it difficult to standardize protocols largely due to incomplete spoiling and inaccuracy in the RF field estimation. In the most comprehensive recent review of T1 estimation methods, it was recommended that sites perform validation in vivo; calibrating their T1 maps with the IR reference technique [99].

Oncological Imaging:

The demonstration that tumour T1 values were higher than benign tissue predates the advent of clinical MRI scanners when Damadian first reported alteration of T1 relaxation times in cancerous tumors in 1971 using nuclear magnetic resonance relaxometry (NMR) [100]. In the early 1980s, vast amounts of experimental quantitative NMR measurements of biological tissues (animal and human, normal and diseased) were published [101-103]. Extrapolating to MR imaging this "NMR-relaxometry-disease-signature" concept was logical and intellectually appealing. Indeed, many seminal papers on quantitative MRI pulse sequence design and theory were published only a few years after the advent of clinical MR imaging. However, the challenges of accurate in vivo T1 estimation in a clinically feasible time frame were (and remain) problematic.

More recently, animal tumour models and in vivo human studies using a variety of methods show that the T1 of tumours was greater than in normal tissue and was considered to reflect predominantly the extracellular space [104-106], with a larger interstitial compartment giving a longer (higher) T1. Conversely in treated tumours, a low tumour T1 has been correlated to increased necrosis [105], low water content, high levels of soluble protein [104], and low proliferation. However, in four different quantitative T1 studies, only decreases were detected in response to successful therapy with radiation or cytotoxics [107-110]. McSheehy et al (2009) recently used animal tumour models to show five anticancer drugs (with different mechanisms of action) caused a rapid decrease in T1 that often preceded and was strongly correlated with the change in tumour size suggesting that a fractional change in T1 (ΔT1) was a useful generic early-response marker. Furthermore, ΔT1 did not occur in tumours resistant to that drug, and ΔT1 was positively correlated with proliferation markers (Ki67 and choline), suggesting utility as a generic marker of early response [110].

1.3.7 Tissue heterogeneity and Post-processing/Textural Analysis MRI:

Rationale:

Tumours are heterogeneous both on genetic and histopathological levels with intratumoral spatial variation in the cellularity, angiogenesis, extravascular extracellular matrix, and areas of necrosis. Tumours with high intratumoral heterogeneity have been shown to have poorer prognosis, which could be secondary to intrinsic aggressive biology or treatment resistance

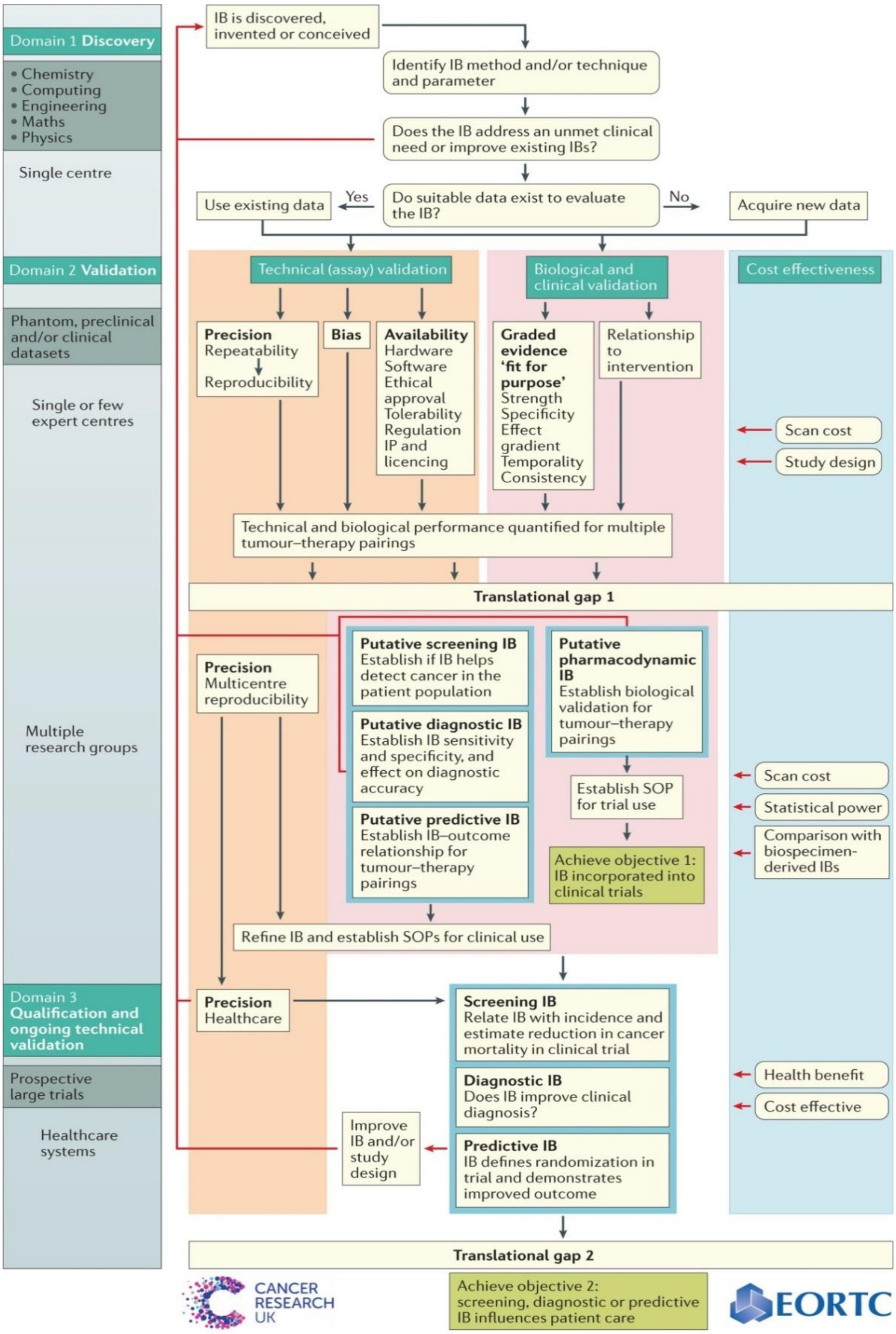
[112, 113]. A non-invasive method of assessing the heterogeneity within a tumour therefore might be of clinical benefit. Texture analysis refers to a variety of mathematical methods that can be used to evaluate the grey-level intensity and position of the pixels within an image to derive so-called 'texture features' that provide a measure of intralesional heterogeneity [113]. Different methods have been applied, including statistical-, model-, and transform-based methods [114-116]. Statistical-based techniques have been most commonly applied and describe the distribution and relationships of grey level values in the image. Three orders of parameters are described in statistical-based texture analysis. First-order statistics relate to grey-level frequency distribution within the region of interest, which can be obtained from the histogram of pixel intensities. There has been exponential growth of studies examining the use of quantitative parameters derived from medical images into the field of radiomics; although many of the concepts of image feature extraction have been around for decades, research output in the field has grown exponentially, with over 1500 publications in 2020 containing the term radiomics [117].

Oncological Imaging:

Textural analysis of MR images remains an area of active research with several theoretical advantages- most importantly the possibility of gaining additional quantitative information from clinically acquired images without the need for prior standardisation. Recent studies have used MR textural analysis for lesion detection, classification, treatment response-evaluation and prediction for example in breast, brain, and rectal cancer [118-120]. For example, Eliat et al [118] found that texture features of DCE MRI may distinguish glioblastoma multiforme from malignant glioneuronal tumors. More recently, De Cecco et al [120] showed texture parameters derived from T2 weighted images of rectal cancer have the potential to act as imaging biomarkers of tumoral response to neoadjuvant chemoradiation.

1.4 Conclusions

Researchers and clinicians from across the biomedical spectrum are increasingly demanding QIBs for incorporation into algorithmic decision making. The imaging community is responding by developing QIBs in numerous modalities across a broad set of functional areas. MRI based QIBs are particularly attractive due to numerous methods for obtaining quantitative data though current development is hampered by a lack of stringent validation and widespread consensus and standardisation compounded by lack of training among radiologists in advanced clinical research techniques. QIB development requires painstaking evaluation with sequential attention to analytical validation, qualification, and utilization of novel techniques and metrics. This has been acknowledged by funding bodies and Figure 2 shows an updated imaging biomarker ‘roadmap’ developed by Cancer Research UK and endorsed by EORTC [80].



Nature Reviews | Clinical Oncology

Figure 1-2 Representative 'roadmap' for development of imaging biomarkers in cancer [80].

1.5 REFERENCES

1. Sullivan DC, Obuchowski NA, Kessler LG, Raunig DL, Gatsonis C, Huang EP, Kondratovich M, McShane LM, Reeves AP, Barboriak DP, Guimaraes AR, Wahl RL. Metrology Standards for Quantitative Imaging Biomarkers. Radiology 2015 Dec; 277(3): 813-25.

2. Kessler LG, Barnhart HX, Buckler AJ, et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. Stat Methods Med Res 2015; 24(1): 9–26.

3. Prescott JW. Quantitative Imaging Biomarkers: The Application of Advanced Image Processing and Analysis to Clinical and Preclinical Decision Making. J Digit Imaging 2013 26(1): 97–108.

4. Smith JJ, Sorensen AG, Thrall JH. Biomarkers in imaging: realizing radiology's future. Radiology 2003 227(3): 633–638.

5. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001 69(3): 89-95.

6. Quantitative Imaging Biomarkers Alliance. http://rsna.org/QIBA.aspx.

7. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res 2015 24(1): 27–67.

8. Obuchowski NA, Reeves AP, Huang EP, et al. Quantitative imaging biomarkers: a review of statistical methods for computer algorithm comparisons. Stat Methods Med Res 2015 24(1): 68–106.

9. Obuchowski NA, Barnhart HX, Buckler AJ, et al. Statistical issues in the comparison of quantitative imaging biomarker algorithms using pulmonary nodule volume as an example. Stat Methods Med Res 2015 24(1): 107–140.

10. European Society of Radiology (ESR). ESR Statement on the Validation of Imaging Biomarkers. Insights Imaging. 2020;11(1):76.

11. Abramson RG, Yankeelov TE. Imaging biomarkers and surrogate endpoints in oncology clinical trials. In: Luna, A, Vilanova JC, Hygino D, Cruz LC, Rossi SE editors. Functional maging in Oncology. Springer Heidelberg: 2014.

12. Katz R. Biomarkers and surrogate markers: an FDA perspective. NeuroRx: the journal of the American Society for Experimental NeuroTherapeutics. 2004; 1(2):189–95.

13. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. J Clin Oncol. 2009 27(24):4027–34.

14. Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points- the challenge of statistical validation. Nature Reviews Clinical oncology 2010 7(6):309–17.

15. Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. Stat Med 2006 25(2): 183–203.

16. World Health Organization. WHO handbook for reporting results of cancer treatment. Geneva: World Health Organization, 1979.

17. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009 45(2): 228-47.

18. Green E, Yothers G, Sargent DJ. Surrogate endpoint validation: statistical elegance versus clinical relevance. Statistical methods in medical research. 2008 17(5):477–86.

19. Institute of Medicine. Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease. National Academies Press; Washington DC: 2010.

20. Abramson RG, Burton KR, Yu JP, Scalzetti EM, Yankeelov TE, Rosenkrantz AB, Mendiratta-Lala M, Bartholmai BJ, Ganeshan D, Lenchik L, Subramaniam RM. Methods and challenges in quantitative imaging biomarker development. Acad Radiol. 2015 22(1):25-32.

21. Redpath TW. Signal---to---noise ratio in MRI. Br J Radiol 1998 71: 704-707.

22. Chang KJ, Kamel IR, Macura KJ, Bluemke DA. 3.0T MR imaging of the abdomen: comparison with 1.5 T. Radiographics 28(7):1983-98.

23. Bernstein M, Huston J, Ward H. Imaging artifacts at 3.0T. J Magn Reson Imaging 2006 24: 735-46.

24. Schmitt F, Grosu D, Mohr C, et al. 3 Tesla MRI: successful results with higher field strengths. Radiologe 2004 44: 31-47.

25. Merkle EM, Dale BM. Abdominal MRI at 3.0T: the basics revisited. AJR Am J Roentgenol 2006 186: 1524-1532.

26. de Bazelaire CMJ, Duhamel GD, Rofsky NM, Alsop DC. MR imaging relaxation times of abdominal and pelvic tissues measured in vivo at 3.0T: preliminary results. Radiology 2004 230: 652-659.

27. Blamire AM. The technology of MRI-the next 10 years? Br J Radiol 2008 81: 601-617.

28. Roemer PB, Edelstein WA, Hayes CE, Souza SP, Mueller OM. The NMR phased array. Magn Reson Med 1990 16: 192-225.

29. Fujita H. New horizons in MR technology: RF coil designs and trends. Magn Reson Med Sci 2007 6: 29-42.

30. Reeder SB, McVeigh ER. The effect of high performance gradients on fast gradient echo imaging. Magn Reson Med 1994 32: 612-621.

31. Poustchi-Amin M, Mirowitz SA, Brown JJ, McKinstry RC, Li T, Technology T. Principles and Applications of Echo-planar Imaging : A Review for the General Radiologist. Radiographics 2001 21: 767-779.

32. Deshmane A, Gulani V, Griswold MA, Seiberlich N. Parallel MR imaging. J Magn Reson Imaging 2012 36: 55-72.

33. Blaimer M, Breuer F, Mueller M, Heidemann RM, Griswold MA, Jakob PM. SMASH, SENSE, PILS, GRAPPA: how to choose the optimal method. Top Magn Reson Imaging 2004 15:223-236.

34. Lauterbur PC. Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. Nature 1973 242(5394): 190-191.

35. Griswold MA, Jakob PM, Heidemann RM, et al. Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA). Magn Reson Med 2002 47: 1202-1210.

36. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: Sensitivity encoding for fast MRI. Magn Reson Med 1999 42: 952-962.

37. Erasmus JJ, Gladish GW, Broemeling L, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. J Clin Oncol 2003 21: 2574–82.

38. Grossi F, Belvedere O, Fasola G, et al. Tumor measurements on computed tomographic images of non-small cell lung cancer were similar among cancer professionals from different specialties. J Clin Epidemiol 2004 57: 804–8.

39. Trillet-Lenoir V, Freyer G, Kaemmerlen P, et al. Assessment of tumour response to chemotherapy for metastatic colorectal cancer: accuracy of the RECIST criteria. Br J Radiol 2002 75: 903–8.

40. Prasad SR, Saini S, Sumner JE, et al. Radiological measurement of breast cancer metastases to lung and liver: comparison between WHO (bidimensional) and RECIST (unidimensional) guidelines. J Comput Assist Tomogr 2003 27: 380–4.

41. Escudier B, Eisen T, Stadler WM, et al. Sorafenib in advanced clear-cell renal cell carcinoma. N Engl J Med 2007 356: 125–34.

42. Yang JC, Haworth L, Sherry RM, et al. A randomized trial of bevacizumab, an anti-vascular endothelial growth factor antibody, for metastatic renal cancer. N Engl J Med 2003 349: 427–34.

43. Kostis WJ, Yankelevitz DF, Reeves AP, Fluture SC, Henschke CI. Small pulmonary nodules: reproducibility of three-dimensional volumetric measurement and estimation of time to follow-up CT. Radiology 2004 231: 446–52.

44. Karrison TG, Maitland ML, Stadler WM, Ratain MJ. Design of phase II cancer trials using a continuous endpoint of change in tumor size: application to a study of sorafenib and erlotinib in non-small-cell lung cancer. J Natl Cancer Inst 2007 99: 1455–61.

45. Chenevert TL, Stegman LD, Taylor JM, et al. Diffusion magnetic resonance imaging: an early surrogate marker of therapeutic efficacy in brain tumors. J Natl Cancer Inst 2000 92: 2029–36.

46. Hamstra DA, Rehemtulla A, Ross BD. Diffusion magnetic resonance imaging: a biomarker for treatment response in oncology. J Clin Oncol 2007; 25: 4104–09.

47. Anderson AW, Zhong J, Petroff OAC, Szafer A, Ransom BR, Prichard JW, et al. Effects of osmotically-driven cell volume changes on diffusion weighted imaging of the rat optic nerve. Magn Reson Med 1995 35: 162–7.

48. Sugahara T, Korogi Y, Kochi M, Ikushima I, Shigematu Y, Hirai T, et al. Usefulness of diffusion-weighted MRI with echo-planar technique in the evaluation of cellularity in gliomas. J Magn Reson Imaging 1999 9: 53–60.

49. Lyng H, Haraldseth O, Rofstad EK. Measurement of cell density and necrotic fraction in human melanoma xenografts by diffusion weighted magnetic resonance imaging. Magn Reson Med 2000 43: 828–36.

50. Oka K, Yakushiji T, Sato H, Yorimitsu S, Hayashida Y, Yamashita Y, Mizuta H. Ability of diffusion-weighted imaging for the differential diagnosis between chronic expanding hematomas and malignant soft tissue tumors. J Magn Reson Imaging. 2008 28(5): 1195–200.

51. Partridge SC, Mullins CD, Kurland BF, Allain MD, DeMartini WB, Eby PR, Lehman CD. Apparent diffusion coefficient values for discriminating benign and malignant breast MRI lesions: effects of lesion type and size. AJR Am J Roentgenol. 2010 194(6): 1664-73.

52. Miller FH, Hammond N, Siddiqi AJ, Shroff S, Khatri G, Wang Y, Merrick LB, Nikolaidis P. Utility of diffusion-weighted MRI in distinguishing benign and malignant hepatic lesions. Journal of Magnetic Resonance Imaging. 2010 32(1): 138-47.

53. Srinivasan A, Dvorak R, Perni K, Rohrer S, Mukherji SK. Differentiation of benign and malignant pathology in the head and neck using 3T apparent diffusion coefficient values: early experience. American Journal of Neuroradiology. 2008 29(1): 40-4.

54. Kim H, Morgan DE, Zeng H, Grizzle WE, Warram JM, Stockard CR, et al. Breast tumor xenografts: diffusion-weighted MR imaging to assess early therapy with novel apoptosis-inducing anti-DR5 antibody. Radiology 2008 248(3): 844–51.

55. Morse DL, Galons JP, Payne CM, Jennings DL, Day S, Xia G, et al. MRI-measured water mobility increases in response to chemotherapy via multiple cell-death mechanisms. NMR Biomed 2007 20(6): 602–14.

56. Moffat BA, Chenevert TL, Lawrence TS, Meyer CR, Johnson TD, Dong Q, et al. Functional diffusion map: a noninvasive MRI biomarker for early stratification of clinical brain tumor response. Proc Natl Acad Sci U S A 2005 102: 5524–9.

57. Hall DE, Moffat BA, Stojanovska J, Johnson TD, Li Z, Hamstra DA, et al. Therapeutic efficacy of DTI-015 using diffusion magnetic resonance imaging as an early surrogate marker. Clin Cancer Res 2004 10(23): 7852–9.

58. Lee KC, Sud S, Meyer CR, Moffat BA, Chenevert TL, Rehemtulla A, et al. An imaging biomarker of early treatment response in prostate cancer that has metastasized to the bone. Cancer Res 2007 67(8): 3524–8.

59. Yankeelov TE, Lepage M, Chakravarthy A, Broome EE, Niermann KJ, Kelley MC, et al. Integration of quantitative DCE-MRI and ADC mapping to monitor treatment response in human breast cancer: initial results. Magn Reson Imaging 2007 25: 1–13.

60. Morse DL, Galons JP, Payne CM, Jennings DL, Day S, Xia G, et al. MRI-measured water mobility increases in response to chemotherapy via multiple cell-death mechanisms. NMR Biomed 2007 20(6): 602–14.

61. Kim H, Morgan DE, Zeng H, Grizzle WE, Warram JM, Stockard CR, et al. Breast tumor xenografts: diffusion-weighted MR imaging to assess early therapy with novel apoptosis-inducing anti-DR5 antibody. Radiology 2008;248(3): 844–51.

62. Dzik-Jurasz A, Domenig C, George M, Wolber J, Padhani A, Brown G, et al. Diffusion MRI for prediction of response of rectal cancer to chemoradiation. Lancet 2002 360(9329): 307–8.

63. Koh DM, Scurr E, Collins D, Kanber B, Norman A, Leach MO, et al. Predicting response of colorectal hepatic metastasis: value of pretreatment apparent diffusion coefficients. AJR Am J Roentgenol 2007 188: 1001–8.

64. Cui Y, Zhang XP, Sun YS, Tang L, Shen L. Apparent diffusion coefficient: potential imaging biomarker for prediction and early detection of response to chemotherapy in hepatic metastases. Radiology 2008 248: 894–900.

65. Seierstad T, Røe K, Olsen DR. Noninvasive monitoring of radiationinduced treatment response using proton magnetic resonance spectroscopy and diffusion-weighted magnetic resonance imaging in a colorectal tumor model. Radiother Oncol 2007 85(2): 187–94.

66. Folkman J. Angiogenesis in cancer, vascular, rheumatoid and other disease. Nat Med 1995 1: 27–31.

67. Yankeelov TE, Gore JC. Dynamic contrast enhanced magnetic resonance imaging in oncology: theory, data acquisition, analysis, and examples. Curr Med Imaging Rev 2007;3:91–107.

68. Tofts PS, Brix G, Buckley DL, et al. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. J Magn Reson Imaging 1999; 10: 223–32.

69. Flickinger FW, Allison JD, Sherry RM, Wright JC. Differentiation of benign from malignant breast masses by time-intensity evaluation of contrast enhanced MRI. Magn Reson Imaging 1993 11: 617–620.

70. Tilanus-Linthorst MM, Obdeijn IM, Bartels KC, de Koning HJ, Oudkerk M. First experiences in screening women at high risk for breast cancer with MR imaging. Breast Cancer Res Treat 2000 63: 53–60.

71. Tilanus-Linthorst MM, Bartels CC, Obdeijn AI, Oudkerk M. Earlier detection of breast cancer by surveillance of women at familial risk. Eur J Cancer 2000 36: 514–519.

72. Kuhl CK, Schmutzler RK, Leutner CC, et al. Breast MR imaging screening in 192 women proved or suspected to be carriers of a breast cancer susceptibility gene: preliminary results. Radiology 2000 215: 267–279.

73. Brown J, Buckley D, Coulthard A, et al. Magnetic resonance imaging screening in women at genetic risk of breast cancer: imaging and analysis protocol for the UK multicentre study. UK MRI Breast Screening Study Advisory Group. Magn Reson Imaging 2000 18: 765–776.

74. Knopp MV, Weiss E, Sinn HP, et al. Pathophysiologic basis of contrast enhancement in breast tumors. J Magn Reson Imaging 1999 10: 260–266.

75. Jager GJ, Ruijter ET, van de Kaa CA, et al. Dynamic turboFLASH subtraction technique for contrast-enhanced MR imaging of the prostate: correlation with histopathologic results. Radiology 1997 203: 645–652.

76. Huch Boni RA, Boner JA, Lutolf UM, et al. Contrast-enhanced endorectal coil MRI in local staging of prostate carcinoma. J Comput Assist Tomogr 1995 19: 232–237.

77. Pickles MD, Lowry M, Manton DJ, Gibbs P, Turnbull LW. Role of dynamic contrast enhanced MRI in monitoring early response of locally advanced breast cancer to neoadjuvant chemotherapy. Breast Cancer Res Treat 2005 91: 1–10.

78. Gong QY, Brunt JN, Romaniuk CS, et al. Contrast enhanced dynamic MRI of cervical carcinoma during radiotherapy: early prediction of tumour regression rate. Br J Radiol 1999 72: 1177–84.

79. George ML, Dzik-Jurasz AS, Padhani AR, et al. Non-invasive methods of assessing angiogenesis and their value in predicting response to treatment in colorectal cancer. Br J Surg 2001 88: 1628–36.

80. O'Connor JP, Jackson A, Parker GJ, Roberts C, Jayson GC. Dynamic contrast-enhanced MRI in clinical trials of antivascular therapies. Nat Rev Clin Oncol. 2012 Feb 14;9(3):167-77.

81. Ohno Y, Nogami M, Higashino T, et al. Prognostic value of dynamic MR imaging for non-small-cell lung cancer patients after chemoradiotherapy. J Magn Reson Imaging 2005 21: 775–83.

82. Tomura N, Omachi K, Sakuma I, et al. Dynamic contrast-enhanced magnetic resonance imaging in radiotherapeutic efficacy in the head and neck tumors. Am J Otolaryngol 2005 26: 163–67.

83. Loncaster JA, Carrington BM, Sykes JR, et al. Prediction of radiotherapy outcome using dynamic contrast enhanced MRI of carcinoma of the cervix. Int J Radiat Oncol Biol Phys 2002 54: 759–67.

84. Padhani AR, Leach MO. Antivascular cancer treatments: functional assessments by dynamic contrast-enhanced magnetic resonance imaging. Abdom Imaging 2005 30: 324–41.

85. Porto L, Kieslich M, Franz K, et al. MR spectroscopy differentiation between high and low grade astrocytomas: a comparison between paediatric and adult tumours. Eur J Paediatr Neurol. 2011 15: 214–21.

86. Kauppinen RA, Peet AC. Using magnetic resonance imaging and spectroscopy in cancer diagnostics and monitoring: preclinical and clinical approaches. Cancer Biol Ther. 2011 12: 665–79.

87. Boonsirikamchai P, Choi S, Frank SJ, et al. MR imaging of prostate cancer in radiation oncology: what radiologists need to know. Radiographics. 2013 33: 741–61.

88. Zhao M, Pipe JG, Bonnett J, Evelhoch JL. Early detection of treatment response by diffusion-weighted 1H-NMR spectroscopy in a murine tumour in vivo. Br J Cancer 1996 73: 61–64.

89. Manton DJ, Chaturvedi A, Hubbard A, et al. Neoadjuvant chemotherapy in breast cancer: early response prediction with quantitative MR imaging and spectroscopy. Br J Cancer 2006 94: 427–35.

90. Preul MC, Caramanos Z, Villemure JG, et al. Using proton magnetic resonance spectroscopic imaging to predict in vivo the response of recurrent malignant gliomas to tamoxifen chemotherapy. Neurosurgery 2000 46: 306–18.

91. Shukla-Dave A, Poptani H, Loevner LA, et al. Prediction of treatment response of head and neck cancers with P-31 MR spectroscopy from pretreatment relative phosphomonoester levels. Acad Radiol 2002 9: 688–94.

92. Vaupel P. The role of hypoxia induced factors in tumor progression. Oncologist 2004 9: 10–17.

93. Powis G, Kirkpatrick L. Hypoxia inducible factor-1{alpha} as a cancer drug target. Mol Cancer Ther 2004 3: 647–654.

94. Robinson SP, Howe FA, Rodrigues LM, Stubbs M, Griffiths JR: Magnetic resonance imaging techniques for monitoring changes in tumor oxygenation and blood flow. Semin Radiat Oncol 1998 8: 197-207.

95. Taylor NJ, Baddeley H, Goodchild KA, Powell ME, Thoumine M, Culver LA, Stirling JJ, Saunders MI, Hoskin PJ, Phillips H, Padhani AR, Griffiths JR. BOLD MRI of human tumor oxygenation during carbogen breathing. J Magn Reson Imaging. 2001 Aug;14(2):156-63.

96. Li SP, Taylor NJ, Makris A, Ah-See ML, Beresford MJ, Stirling JJ, d'Arcy JA, Collins DJ, Padhani AR. Primary human breast adenocarcinoma: imaging and histologic correlates of intrinsic susceptibility-weighted MR imaging before and during chemotherapy. Radiology 2010 257:643-652.

97. Alonzi R, Padhani AR, Maxwell RJ, Taylor NJ, Stirling JJ, Wilson JI, d'Arcy JA, Collins DJ, Saunders MI, Hoskin PJ. Carbogen breathing increases prostate cancer oxygenation: a translational MRI study in murine xenografts and humans. Br J Cancer 2009 100: 644-648.

98. Kingsley PB. Methods of measuring spin-lattice (T1) relaxation times: an annotated bibliography. Concepts Magn Reson 1999 11: 243–276.

99. Stikov N, Boudreau M, Levesque IR, Tardif CL, Barral JK, Pike GB. On the accuracy of T1 mapping: searching for common ground. Magn Reson Med 2015 73(2): 514-22.

100. Damadian R. Tumor detection by nuclear magnetic resonance. Science 1971 171(3976): 1151-3.

101. Bottomley PA. In vivo tumor discrimination in a rat by proton nuclear magnetic resonance imaging. Cancer Res 1979 39(2 Pt 1): 468-70.

102. Bottomley PA, Hart HR, Edelstein WA, Schenck JF, Smith LS, Leue WM, Mueller OM, Redington RW. NMR imaging/spectroscopy system to study both anatomy and metabolism. Lancet 1983 2(8344): 273-4.

103. Brant-Zawadzki M, Bartkowski HM, Ortendahl DA, Pitts LH, Hylton NM, Nishimura MC, Crooks LE. NMR in experimental cerebral edema: value of T1 and T2 calculations. AJNR Am J Neuroradiol 1984 5(2): 125-9.

104. Jakobsen I, Kaalhus O, Lyng H, Rofstad EK. Detection of necrosis in human tumour xenografts by proton magnetic resonance imaging. Br J Cancer 1995 71: 456-61.

105. Gambarota G, Veltien A, van Laarhoven H, et al. Measurements of T1 and T2 relaxation times of colon cancer metastases in rat liver at 7 T. Magn Res Mat Phys Biol Med 2004 17: 281-7.

106. Thomsen C, Sørensen PG, Karle H, Christoffersen P, Henriksen O. Prolonged bone marrow T1-relaxation in acute leukaemia. In vivo tissue characterization by magnetic resonance imaging. Magn Reson Imag 1987 5:251-7.

107. Jensen KE, Sørensen PG, Thomsen C, et al. Magnetic resonance imaging of the bone marrow in patients with acute leukemia during and after chemotherapy. Acta Radiologica 1990 31: 361-9.

108. Santoni R, Bucciolini M, Chiostrini C, Cionini L, Renzi R. Quantitative magnetic resonance imaging in cervical carcinoma: a report on 30 cases. Br J Radiol 1991;64:498-504.

109. Smith SR, Roberts N, Percy DF, Edwards RH. Detection of bone marrow abnormalities in patients with Hodgkin's disease by T1 mapping of MR images of lumbar vertebral bone marrow. Br J Cancer 1992 65: 246-51.

110. McSheehy PM, Weidensteiner C, Cannet C, Ferretti S, Laurent D, Ruetz S, Stumm M, Allegrini PR. Quantified tumor t1 is a generic early-response imaging biomarker for chemotherapy reflecting cell viability. Clin Cancer Res. 2010 Jan 1;16(1):212-25.

111. Hockel M, Knoop C, Schlenger K et al. Intratumoral pO2 predicts survival in advanced cancer of the uterine cervix. Radiother Oncol 1993 26(1): 45-50.

112. Yang Z, Tang LH, Klimstra DS. Effect of tumor heterogeneity on the assessment of Ki67 labeling index in well differentiated neuroendocrine tumors metastatic to the liver: implications for prognostic stratification. Am J Surg Pathol 2011 35(6): 853-860.

113. Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. Clin Radiol 2004 59(12): 1061-1069.

114. Al-Kadi OS, Watson D. Texture analysis of aggressive and nonaggressive lung tumor CE CT images. IEEE Trans Biomed Eng 2008 55(7): 1822-1830.

115. Ganeshan B, Miles KA, Young RC, Chatwin CR. Hepatic entropy and uniformity: additional parameters that can potentially increase the effectiveness of contrast enhancement during abdominal CT. Clin Radiol 2007 62(8): 761-768.

116. Brown RA, Frayne R. A comparison of texture quantification techniques based on the Fourier and S transforms. Med Phys 2008 35(11): 4998–5008.

117. Shur JD, Doran SJ, Kumar S, Ap Dafydd D, Downey K, O'Connor JPB, Papanikolaou N, Messiou C, Koh DM, Orton MR. Radiomics in Oncology: A Practical Guide. Radiographics. 2021 Oct;41(6):1717-1732.

118. Eliat PA, Olivie D, Saikali S, Carsin B, Saint-Jalmes H, de Certaines JD. Can dynamic contrast-enhanced magnetic resonance imaging combined with texture analysis differentiate malignant glioneuronal tumors from other glioblastoma? Neurol Res Int 2012 195176.

119. Parikh J, Selmi M, Charles-Edwards G, Glendenning J, Ganeshan B, Verma H, Mansi J, Harries M, Tutt A, Goh V. Changes in primary breast cancer heterogeneity may augment midtreatment MR imaging assessment of response to neoadjuvant chemotherapy. Radiology 2014 272(1): 100-12.

120. De Cecco CN, Ganeshan B, Ciolina M, Rengo M, Meinel FG, Musio D, De Felice F, Raffetto N, Tombolini V, Laghi A. Texture analysis as imaging biomarker of tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3-T magnetic resonance. Invest Radiol 2015 50(4): 239-45.

SECTION B: DEVELOPMENT OF NOVEL WHOLE-BODY MRI T1 AS A BIOMARKER FOR TUMOUR RESPONSE ASSESSMENT

Chapter 2 Technical/Analytical Validation of a Novel Quantitative MR Biomarker in Tumour Response Assessment and Prediction as part of whole body MRI assessment (T1 mapping)

CONTRIBUTION:

Primary author responsible for conception, data acquisition/analysis, drafting, revision, final approval and accuracy/integrity.

2.1 INTRODUCTION

2.1.1 Context

In 1971, Damadian observed a distinct difference in relaxation properties between normal and cancerous tissues with pulsed magnetic resonance [1]. During the early evolution from 'nuclear magnetic resonance (NMR)' to modern MRI techniques, intense efforts utilising in vitro and in vivo animal models were made to attempt to exploit these differences and further oncological clinical utility [2-4]; however these in most cases failed to adequately reflect in vivo tumours and there was increasing recognition that relaxation measurement techniques were difficult to quantitatively standardise across imaging platforms and field strengths within clinically feasible scanning conditions. Practical clinical MRI went on to provide excellent tissue contrast, but the signal intensity at any pixel could only be interpreted qualitatively and assigning meaning based upon underlying biophysical or physiological properties with respect to image signal intensity alone remained problematic. MR signal is influenced by several intrinsic contrast mechanisms, such as the density of water protons and proton relaxation times (T1, T2, T2*). Interpreting MR images, therefore, requires selection of an appropriate sequence and a thorough understanding of the corresponding signal contrast in relation to underlying pathophysiology whilst remaining cognisant of the characteristics of the imaging system such as coil sensitivities. Evaluation remains largely qualitative and subjective, and comparisons across different scanner hardware and acquisition sequences are difficult, typically being restricted to descriptions of relative contrast between different anatomical regions or tissues. To obtain information on tissue structure and function in a quantitative manner, independent of scanner hardware and sequence parameters, the contributions of different contrast mechanisms from the MR signal need to be extracted.

Quantitative T1 mapping has received renewed interest for a variety of applications. The T1 parameter is an intrinsic MR property of tissue, and mapping T1 in vivo is useful in several ways. Firstly, knowledge of T1 helps to optimise MR protocols, e.g. by setting Ernst angle appropriately. In addition, it provides a tool to evaluate contrast uptake, blood perfusion and volume, as well as disease progression during a longitudinal study. Furthermore, it is often desirable to compare T1 measurements across subjects and across scanners. Although there are many techniques for T1 mapping [5], there is also a wide range of reported T1 values in tissue [6], an inconsistency that raises the issue of reproducibility and standardisation.

2.1.2 Techniques:

The measurement of the T1 relaxation time requires measurement of the spin magnetization component in the direction of the magnetic field ("longitudinal" component) after perturbation from its equilibrium state (e.g. by applying an RF pulse to invert the magnetization). T1 is the decay constant for the exponential recovery of the longitudinal magnetization toward its equilibrium state. The rate at which the spin magnetization recovers to its equilibrium state depends on mechanisms that allow the spins to exchange energy with their surroundings. For example, interactions between the inverted magnetic spin moments and magnetic dipole moments of neighboring molecules that undergo tumbling motion can provide such a relaxation mechanism, which is particularly effective if the tumbling motion occurs near the resonance frequency of the spin moments. Another relaxation mechanism is provided by interactions with the paramagnetic moments in contrast agent molecules, which are particularly effective in shortening T1.

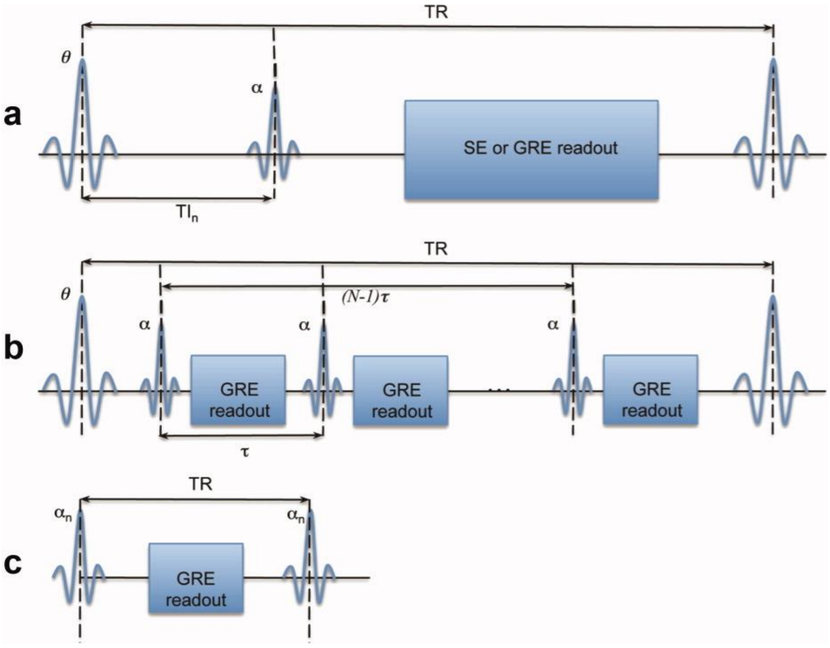


Figure 2-1 schematic of T1 mapping sequences: (a) Inversion Recovery, (b) Look-Locker, (c) Variable Flip Angle (from [7]).

The gold standard for T1 mapping is known as inversion recovery (IR) (Fig. 2.1a) and consists of inverting the longitudinal magnetization Mz and sampling the MR signal as it recovers according to the Bloch Equation:

$$\frac{d}{dt}M_z(t) = \frac{[M_0 - M_z(t)]}{T1}$$

The IR sequence consists of two RF pulses, separated by an inversion time TI. This sequence is repeated N times for each phase encode, varying the TI to produce N samples along the T1 recovery curve. The first RF pulse inverts the magnetization Mz, which then recovers with relaxation time T1. The second pulse tips the recovered longitudinal magnetization into the transverse plane. A variation of this approach is a saturation recovery sequence, where instead of inversion, the first pulse is a saturation pulse that tips the magnetization into the transverse plane and there is no second pulse. For either approach, the sequence is repeated several times, each time with a unique TI or time to repetition (TR) as appropriate, to sample to recovery curve. The recovery curve is then fitted to an exponential model, which is an accurate representation only when several assumptions are met, including a perfect inversion pulse, constant temperature, or $TR \gg T1$. While these assumptions are often justified, care must be taken to always pick a model that corresponds to the scheme acquisition [8]. While inversion recovery and saturation recovery offer accurate and precise T1 measurements, both techniques are too slow to be feasible in clinical practice since long TRs are required and acquire only one phase encode per TR. To speed up the acquisition, several variants have been proposed, such as following the inversion pulse by a fast spin-echo (FSE) [9] or echo-planar imaging (EPI) readout [10].

The Look-Locker (LL) method [11] is a rapid technique that measures T1 from a single recovery of longitudinal magnetization. It overcomes the limitation of the conventional IR method of requiring a long delay (on the order of T1) for longitudinal magnetization to recover until the next inversion pulse is played for subsequent readout. The basic sequence (Fig 2.1b) consists of only a single inversion pulse followed by a series of very small angle excitation (α) pulses with gradient echo readouts to sample the T1 curve. Because small angle RF pulses are used, the longitudinal magnetization is minimally disrupted during T1 recovery with no wait-period until equilibrium is reached as sampling is performed in a continuous manner. However, if separation between α pulses is less than T2, the T1 signal will be corrupted by residual transverse magnetization gathered from previous α pulses. To avoid this, either the separation between the α pulses needs to be > 5 × T2 or gradient spoiling needs to be used to 'crush' any residual transverse magnetization. It is also important to note that due to continuous perturbation of the magnetization by successive α pulses, the recovery is driven into equilibrium more quickly, resulting in an "effective T1" or T1* and 'actual T1' back calculated. Several other variants of the LL method have been developed targeted for specialized applications (e.g. MOdified LL Inversion recovery sequence (MOLLI) has been proposed by Messroghli et al for high resolution T1 mapping of the heart [12].

A different approach to rapid T1 measurement is to vary the flip angle of a gradient recalled sequence [13] (Fig. 2.1c). This method offers volumetric, high spatial resolution T1 mapping in significantly less time than conventional methods. The time gain stems from the use of a short TR, unlike the conventional approach where TRs must be long to allow full signal recovery. The main source of error, however, is imperfect tipping of magnetization due to an inhomogeneous RF field. Imperfections in the transmit field of the RF coil, which leads to certain locations in the imaging volume experiencing more or less RF power and, therefore, an imperfect RF flip angle for tipping magnetization and for refocusing. RF inhomogeneity affects all types of relaxation measurements. For T1 quantification, it represents the most common cause of errors and has an especially pronounced effect on the variable flip angle approach. To correct for this error, a straight-forward solution is to map the RF B1 field to determine the true flip angles, and the correct flip angles can then be used in calculating relaxation times [14]. Slice profile effects are another source of inaccuracy. Ideally, all the spins in the selected imaging slice should be inverted (in the case of a 180-degree pulse), and all spins outside the slice should remain unperturbed. In reality, the slice profile is never perfectly rectangular, which means that the actual flip angle falls off over a transition region at the edges of the slice. Hence, the effective flip angle is not the nominal angle but is the integral of the inversion profile over the slice thickness. Slice profile imperfection affects both T1 and T2 measurements, even for gold standard measurements [15]. This problem can be alleviated somewhat by using nonselective pulses or interleaving even and odd slices in a multislice slice-selective excitation. Partial volume effects, magnetization transfer effects, temperature effects (T1 increases 2–3% per degree Celsius), and movement due to actual subject motion or bulk flow (e.g., blood flow that artifactually shortens T1) all can lead to errors. A final consideration to obtaining accurate measurements is having a good sense of target range of T1 or T2 values, because acquisition parameters (e.g., TR, TE, flip angles) can only be optimized for a limited range of relaxation times.

To calculate relaxation times, the data acquired from the MRI experiment is fitted to some form of mathematical model. In many cases, the exponential function is used, as it describes many biophysical phenomena, including T1 and T2 relaxation. Estimation of T1 generally involves fitting a monoexponential function to the T1 relaxation curves. With conventional IR T1 measurement, an exponential fit to signal recovery at various TIs is performed. The ideal scenario is where there is no noise with a single relaxation time in the imaging voxel, data fitting theoretically can be achieved using two data points on the monoexponential curve [7]. In the presence of noise, inherent uncertainty in each measured data point ideally necessitates consideration of acquiring at least three data points, with attention paid to how far apart they should be spaced to yield reliable estimates of T1. One may increase the number of measured data points, however, there is a practical limit beyond which a greater number of measurements or averages will not yield significant additional benefit (the optimal balance between these two considerations can be determined through simulations). It is acknowledged that monoexponential modelling cannot describe all situations, however for T1, multicomponent measurements are generally difficult to obtain, because exchange is fast compared with the T1 relaxation rate.

For this technical/analytic process we attempted to demonstrate the accuracy (compared with gold standard IR sequences), reproducibility and repeatability of a clinically feasible variable flip angle acquisition which would permit whole body volumetric imaging in patients with metastatic cancer.

To do this in stepwise fashion, we first validated a rapid multishot turbo spin echo sequence inversion recovery (MS-TSE-IR) against gold standard inversion recovery SE sequence for the longitudinal relaxation times (taking approximately two hours for a single slice) both on a dedicated phantom and one volunteer single sitting in vivo. This was to provide an accurate surrogate reference T1 value for a larger accuracy and precision study. Once this was established, we used the faster MS-TSE IR sequence as ground truth in vivo, to validate a variable flip angle GRE sequence in healthy volunteers in a single anatomic ‘station’ (pelvis-chosen to minimise artefact e.g. respiration and afford a range of tissue types) for accuracy, repeatability and reproducibility. Finally we conducted another study with healthy volunteers on a whole body basis with multiple ‘stations’ using VFA GRE sequences to demonstrate repeatability and reproducibility as well as examine which dual flip angle selection would be most robust to take forward to clinical whole body use in cancer patients by providing an acceptable balance of extensive spatial coverage, accuracy, and repeatability in T1 quantification.

2.2 MATERIALS AND METHODS

Institutional review board approval was granted and healthy volunteers’ informed consent (written and verbal) was obtained prior to entry into this prospective validation study (07/Q0502/15).

2.2.1 MRI Techniques overview:

Imaging was performed using a single 3.0 T wide-bore MR scanner (Ingenia; Phillips Healthcare, Best, Netherland) using the manufacturer head coil, two anterior surface coils (2*16 elements) and table embedded posterior coils. Where healthy subjects were imaged, all volunteers were imaged supine with arms by their side. Where referenced, pelvis only acquisitions were coronal and centred between the anterior superior iliac spine at the top and femur mid shaft at the bottom of the image/slice and at the level of the femoral heads in the anteroposterior plane. For the first two validation studies, a bag of saline was placed between the legs as cranially as possible to provide a higher T1 value substrate for analysis. Where referenced, for the healthy subjects ‘full body’ coverage (vertex to mid-thigh) was obtained through a multi-station acquisition of contiguous body regions.

Where referenced, a Eurospin II TO5 (Diagnostic Sonar LTD, Livingston, Scotland) phantom consisting of 12 tubes providing calibrated T1 (and T2) values for reference was also employed (range depending on exact temperature ~205ms to 1950ms).

Inversion recovery parameters are given in the relevant sections, however the 3D variable flip angle (VFA) sequences were the same across the validation acquisitions and these parameters together with the B1 map shown in Table 2.1 below. Volumetric T1 maps were generated using variable flip angle 2-point modified (m)Dixon 'in phase' images by a linear fitting algorithm [16] with MATLAB (v7.13) incorporating a separately acquired dual TR B1 map as per well described methods [17]. Where required, the VFA slice (e.g. in pelvis) was picked and registered to match the single-slice IR images in early validation.

Parameters	**mDixon-variable (2-8 separate) flip angles**	**B1 map**
Plane	Coronal	Coronal
TE (ms)	1.15/2.3	2.8
TR (ms)	3.5	10/50 (extension 90)
Field of view	500x300	500x300
Voxel size (mm*mm)	2.1x2.1	2.1x2.1
Number of slices	120	60
Slice thickness (mm)	5	10
Acquisition matrix	240x240	120x120
Echo train length	2	1
Acceleration factor (SENSE)	2	3
Pixel bandwidth (Hz)	1992	3356
Scan time (s)	17 per FA	172 (per station)
Number of stations	Up to 4 (5)	Up to 4(5)
Total FH coverage (mm*mm)	1388 (10% overlap)	1388

Table 2.1 Table showing whole body MRI (WBMR) parameters. T2-TSE: T2-weighted turbo spin echo, mDixon: modified Dixon, DWI diffusion weighted imaging, TE: time of echo, TR: repetition time, SENSE: sensitivity encoding.

2.2.2 Validation of accuracy and precision:

Assessment with gold standard inversion recovery in vitro and in vivo:

Gold standard IR T1 single slice maps were fitted from six IR spin-echo scans (TIs: 150, 500, 1000, 2000, 3000, 4000ms) and TE/TR: 30/15,000 ms), using a nonselective inversion pulse with FoV 376mm/160x160, and slice thickness 3mm. TR was long to allow full relaxation in high T1 structures and minimise measurement error. Multishot IR TSE sequence with matched parameters but allowing more time efficient single slice acquisition was also performed with the same TIs. Section-selective adiabatic inversion pulse was used to ensure good inversion throughout the field of view, even in the presence of variations in the amplitude of the radiofrequency field. The inversion was performed over a section thickness twice that of the imaged section thickness to avoid the imperfect inversion near the edges of the inverted section.

The complex IR data were fitted (Figure 2.2) by considering the general IR sequence [θ_1-TI-θ_2-(TRTI)]$_n$, where θ_1 and θ_2 are the RF pulses (nominally 180° and 90°, respectively), TI is the inversion time, and TR the repetition time. Sampling the complex signal Sn (TI_n) at different time points TI_n gives an estimate of T1 when data are fitted to:

$$S_n = a + be^{-\frac{TI_n}{T_1}}$$

$$S_n = a + be^{-\frac{TI_n}{T1}}$$

This model has five free parameters because a and b are complex-valued and T1 is real-valued [8]. Finally matched field of view VFA volumetric acquisition was also performed using parameters previously given and up to eight flip angles (2.5-20 degrees inclusive, in 2.5 degree increments).

These three sequences (gold standard IR, MS TSE IR and VFA 3D) were performed on the T1 phantom, as well as a single subject's (36 year old male) pelvis. For the healthy volunteer, different tissues were chosen by the same radiologist (7 years' experience) on the single slice IR images (Figure 2.3) and regions of interest (ROI) of identical size were constructed and transferred to the respective parametric maps.

Assessment with MS-TSE IR in vivo single station (accuracy and precision):

Nine healthy volunteers (five males, average age 34.7 years and four females, average age 29.9 years) were scanned three times in two sittings- two scans were performed one after the other and then a final third scan approximately a week after the first (median 8 days;

range 6 to 9 days). Each scan was only through the subject’s pelvis and involved a single slice MS-TSE-IR sequence (to evaluate subsequent VFA 3Daccuracy) as described above and an eight flip angle VFA 3D acquisition also as above (also assessing precision- repeatability and reproducibility). The same radiologist contoured ROIs (using in phase FA 15 degree images for VFA 3D) as before and these were compared across the sequences, and subjects within and across sessions.

2.2.3 Validation of precision and dual flip angle selection on whole body basis:

Nine (different) subjects (six males, average age 28.4 years and three females, average age 31.3 years) were scanned three times in two sittings- two scans were performed one after the other and then a final third scan approximately a week after the first (median 7 days; range 5 to 9 days). Each scan comprised a multi-station whole body acquisition with an eight flip angle VFA 3D acquisition at each station described above. Parametric maps of each image slice at each station were constructed as described previously using all eight flip angles as well as flip angle pairs, to which ROIs from different tissue types were transferred from in phase (FA=15 degrees) images and both were assessed within subjects within and across sessions.

2.2.4 Statistical Analysis:

Statistical analysis was performed using GraphPad Prism v6.0 (La Jolla, California, USA) and using SPSS version 22 (IBM, New York, USA). Statistical significance was defined as P < 0.05.

Accuracy

Agreement of VFA and MS-TSE-IR T1 measurements with reference gold standard IR T1 measurements was tested by Lin's concordance correlation coefficient [18, 19] and Bland-Altman statistics. The accuracy error was computed as the percentage difference between T1 measured with the test protocols during the first scanning session (where multiple), and reference gold standard IR T1 for the first experiment (or MS-TSE-IR in second). Smaller values represent higher accuracy.

Precision:

Test-retest repeatability was assessed by the precision error, calculated as the percentage difference of T1 values measured in duplicate relative to the mean of the two measured values (at the same sitting). Smaller values represent higher repeatability. Reproducibility across sessions (temporal) for the T1 measured in each ROI with the eight flip angle VFA protocol was assessed by intraclass correlation coefficient (ICC) and coefficient of variation (% CV = 100 x standard deviation/mean calculated between sessions, for each reference T1 value) and the mean of the first session was compared to the second session for each tissue type. Levels of agreement for ICCs and (and Lin's concordance correlation coefficient) will be interpreted according to Landis and Koch [20]: 0.0–0.20 no to slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, and 0.81–1 almost perfect agreement. Flip angle pairings were also assessed as above to find the best performing pair for specified T1 ranges (which could be used in future studies as reference for selecting appropriate FA pair by target T1 value(s)).

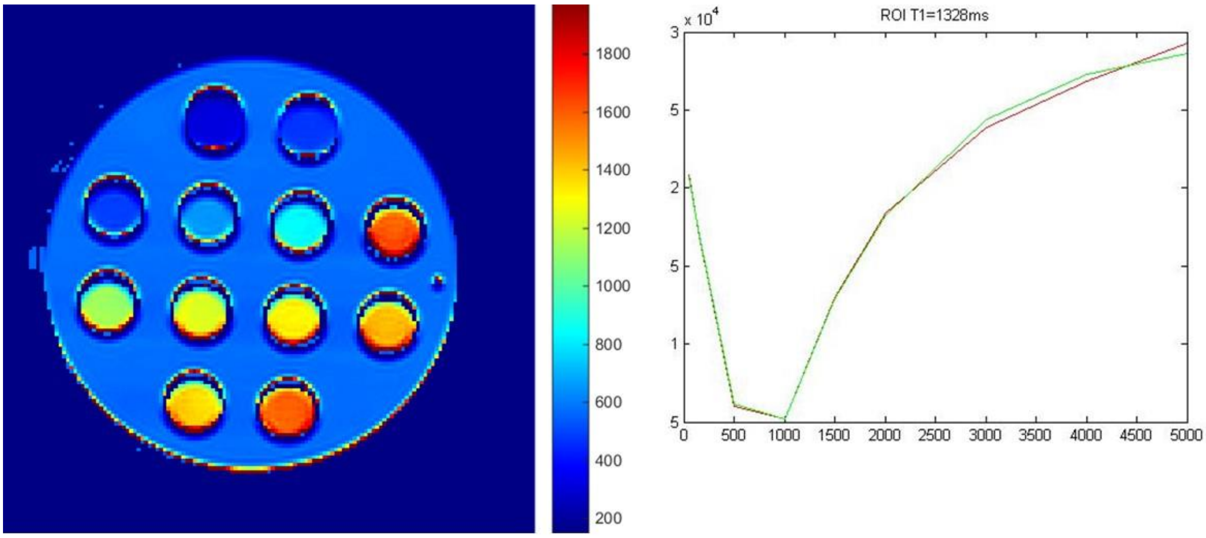


Figure 2-2 Showing gold standard IR derived parametric T1 map with colour look up table applied (CLUT) performed on T1 phantom T1 map from 'real' data with fitting (green) and data points (red)

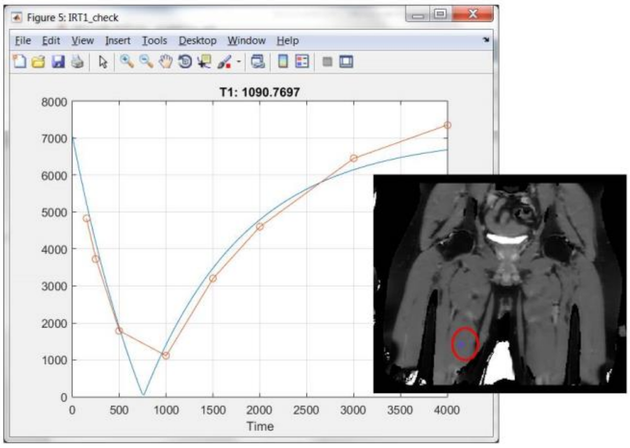


Figure 2-3 Showing MS TSE IR sequence derived parametric T1 map with colour look up table applied (CLUT) performed 36 year old male volunteer pelvis with fitting of muscle ROI.

2.3 RESULTS

Validation of accuracy:

Assessment with gold standard inversion recovery in vitro and in vivo:

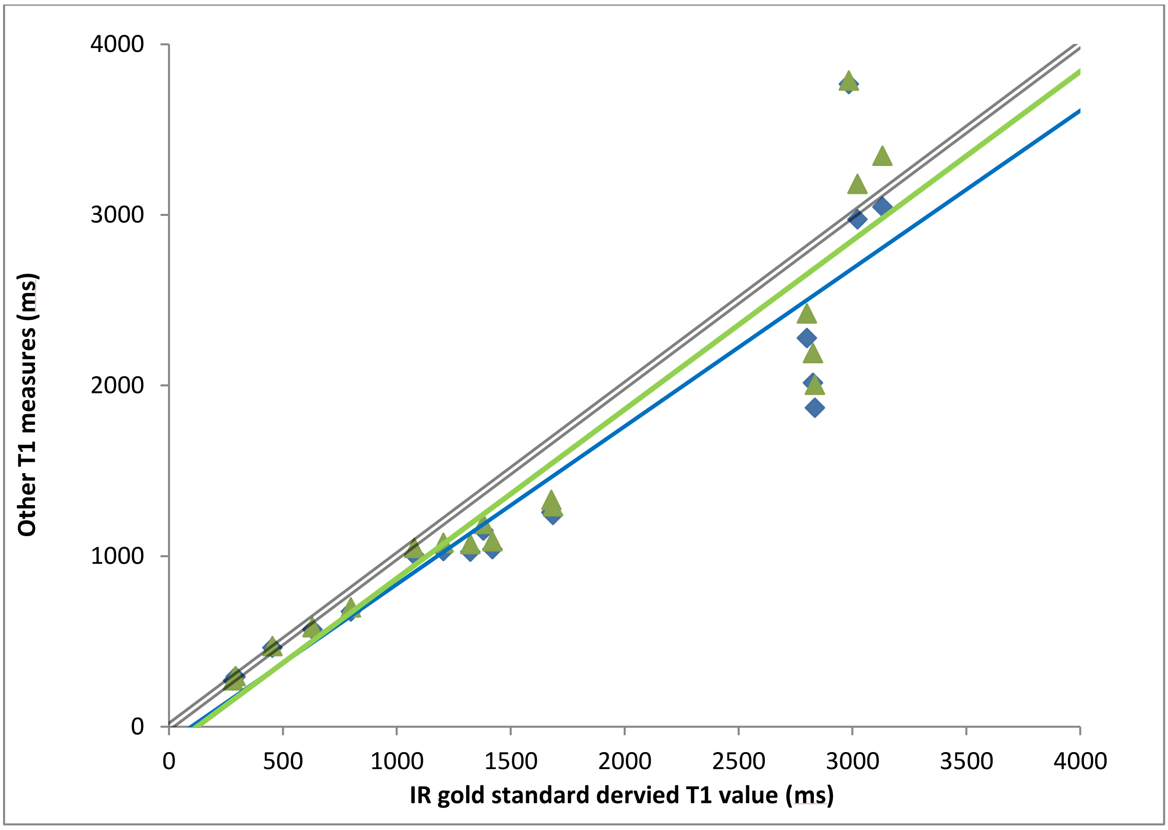


Figure 2-4 Correlation between gold standard IR derived T1 values versus MS-TSE-IR (green) and VFA 3D (blue) with y=x line for phantom data.

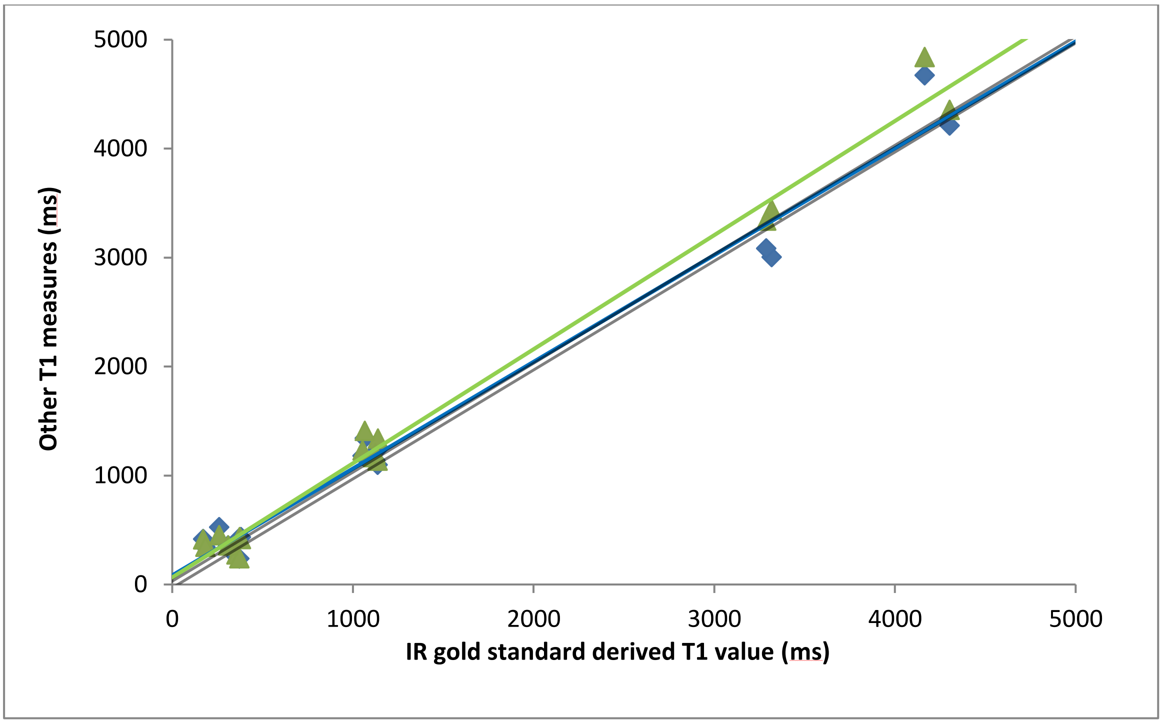


Figure 2-5 Correlation between gold standard IR derived T1 values versus MS-TSE-IR (green) and VFA 3D (blue) with y=x line for in vivo pelvis data.

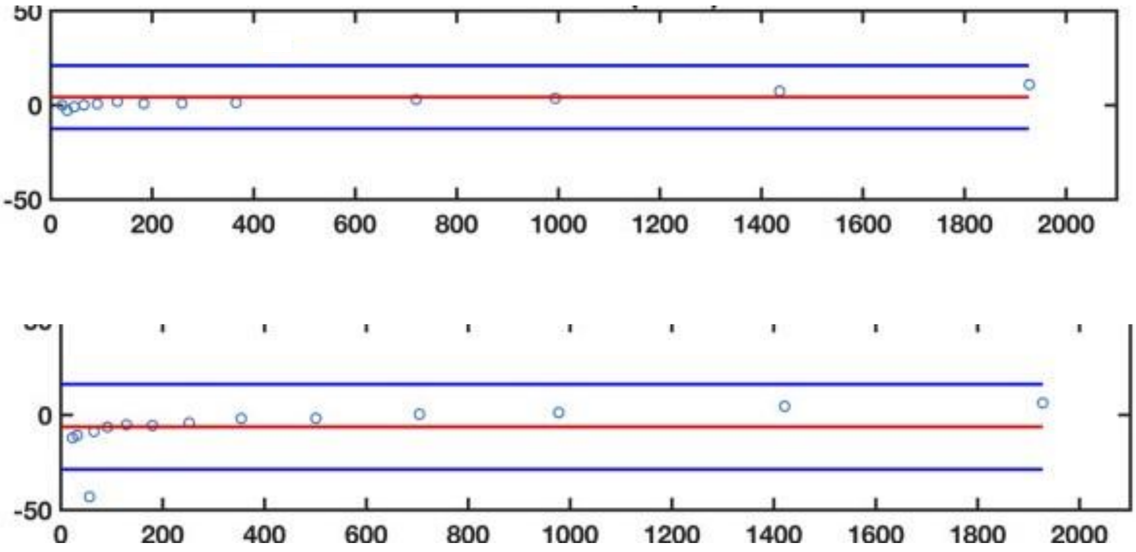


Figure 2-6 Bland-Altman plots, top MS-TSE-IR and bottom VFA-3D derived T1 values, in phantom against gold standard IR derived T1 (y-axis is % difference and x-axis is average T1 (ms)).

Figures 2.4 and 2.5 show phantom and in vivo pelvic correlation of the two test T1 methods (MS-TSE-IR and VFA-3D) with gold standard IR T1 derivation demonstrating very good (almost y=x) agreement for each singular experiment. For phantom scanning, there was excellent concordance between MS-TSE-IR and gold standard IR measurements (concordance correlation coefficient > 0.99; P < 0.001), with some deviations from unity line observed at T1<150ms. Bland-Altman plots (Figure 2.6) show near-zero bias, limits of agreement (−10%, 10%). The median accuracy error was 1.2 % (IQR 1.4%). For the VFA 3D acquisition concordance with gold standard IR was also excellent (concordance correlation coefficient 0.91, P<0.001) in phantom scanning, the median accuracy error was 5.2% (IQR 14.2%), near zero bias across the range of T1 values with some underestimation of low T1 values and overestimation at high T1 values.

With respect to in vivo accuracy validation on a single healthy volunteer pelvis, similar observations were seen. In total 18 ROIs were drawn per sequence across the single slice analysed (6 muscle, 4 bone, 4 subcutaneous fat, 2 urinary bladder, and 2 saline bag). Comparing the MS-TSE-IR versus gold standard IR derived T1 values for these structure, concordance was excellent (concordance correlation coefficient >0.99; p<0.0001) with little deviation from unity line, near zero bias and limits of agreement (-15%, 15%). The median accuracy error was 0.6% (0.7% IQR). For the VFA 3D acquisition concordance with gold standed IR T1 values was also excellent (concordance correlation coefficient 0.94; p<0.001). The median accuracy error was 8.3% (IQR 21.4%) with overall near zero bias with minor deviations around low and high T1 values (limits of agreement -20%, 20%).

Assessment with MS-TSE IR in vivo single station (accuracy and precision):

T1 values derived from the MS-TSE-IR single slice parametric map ROIs averaged across the test-retest scans were used for as reference T1 values against which the accuracy of the VFA 3D acquisition was compared (given this- accuracy of the MS-TSE-IR was not assessable though precision could be given the multiple tests). Again each slice had 18 ROIs across the same range of tissue types.

Precision was assessed in terms of repeatability using the first sitting test-retest scan ROI values and reproducibility comparing the first sitting scans with the temporally separated second sitting scan ROI T1 values performed one week later as in Methods section. Furthermore, the T1 values were binned into three ranges low (T1<500ms), intermediate (500-1500ms), and high (>1500ms) for presentation purposes so as to gain an appreciation of T1 variation across these ranges and is presented in Table 2.2.

T1 reference	Substrate tissues (mean of median T1	Accuracy Error % (IQR)	Precision Error % (IQR)

value (ms)	value)	MS-TSE-IR	VFA 3D	MS-TSE-IR	VFA 3D
<500ms	Fat (424ms) Bone (456ms)		4.5 (18.4)	1.4 (0.7)	6.7 (4.3)
500-1500ms	Muscle (1325ms)		9.2 (6.4)	1.1 (0.9)	8.9 (2.4)
>1500ms	Urinary bladder (3271ms) Saline (3048ms)		7.3 (10.2)	1.3 (1.0)	6.2 (3.7)
Overall			8.5 (11.4)	1.3 (0.8)	7.9 (3.5)

Table 2.2 Accuracy and precision (test-retest) errors (%) for VFA 3D protocol, for ranges of reference T1 values in vivo. Values are given as median (interquartile range).

As shown, there was less precision error in the MS-TSE-IR values. Accuracy and precision for the VFA 3D method on the single station assessment was good with both <15%. For temporal reproducibility, ROIs from the two scans from the first sitting were averaged and compared to the scan ~ one week later. For both MS-TSE-IR and VFA-3D the T1 value agreement between scans one week apart was excellent (ICC 0.99; p<0.001 and ICC 0.95; p<0.001 respectively). Root mean square co-efficient of variation for MS-TSE-IR T1 values was 4.75%, (range 1.4–13.7%) and for VFA-3D T1 values RMS CV was 9.24%, range: 0.13–14.6%).

Validation of whole body precision and dual flip angle selection:

Whole body VFA 3D scans were assessed using a total of 20 ROIs per subject per scan (4 muscle paravertebral, 4 subcutaneous fat, 2 bone marrow vertebral T6 and L4, 2 liver, 2 spleen, 2 renal cortex, 2 urinary bladder, 2 brain white matter) transferred to the relevant parametric T1 map from in phase Dixon images (FA 15 degrees for best tissue contrast). We present data for precision based on T1 values generated from all 8 flip angles as well as four best performing (and most sensible based on fitting parameters) dual flip angle selections and separated according to T1 value ranges (as well as overall). Repeatability was performed on the first sitting scans for test-retest T1 values and reproducibility data between temporally separated first and second sitting T1 values. Both precision error and RMS-CoV values were in general lower for dual flip angle pairs and dual flip angle selection of 2.5 and 15 degrees seemed to perform slightly better in precision terms though the differences between these pairs were not significant. This is summarised in Table 2.3.

Flip Angles	Repeatability (test-retest)	Reproducibility (between sessions)

used	Precision Error % (IQR)				Root mean square CoV (range)			
	Overall T1 range	Low T1 <500ms	Intermediate T1 (500-1500ms)	High T1 (>1500ms)	Overall T1 range	Low T1 <500ms	Intermediate T1 (500-1500ms)	High T1 (>1500ms)
All (8FAs)	**7.4 (3.3)**	**6.3 (4.1)**	**7.2 (2.4)**	**7.7 (3.9)**	**8.4 (1.4-14.8)**	**6.7 (0.1-15.7)**	**10.1 (2.0-20.2)**	**7.4 (2.3-17.4)**
2.5 + 12.5	5.4 (4.2)	4.2 (3.7)	5.7 (4.2)	5.1 (4.7)	7.2 (2.5-18.2)	5.7 (1.2-15.2)	9.5 (4.5-23.1)	8.8 (3.4-20.1)
2.5 + 15	**5.2 (3.9)**	**4.4 (3.9)**	**5.4 (2.8)**	**4.7 (2.7)**	**5.9 (0.7-11.2)**	**5.9 (1.3-14.3)**	**6.4 (0.5-11.3)**	**4.9 (1.5-10.4)**
2.5 + 17.5	7.4 (2.9)	5.6 (4.5)	7.9 (4.2)	6.1 (3.5)	6.7 (2.4-16.7)	6.3 (2.5-13.3)	7.8 (4.2-14.4)	7.0 (2.8-16.7)
2.5 + 20	9.1 (5.3)	6.7 (5.3)	11.4 (5.7)	8.9 (5.1)	11.1 (3.1-23.2)	9.3 (1.9-21.4)	12.4 (3.5-30.3)	11.5 (4.0-24.3)

Table 2.3 Summarises the precision values for whole body test-retest (repeatability) and across session (reproducibility) T1 values calculated from all eight flip angles and several dual flip angle pairs. In bold are the all 8FA measures and best performing dual flip angle pair (2.5 and 15 degrees)

2.4 DISCUSSION

Accurate, repeatable, and reproducible quantification of T1 is highly desired for a range of practical applications as discussed in the introduction. Here we demonstrate excellent agreement between gold standard IR measurement and both the faster MS-TSE-IR and VFA 3D sequence derived T1 values across a range of values both in vitro and in vivo. We justify the use of the MS-TSE-IR T1 methodology as means of assessing the VFA-3D accuracy in a larger experiment to assess the accuracy, repeatability and reproducibility of VFA-3D method in vivo (and this is reinforced by very low precision error percentage for the MS-TSE-IR technique). Finally we demonstrate excellent precision of the whole body VFA 3D acquisition derived T1 values in test-retest and across session bases and can be fairly confident (at least on this single centre single platform experiment) that the measures for a clinical qualification study should be fairly robust both in accuracy and precision. When moving from VFA to dual flip angle T1 measurement (and forearmed with knowledge that tumour typically have very high T1 relaxation times at baseline) we do not seem to lose significantly in terms of accuracy or precision with suitably chosen pairing (e.g. in this experiment 2.5 and 15 degrees seemed optimum but was not significantly different). Importantly preclinical studies have suggested changes of the order of ~20% in T1 relaxation with successful therapy- this technique should be able to detect this change with some certainty.

There are several limitations of the study. We were unable to use gold standard IR T1 estimation beyond phantom and a single volunteer- this was not feasible given the time take to acquire a single slice and necessitated using the multishot sequence to approximate this for accuracy/alternate verification (and we did not use NMR techniques for ultra-accurate T1 derivation). Secondly there are a number of areas which could have been explored to optimise techniques- notably for VFA e.g. B1 mapping optimisation, actual flip angle estimation, parallel imaging optimisation etc. Some of these were touched upon but not formally evaluated in dedicated experiments. Most obviously, this study was performed on a single scanner in a single centre. There are now T1 validation studies being performed in multicentre setting [21], which are providing excellent resources to further this field and clearly if quantitative T1 mapping is to be used clinically this should occur in concert with clinical qualification.

2.5 REFERENCES

1. Damadian R. Tumor detection by nuclear magnetic resonance. Science 1971; 171:1151-1153.

2. Bottomley PA, Foster TH, Arsinger RE, Pfeifer LM. A review of normal tissue hydrogen NMR relaxation times and relaxation mechanisms from 1-100 MHz: dependence on tissue type, NMR frequency, temperature, species, excision, and age. Med Phys 1984; 11:425-448.

3. Beall PT, Amtey SR. Kasturi SR. NMR handbook for biomedical applications. New York: Pergamon, 1984.

4. Hazlewood CF. Cleveland G, Medina D. Relationship between hydration and proton nuclear magnetic resonance relaxation times in tissues of tumor-bearing and non-tumor-bearing mice: implications for cancer detection. JNCI 1974; 52:1849-1853.

5. Kingsley PB. Methods of measuring spin-lattice (T1) relaxation times: an annotated bibliography. Concepts Magn Reson 1999;11:243–276.

6. Tofts P. Quantitative MRI of the brain: measuring changes caused by disease. Chichester, West Sussex: Wiley, 2003.

7. Cheng HL, Stikov N, Ghugre NR, Wright GA. Practical medical applications of quantitative MR relaxometry. J Magn Reson Imaging. 2012 Oct;36(4):805-24.

8. Barral JK, Gudmundson E, Stikov N, Etezadi-Amoli M, Stoica P, Nishimura DG. A robust methodology for in vivo T1 mapping. Magn Reson Med 2011; 64: 1057–1067.

9. Zhu DC, Penn RD. Full-brain T1 mapping through inversion recovery fast spin echo imaging with time-efficient slice ordering. Magn Reson Med 2005; 54: 725–731.

10. Gowland PA, Leach MO. A simple method for the restoration of signal polarity in multi-image inversion recovery sequences for measuring T1. Magn Reson Med 1991; 18: 224–231.

11. Look DC, Locker DR. Time saving in measurement of NMR and EPR relaxation times. Rev Sci Instrum 1970; 41: 250–251.

12. Messroghli DR, Radjenovic A, Kozerke S, Higgins DM, Sivananthan MU, Ridgway JP. Modified Look-Locker inversion recovery (MOLLI) for high-resolution T1 mapping of the heart. Magn Reson Med 2004; 52: 141–146.

13. Deoni SC, Rutt BK, Peters TM. Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state. Magn Reson Med 2003; 49: 515–526.

14. Cheng HL, Wright GA. Rapid high-resolution T(1) mapping by variable flip angles: accurate and precise measurements in the presence of radiofrequency field inhomogeneity. Magn Reson Med 2006; 55: 566–574.

15. Kingsley PB, Ogg RJ, Reddick WE, Steen RG. Correction of errors caused by imperfect inversion pulses in MR imaging measurement of T1 relaxation times. Magn Reson Imaging 1998; 16: 1049–1055.

16. Wang D, Shi L, Wang YX, Yuan J, et al. Concatenated and parallel optimization for the estimation of T1 map in FLASH MRI with multiple flip angles. Magn Reson Med 2010 63(5): 1431-6.

17. Yarnykh VL. Optimal radiofrequency and gradient spoiling for improved accuracy of T1 and B1 measurements using fast steady-state techniques. Magn Reson Med 2010 63(6): 1610-26.

18. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45:255–268.

19. Lin LI-K. A note on the concordance correlation coefficient biometrics. Stata J 2000;56:324–325

20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–174

21. Bane O, Hectors SJ, Wagner M, Arlinghaus LL, Aryal MP, Cao Y, Chenevert TL, Fennessy F, Huang, W, Hylton NM, Kalpathy-Cramer J, Keenan KE, Malyarenko DI, Mulkern RV, Newitt DC, Russek SE, Stupic KF, Tudorica A, Wilmes LJ, Yankeelov TE, Yen Y-F, Boss MA, Taouli B. Accuracy, repeatability, and interplatform reproducibility of T1 quantification methods used for DCE-MRI: Results from a multicenter phantom study. Magn. Reson. Med 2018;79: 2564-2575.

Chapter 3 Clinical Validation (Qualification) of Novel Quantitative MR Biomarker in Tumour Response Assessment and Prediction as part of whole body MRI assessment- sarcoma (soft tissue and bone) and breast cancer

CONTRIBUTION:

Primary author responsible for conception, data acquisition/analysis, drafting, revision, final approval and accuracy/integrity.

3.1 RUNNING TITLE

In vivo measurement of tumour T1 relaxation time using whole body multiple flip angle method can predict response to chemotherapy in local bone sarcoma and metastatic soft tissue sarcoma and breast cancer.

3.2 INTRODUCTION

Oncological therapies are associated with variable efficacy, side-effects and financial cost. The current standard for assessing treatment response of tumours based on imaging is one-dimensional size measurement as defined by the Response Evaluation Criteria In Solid Tumours (RECIST) 1.1 [1]. Whilst acceptable for certain solid tumour and conventional chemotherapies, the RECIST criteria fail to adequately characterise response of cystic lesions or indeed the response of solid lesions to newer anti-angiogenic treatments [2]. In addition specific metastatic deposits (e.g. bone disease) are non-measurable. One proposed improvement on the RECIST criteria was the Choi criteria which incorporated density and intensity changes by CT (and MR) imaging, in addition to changes in size [3]. Finally, to address the inadequacies of the RECIST and Choi criteria, the use of a standardized uptake value (SUV) obtained by PET was incorporated into response criteria to better assess tumour metabolism [4]. Thus, the revised RECIST 1.1 criteria endorsed the functional assessment of tumours following treatment as complementary to anatomic evaluations [5]. However, it is recognised that the above conventional re-staging strategies may be sub-optimal for response assessment and more reliance can be placed upon clinical outcome measures such as progression free survival. Furthermore, recent efforts have focused on quantitative parametric measurement to predict response prior to, or early during the course of, systemic therapy to mitigate serious side-effects, morbidity and costs of treatment as well as expedite more successful treatment strategies in a personalised approach.

Soft-tissue sarcomas (STS) are a heterogeneous group of neoplasms arising from cells of embryonic mesenchymal origin. STS account for less than 1% of all malignant neoplasms and rates of STS range from 4.5 to 6.5 cases per 100,000 person-years [6]. Staging of STS is most commonly performed using the American Joint Committee on Cancer (AJCC) system. STS stage is determined by the size of the tumour, the histological grade, and whether there is spread to lymph nodes or distant sites. The overall five-year survival rate in patients with soft tissue sarcomas of all stages remains only 50% to 60% [7]. Approximately half of all STS patients with intermediate or high-grade tumours develop metastatic disease requiring systemic treatment. In almost all cases the treatment intention for systemic/metastatic disease is palliative. Approximately 50% of patients develop distant metastases and eventually die of disseminated disease; with a median survival of approximately 12 months from diagnosis of metastases [8]. The management of advanced disease is complex and published response rates to chemotherapy vary enormously; from 10–50% depending on the drugs used, patient selection and histological subtype [9].

Primary bone tumours are rare, accounting for less than 1% of cancers in adults [10]. There are on average 427 new cases per year in England and Wales [11] and there is relatively high incidence in children and adolescents (accounting for approximately 5% of all childhood cancers in European Countries) [12], but can arise at any age. Osteosarcoma (OS) is the most frequent primary cancer of bone (incidence 0.2-0.3/100,000/year) [13], UK rates are 0.27/100,000 population in England with an average of between 124–150 cases per year in England and Wales. Ewing sarcoma (ES; including primitive neuroectodermal tumour of bone) is the second most common primary malignant bone cancer in children and adolescents, but is also seen in adults. The median age at diagnosis is also around 15 years with UK all age standardised rates report 0.11-0.12/100,000 population in the UK (1979–2004)] with approximately 65–75 new cases per year in the UK [10]. Bone sarcoma is a potentially curable disease with surgery and chemotherapy being the mainstays of treatment- notably preoperative ‘neoadjuvant’ systemic combination chemotherapy, local surgery, and postoperative ‘adjuvant’ chemotherapy is the current treatment paradigm [14] with the aim of decreasing the incidence of a subsequent distant relapse [15]. The current NHS standard first-line chemotherapy for OS is a combination of high-dose methotrexate, doxorubicin and cisplatin (MAP) given over 5 weeks with patients receiving a total of 6 cycles of treatment. Patients with extremity/resectable tumours receive 2 cycles of MAP prior to surgery and undergo MRI scanning at 8 weeks to assess radiological response to treatment and aid surgical planning (subsequently receiving 4 cycles of post-operative chemotherapy). The histological response at resection (week 11) is the most important prognostic factor in determining long-term outcome for patients with localized disease at diagnosis [16]. Internationally, the standard treatment of ES is not defined. Current trial treatment in ES (with locally confined or pulmonary/pleural metastatic ES) in our institution involves randomisation to two different ‘induction’ (pre-local therapy treatment and assessment of response utilising histology) prior to ‘consolidation’ treatment post local therapy.

The National Institute for Health and Clinical Excellence (NICE) have published guidelines on the treatment of advanced breast cancer in 2009 to help standardise clinical practice in the UK [17], comprising the small proportion are diagnosed in the advanced stages, when the tumour has spread significantly within the breast or metastatic (mBC). Approximately 5% to 10% of breast cancers are metastatic at diagnosis; of these, approximately one-fifth will survive 5 years [18]. Notwithstanding the increasing use of systemic treatments in the adjuvant setting, the median survival has steadily improved over the past 30 years from 12-18 months to 24-36 months as treatment options have increased [19]. The exact choice of hormonal/chemotherapy regimen will depend on a number of factors and which regimen should be used in which circumstance is not generalizable [20] (except to note the National/International guidance). Chemotherapy response assessment is usually performed every two or four cycles of chemotherapy. The main aim of these assessments is to exclude progressive disease, in particular in patients for whom further treatment options exist or those who experience significant toxic effects from their treatment.

Quantitation of T1 may be used to characterise normal and pathological tissue and there is increasing interest in its use in producing robust dynamic contrast enhanced (DCE) quantitative parameters. Whole body (WB) T1 quantitation presents several difficulties (most notably respiratory artefact and large field of view requirement). Several techniques have been explored; the most commonly used being inversion recovery (IR; which remains the reference standard method), multiple flip angle (MFA), and Look-Locker (LL) techniques. Several studies using different T1 estimation methodologies showed that the T1 of experimental tumours was greater than in normal tissue and was considered to reflect predominantly the extracellular space [21-23], with a larger interstitial compartment giving a longer (higher) T1. In untreated tumours, a low tumour T1 has been correlated to increased necrosis [22], low water content, high levels of soluble protein [21], and low proliferation; these baseline characteristics have also been associated with poorer response to systemic therapy perhaps in part due to more chaotic perfusion and drug delivery for example. In four different quantitative T1 studies, only decreases were detected in response to successful therapy with radiation or cytotoxics [24-27]. McSheehy et al (2009) recently used animal tumour models to show five anticancer drugs (with different mechanisms of action) caused a rapid decrease in T1 that often preceded and was strongly correlated with the change in tumour size suggesting that a fractional change in T1 (ΔT1) was a useful generic early-response marker. Furthermore, ΔT1 did not occur in tumours resistant to that drug, and ΔT1 was positively correlated with proliferation markers (Ki67 and choline), suggesting that ΔT1 indirectly reflected the number of viable cells in a solid tumour [27]. Repeatability of a whole body B1 corrected multiple flip angle (MFA) method has been reported in healthy volunteers with co-efficients of variation across tissue types between 5-10% when selecting two appropriate flip angles [28].

This study evaluated tumoral T1 relaxation times using multiple flip angle method prior to, and early during, systemic chemotherapy for prediction of response to therapy in three cancer cohorts.

3.3 MATERIALS AND METHODS

Institutional review board approval was granted and patients' informed consent (written and verbal) was obtained prior to entry into this prospective study (R&D No: 13/LO/0758).

3.3.1 Patient cohort:

Metastatic soft tissue sarcoma (mSTS) or breast cancer (mBC) patients with histologically confirmed disease were recruited provided they were aged 18 years or above, with highly likely metastatic disease prior to systemic chemotherapy. Bone tumour (either OS or ES) patients with histological confirmed disease were recruited provided they were aged ≥13 years and considered by referring clinician to derive benefit from chemotherapy usually prior to likely resection in localised or low volume oligometastic disease. Patients screened between 14th August 2014 to 20th July 2016 were eligible (n=98 patients). Patients who declined trial entry, received incomplete CRT, were unable to undergo MRI, were pregnant or with prior malignancy were excluded (n=64). Furthermore, those with incomplete two year clinical, imaging and (where applicable) histopathological follow-up were excluded for subsequent analysis (n=5), as were those with inadequate/incomplete baseline and follow up MR datasets (n=4). In total, 25 patients were accrued (mean age 54.4 years; range 15-78 years) of whom ten were male (mean age 57.7 years; range 33-79 years; 10/25 male) prior to commencing systemic chemotherapy (median interval 4 days; range 0-13 days). Patient demographics and summary of primary disease and lesion distribution is outlined in Table 1.

Patient group	Total Number Patients (%)	Mean age (yrs) (range)	Female (%)	Tumour type (%)	Total lesions & locations
Total	25 (100%)	54.4 (15-78)	15 (60%)	Breast- 6 (24%) STS- 10 (40%) BS- 9 (ES 6/OS 3) (36%)	95 liver 33 muscle/soft tissue 29 extraosseous 11 lung 6 pleural 6 nodal 5 bone 5
Partial Response	12 (48%)	52.2 (18-78)	9 (75%)	Breast- 4 (33%) STS- 2 (17%) BS- 6 (ES 4/OS 2) (50%)	32 liver 14 muscle/soft tissue 7 extraosseous 6 lung 2 bone 3
Non-response	13 (52%)	56.2 (15-63)	6 (46%)	Breast- 2 (15%) STS- 8 (62%) BS- 3 (ES 2/OS 1) (23%)	63 liver 19 muscle/soft tissue 22 extraosseous 5 lung 4 pleural 6 nodal 5 bone 2

Table 3.1 Summary of patient demographics, primary tumour and individual lesion distribution (STS; soft tissue sarcoma, BS; bone sarcoma; ES; Ewing sarcoma, OS; osteosarcoma).

3.3.2 Whole body magnetic resonance imaging technique:

Imaging was performed using a single 3.0 T wide-bore MR scanner (Ingenia; Phillips Healthcare, Best, Netherland) using the manufacturer head coil, two anterior surface coils (2*16 elements) and table embedded posterior coils.

All subjects were imaged supine with arms by their side. 'Full body' coverage (vertex to mid-thigh) was obtained through a multi-station acquisition of contiguous body regions. Coronal WB-MRI T1 weighted mDixon imaging at two flip angles (2.5 and 15 degrees) with coronal dual time to repetition (TR) B1 mapping over the same stations were performed and complimented by axial T2 weighted turbo spin echo (TSE), axial DW WB-MRI (using 2 b-values: b50 and 900 s/mm2). Total scan time was approximately 60 minutes. MR parameters are given in Table 2.

Parameters	**T2-TSE**	**DWI (b50, 900)**	**mDixon- 2.5+15 flip angles**	**B1 map**
Plane	Transverse	Transverse	Coronal	Coronal
TE (ms)	80	69	1.15/2.3	2.8
TR (ms)	1228	6300	3.5	10/50 (extension 90)
Field of view	500x300	500x306	500x300	500x300
Voxel size (mm*mm)	1x1	4x4.2	2.1x2.1	2.1x2.1
Number of slices	40	40	120	60
Slice thickness (mm)	5	5	5	10
Acquisition matrix	500x286	124x72	240x240	120x120
Echo train length	91	39	2	1
Acceleration factor (SENSE)	2	2.5	2	3
Pixel bandwidth (Hz)	537	3369	1992	3356
Scan time (s)	47	152	17	172

Number of stations	5	5	4 (5)	4(5)
Total FH coverage (mm*mm)	987.75 (10% overlap)	987.75 (10% overlap)	1388 (10% overlap)	1388

Table 3.2 Table showing whole body MRI (WBMR) parameters. T2-TSE: T2-weighted turbo spin echo, mDixon: modified Dixon, DWI diffusion weighted imaging, TE: time of echo, TR: repetition time, SENSE: sensitivity encoding.

3.3.3 Reference Standard Response Assessment:

Two experienced radiologists (with seven and thirteen years' experience respectively) in consensus identified 'measurable' soft tissue lesions in each patient with the benefit of baseline and post treatment conventional imaging (though blinded to subsequent outcome information) recording single largest diameter of each lesion using anatomic sequences according to established 'Response Evaluation Criteria in Solid Tumors' (RECIST 1.1) 'target' lesion criteria [1]. Lesions were followed to the end of treatment (nominally post six cycles of chemotherapy in the case of metastatic disease and two to four cycles in bone tumours subsequently for resection) and categorised into 'partial response' (>30% decrease in maximum diameter), progressive disease (>20% increase) or stable disease [1]. Partial responding (PR) lesions were compared to non-responding lesions (NR; i.e. progressive disease and/or stable disease) in all analyses. Bone lesions without significant extra-osseous soft tissue component fall outside 'measurable' lesion status by RECIST 1.1 but were treated in a similar manner for the purposes of this analysis however in the case of osteosarcoma or Ewing's sarcoma bone lesions without extra-osseous component (more commonly OS) where the primary lesion was resected post systemic therapy (OS 3/3 patients and ES 4/6 patients) any response assessment by size was superseded by histological analysis to categorise response. The histological response at resection (nominally week 11; usually after 2 or more cycles of chemotherapy) is the most important prognostic factor in determining long-term outcome for patients with localized disease at diagnosis. Those with a "good response", i.e. >90% necrosis on histological analysis or "poor response", i.e. <90% necrosis, have 5-year event free survivals of approximately 75% and 40% respectively [16].

In total twenty five patients had undergone sufficient follow-up for response categorisation. In total 95 lesions (median 4 per patient; range 1-8) were considered; of which 32 lesions in twelve patients demonstrated end of treatment partial response. All twenty five patients underwent a baseline research WBMR as well as second ‘early’ response assessment post two cycles of systemic therapy.

3.3.4 Image Analysis

Volumetric T1 maps (Fig 3.1) were generated using dual flip angle (2.5 and 15 degrees) 2-point modified (m)Dixon 'in phase' images by a linear fitting algorithm [29] with MATLAB (v7.13) incorporating a separately acquired dual TR B1 map [30]. Apparent diffusion coefficient (ADC) maps were calculated by mono-exponential curve fitting of mean signal intensities of all b-values from diffusion weighted sequences.

A third radiologist (with 9 years’ experience) aware of the location of each lesion though unaware of the parametric maps or follow-up data referenced the anatomical images to volumetrically contour the pre-defined lesions on T1 weighted modified Dixon in phase images (flip angle 15) and diffusion weighted (b value 50ms) images using dedicated software (Jim 5.0, Xinapse systems, Thorpe, Waterville, UK) excluding areas of necrosis. These segmented volumes were then transferred to T1 and monoexponential ADC parametric maps respectively.

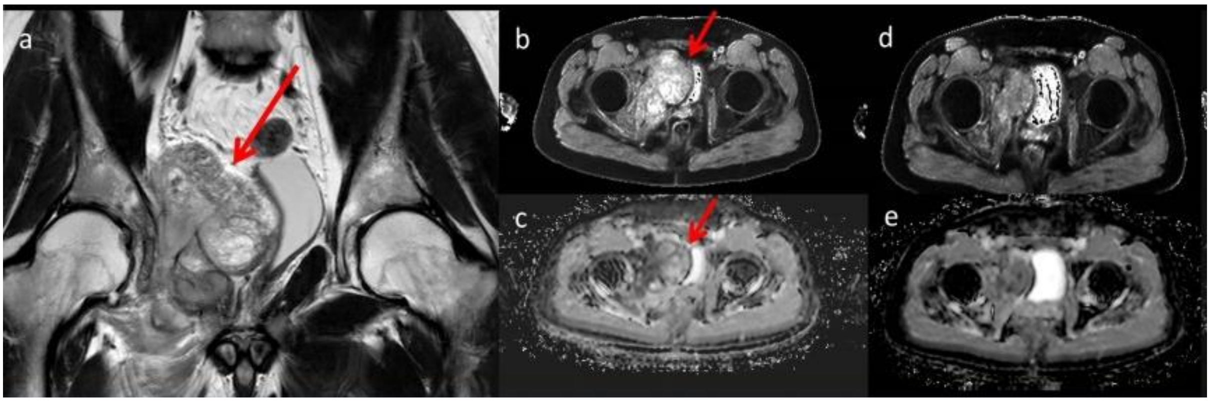


Figure 3-1 Twenty four year old male patient with extra-osseus Ewing sarcoma right pelvis undergoing partial response by the end of two cycles of chemotherapy: (a) baseline (pre-treatment T2 TSE coronal, (b) baseline T1 map (axial), (c) baseline ADC map (d) axial T1 map post 2 cycles of chemotherapyand € axial ADC map post 2 cycles of chemotherapy.

3.3.5 Statistical Analysis

All analyses compared the two outcome groups (i.e. PR versus NR) to assess for significant pretreatment differences between histographic T1 and ADC metrics and were performed on an all lesion basis. Additionally, differences in early post second cycle T1 and ADC values and percentage change from baseline were analysed. Lesional histographic T1 and ADC differences were analysed using a linear mixed model (Stata v13, StataCorp LLC, College Station, Texas, USA) to account for multiple samples per patient. Fitting was performed using response as the fixed factor and patient as the random factor. Means and 95% confidence intervals (CI) were estimated from the model for each patient response group. Where the data was right skewed the log measurement was used. Where the data was right skewed including negative values, a small constant was added to all values before the log transformation was applied with estimated means and 95% CIs back-transformed. Additionally, per lesion differences in early post second cycle T1 and ADC values and percentage change from baseline were analysed using univariate parameters and in combination using multivariate ROC-AUC analysis accuracy for prediction of partial response determined by receiver operating characteristic (ROC) area under curve (AUC) analysis and 95% confidence intervals.

3.4 RESULTS

3.4.1 Histographic parameter changes between baseline and cycle 2:

Table 3.3 summarises histographic values (median, skewness and kurtosis) and ranges of baseline and post 2 cycle T1 and ADC values and percentage changes post 2 cycles in both partial responding and non-responding lesions. Percentage change in T1 value after two cycles chemotherapy in partial responding lesions was significantly lower at -22% versus -1% in non-responders (p=0.001) and was the best predictor of response to treatment. ADC percentage change post 2 cycles of chemotherapy was significantly higher at +70% in partially responding lesions versus 9.8% in non-responding lesions (p=0.003). The actual ADC value post 2 cycles was also significantly higher in partial responding lesions at 1.77 versus 1.28x10-3mm2/s in lesions demonstrating no response (p=0.001). By contrast, the absolute T1 value post 2 cycles of chemotherapy was not significantly different when comparing partial responding to non-responding lesions (p=0.17). These findings are demonstrated in Figure 2. Percentage change in lesion T1 values after two cycles was the best predictor of partial response with ROC-AUC 0.89. This was followed by the actual ADC value post 2 cycles, percentage change in ADC from baseline and baseline ADC with ROC-AUC of 0.86, 0.84 and 0.70 respectively. These are shown in Figure 3.3. The best performing bivariate pairing was percentage change in T1 post two cycles combined with ADC value post two cycles giving ROC-AUC 0.90 (0.78-0.96) which did not significantly alter accuracy compared with best univariate analysis.

Patient group	**All Patients**	**Partial Response**	**Non-response**	**P value (linear mixed)**
Median % change longest plane diameter post 6 cycles	-2.00% (-78% to 119%)	-59% (-78% to 30%)	9.40% (-28% to 119%)	<0.001
Baseline mean of median T1 (ms)	2467 (1035-3998)	2765 (1678-3126)	2297 (1035-3999)	0.09
Baseline mean of T1 kurtosis	0.23	0.18	0.25	0.53
Baseline mean of T1 skewness	0.65	0.71	0.61	0.27
Baseline mean of median ADC (x10-6mm2/s)	1063 (456-2507)	885 (456-1450)	1115 (578-2507)	0.02
Baseline mean of ADC kurtosis	0.90	0.72	0.88	0.31
Baseline mean of ADC skewness	0.65	0.70	0.58	0.33
Post 2 cycles mean of median T1 (ms)	2271 (937-3141)	2011 (937-2924)	2433 (1068-3141)	0.17
Post 2 cycles mean of T1 kurtosis	0.56	0.62	0.45	0.34
Post 2 cycles mean of T1 skewness	0.24	0.32	0.19	0.15
Post 2 cycles mean of median ADC (x10-6mm2/s)	1571 (966-2485)	1770 (1285-2485)	1282 (966-1988)	0.001
Post 2 cycles mean of ADC kurtosis	0.68	0.64	0.72	0.18
Post 2 cycles mean of ADC skewness	0.35	0.29	0.38	0.63
Post 2 cycles median % change T1	-6.40% (-50% to 24%)	-22% (-50% to -3.4%)	-1.00% (-35% to 24%)	0.001
Post 2 cycles median % change ADC	43% (-8.6% to 182%)	70% (11% to 182%)	9.80% (-8.6% to 91%)	0.003

Table 3.3 Summarising histographic T1 and ADC values at baseline, after two cycles of chemotherapy and percentage change compared to baseline (parentheses range) as well as uni-dimensional size change denoting response status. Panes coloured yellow denote parameters demonstrating significant differences between 'partial' responding and non-responding lesions (p<0.05)

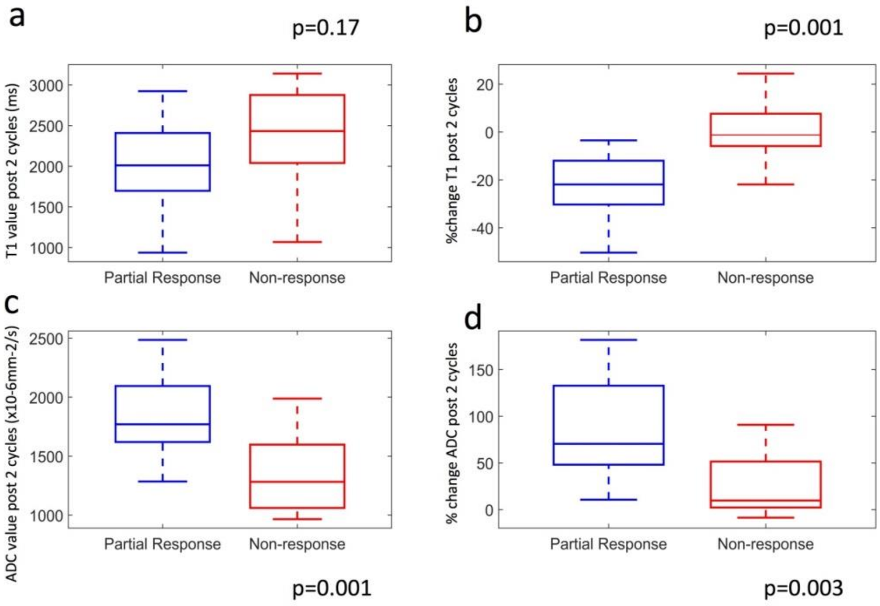


Figure 3-2 Box plots showing differences between responding and non-responding lesions (a) T1 value post 2-cycles, (b) Percentage change in T1 value post 2-cycles, (c) ADC value post 2 cycles, (d) Percentage change ADC value post 2 cycles. Box indicates interquartile range, line median and whiskers most deviated range. Linear mixed method p-values shown.

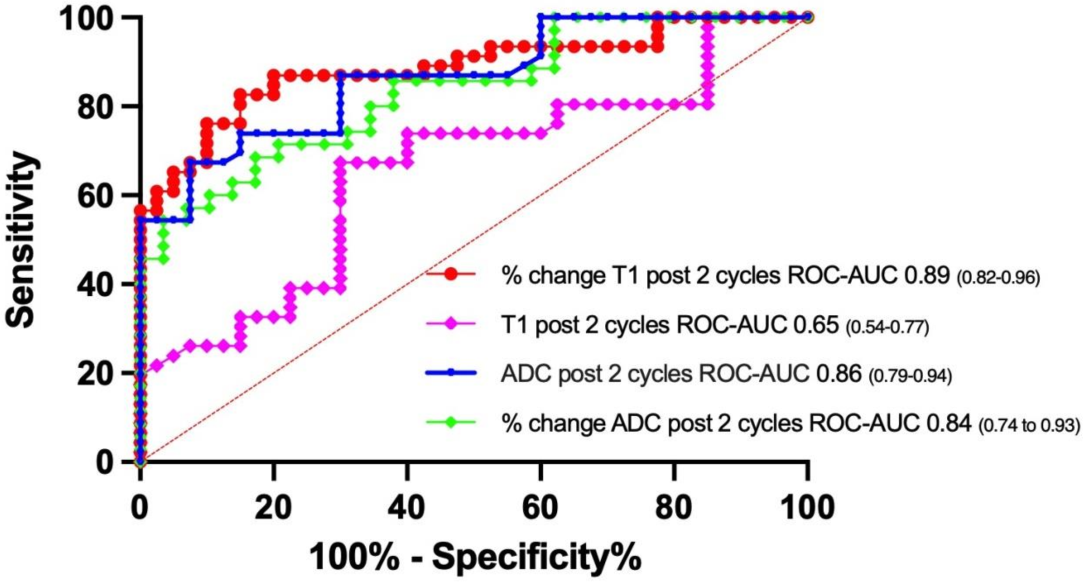


Figure 3-3 Receiver operating characteristics of the actual T1 and ADC values after 2 cycles of chemotherapy as well as percentage change for prediction of response with area under curve (AUC) values as shown (parentheses indicate 95% confidence intervals).

3.4.2 Baseline histographic parameters as predictors of lesion response:

In terms of prediction of response by comparison of baseline T1 and ADC values, partial responding lesions had slightly higher median T1 value of 2765ms versus 2267ms in non-responding lesions, though this did not reach statistical significance (p=0.09). Baseline median ADC values were significantly lower in lesions going on to demonstrate partial response 0.89x10-3mm2/s compared with 1.12x10-3mm2/s in non-responding lesions (p=0.02). Other histographic parameters for both T1 and ADC were not significantly different at baseline (or post 2 cycles of chemotherapy).

3.5 DISCUSSION

This study has demonstrated that prediction of eventual tumour response by early lesion T1 value change during systemic therapy is feasible with whole-body coverage, and is comparable in performance with more widely studied quantitative imaging biomarkers such as ADC. The finding that those tumours which demonstrate partial response to chemotherapy by the end of treatment show decreases in T1 of the order of 20% confirms preclinical tumour model work [27]. This effect has been postulated to reflect a decrease in remaining viable/proliferating tumour cells due to cell destruction with consequent release of proteins and/or metals (causing T1 relaxation) [31]. Interestingly, baseline ADC was significantly lower in those lesions which underwent partial response to chemotherapy compared with non-responding lesions (as has been demonstrated in other in vivo studies) suggesting lesion baseline ADC may have value in response prediction; however baseline T1 did not differ between the two groups. This may be the result of pooling of variable soft tissue and bone lesions (both of which did demonstrate a wider range of T1 values) and/or reflect study size to which ADC may be more robust. ADC has been extensively studied as an imaging biomarker to assess and predict response. Regardless of the definition of response used in these reports (i.e. good or complete response), all studies reported an increase in mean tumour ADC after systemic therapy [32] which is thought to be due to radiation-induced cellular damage and necrosis. Disruption of cell membranes reduces the diffusion restriction and therefore increases the ADC. Our study is in line with observations that both the final post therapy ADC and the relative increase in ADC were typically higher in the favourable response groups, with statistically significant results in the majority of studies [33]. The current study did not demonstrate significant differences in responding lesional histographic parameters compared with non-responding lesions. Histogram analysis has been the subject of recent work whereby the whole spectrum of ADC values within the tumour are analyzed, allowing extraction of not only mean (or median) values but also additional parameters reflective of this distribution. Based on the limited evidence available these parameters do not seem to offer a clear additional benefit thus far [34].

To our knowledge this is the first study to globally assess tumour response on a whole body basis in humans with the aim of assessing baseline and change in tumoral T1 values to predict response to systemic chemotherapy. In line with discussed biomarker validation theory, this contributes to the qualification process of T1 tumour value as a biomarker for response assessment. Prior work (contributing to analytical validation) has demonstrated that the method used in our study is adequately reproducible and repeatable across a range of T1 values to detect changes in T1 which might predict tumour response to systemic treatment before conventional size assessment based on these findings [30].

This study has several limitations including the pooling of lesions across a range on tumour types which may confound resulting changes in quantitative parameters. However this approach may make the findings more robust in terms of the ability of this QIB to differentiate response in lesions early during therapy. Furthermore we have accounted for multiple lesions per patient in our analysis to compensate for multiple samples. In addition for the purposes of this analysis we opted for binary categorisation of response as opposed to stratification of response as described by traditional tumour response criteria. One of the purposes of examining T1 as a QIB was the promise of potential use in assessing response in non-measurable disease (e.g. bone deposits) - these though low in number, these lesions followed the general trend of high baseline T1 predicting response as well as decreasing T1 early in treatment eventually resulting in response to therapy.

The additional quantitative T1 mapping information described in this study can be obtained on a whole body basis within a clinically feasible total scan duration (adding ~20mins for the T1 mapping component to a multiparametric whole body MR acquisition) and may provide complementary biophysical information to direct therapy in terms of prediction and earlier response assessment. Future work is, however, required to understand the contrast mechanisms underpinning the observed T1 changes as well as differences and in common with most ongoing quantitative studies, there is a clear remit for larger cohort multicentre studies across multiple platforms to establish repeatability and reproducibility for clinical utility.

3.6 REFERENCES

1. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009 45(2): 228-47.

2. A.R. Padhani. Functional MRI for anticancer therapy assessment. European Journal of Cancer 2002 38: 2116-2127.

3. Stacchiotti S, Collini P, Messina A, Morosi C, Barisella M, Bertulli R, Piovesan C, Dileo P, Torri V, Gronchi A, Casali PG. High-grade soft-tissue sarcomas: tumor response assessment--pilot study to assess the correlation between radiologic and pathologic response by using RECIST and Choi criteria. Radiology 2009 251(2):447-56.

4. Choi H. Response evaluation of gastrointestinal stromal tumors. Oncologist 2008 13(Suppl 2):4-7.

5. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. J Nucl Med. 2009 May;50 Suppl 1(Suppl 1):122S-50S. doi: 10.2967/jnumed.108.057307. PMID: 19403881; PMCID: PMC2755245.

6. Toro J, Travis L, Wu HJ, et al. Incidence patterns of soft tissue sarcomas, regardless of primary site, in the Surveillance, Epidemiology, and End Results Program, 1978–2001: an analysis of 26,758 cases. Int J Cancer 2006 119: 2922-30.

7. Pisters P. Staging and Prognosis, in Pollock RE (ed). American Cancer Society Atlas of Clinical Oncology: Soft Tissue Sarcomas. Hamilton, Ontario: BC Decker, Inc.; 2002:80–88.

8. Pervaiz N, Colterjohn N, Farrokhyar F, Tozer R, Figueredo A, Ghert M. A systematic meta-analysis of randomized controlled trials of adjuvant chemotherapy for localized resectable soft-tissue sarcoma. Cancer 2008 113(3): 573–581.

9. van Glabbeke M, van Oosterom AT, Oosterhuis JW et al. Prognostic factors for the outcome of chemotherapy in advanced soft tissue sarcoma: an analysis of 2,185 patients treated with anthracycline- containing first-line regimens—an European organization for research and treatment of cancer soft tissue and bone sarcoma group study. Journal of Clinical Oncology 1999 17(1): 150-157.

10. Mirabello L, Troisi RJ, Savage SA. International osteosarcoma incidence patterns in children and adolescents, middle ages and elderly persons. International Journal of Cancer 2009 125(1): 229-234.

11. National Institute for Health and Clinical Excellence (NICE). Guidance on Cancer Services—Improving Outcomes for People with Sarcoma: the Manual [Needs assessment, Evidence review, List of recommendations] 2006, (Developed by the National Collaborating Centre for Cancer).

12. Stiller CA, Craft AW, Corazziari I. Survival of children with bone sarcoma in Europe since 1978: results from the EUROCARE study. European Journal of Cancer 2001 37(6): 760-766.

13. Bielack S, Carrle D, and Casali PG. Osteosarcoma: ESMO clinical recommendations for diagnosis, treatment and follow-up. Annals of Oncology 2009 20(4): iv137–iv139.

14. Carrle D, Bielack SS. Current strategies of chemotherapy in osteosarcoma. International Orthopaedics 2006 30(6): 445-451.

15. Burgert Jr. EO, Nesbit ME, Garnsey LA et al. Multimodal therapy for the management of nonpelvic, localized Ewing’s sarcoma of bone: intergroup study IESSII. Journal of Clinical Oncology 1990 8(9): 1514-1524.

16. Bielack SS, Kempf-Bielack B, Delling G et al. Prognostic Factors in High-Grade Osteosarcoma of the Extremities or Trunk: An Analysis of 1,702 Patients Treated on Neoadjuvant Cooperative Osteosarcoma Study Group Protocols J Clin Oncol 2002 20: 776-790.

17. . National Institute for Health and Clinical Excellence. Advanced breast cancer: diagnosis and treatment. Clinical Guidelines CG81. Available at: http://www.nice.org.uk/TA81. [accessed 16.03.14].

18. Office for National Statistics (2008) Cancer statistics registrations: registrations of cancer diagnosed in 2005, England. Series MB1 number 36. London: Office for National Statistics.

19. Welsh Cancer Intelligence and Surveillance Unit (2008) Cancer incidence in Wales 1992-2002. Cardiff: Welsh Cancer Intelligence and Surveillance Unit.

20. LCA Breast Cancer Clinical Guidelines - London Cancer Alliance October 2013. http://www.londoncanceralliance.nhs.uk/media/59178/FINAL%20LCA_BreastGuidelines_231013%20revised%20(3).pdf

21. Jakobsen I, Kaalhus O, Lyng H, Rofstad EK. Detection of necrosis in human tumour xenografts by proton magnetic resonance imaging. Br J Cancer 1995 71: 456-61.

22. Gambarota G, Veltien A, van Laarhoven H, et al. Measurements of T1 and T2 relaxation times of colon cancer metastases in rat liver at 7 T. Magn Res Mat Phys Biol Med 2004 17: 281-7.

23. Thomsen C, Sørensen PG, Karle H, Christoffersen P, Henriksen O. Prolonged bone marrow T1-relaxation in acute leukaemia. In vivo tissue characterization by magnetic resonance imaging. Magn Reson Imag 1987 5: 251-7.

24. Jensen KE, Sørensen PG, Thomsen C, et al. Magnetic resonance imaging of the bone marrow in patients with acute leukemia during and after chemotherapy. Acta Radiologica 1990 31:361-9.

25. Santoni R, Bucciolini M, Chiostrini C, Cionini L, Renzi R. Quantitative magnetic resonance imaging in cervical carcinoma: a report on 30 cases. Br J Radiol 1991 64: 498-504.

26. Smith SR, Roberts N, Percy DF, Edwards RH. Detection of bone marrow abnormalities in patients with Hodgkin's disease by T1 mapping of MR images of lumbar vertebral bone marrow. Br J Cancer 1992 65: 246-51.

27. McSheehy PM1, Weidensteiner C, Cannet C, Ferretti S, Laurent D, Ruetz S, Stumm M, Allegrini PR. Quantified tumor t1 is a generic early-response imaging biomarker for chemotherapy reflecting cell viability. Clin Cancer Res. 2010 16(1): 212-25.

28. Barnes A et al. ISMRM 2014 #3241.

29. Wang D, Shi L, Wang YX, Yuan J, et al. Concatenated and parallel optimization for the estimation of T1 map in FLASH MRI with multiple flip angles. Magn Reson Med 2010 63(5): 1431-6.

30. Yarnykh VL. Optimal radiofrequency and gradient spoiling for improved accuracy of T1 and B1 measurements using fast steady-state techniques. Magn Reson Med 2010 63(6): 1610-26.

31. Weidensteiner C, Allegrini PR, Sticker-Jantscheff M, et al. Tumour T1 changes in vivo are highly predictive of response to chemotherapy and reflect the number of viable tumour cells--a preclinical MR study in mice. BMC Cancer. 2014 Feb 14;14:88.

32. Sathyakumar K , Chandramohan A , Masih D , Jesudasan MR , Pulimood A , Eapen A . Best MRI predictors of complete response to neoadjuvant chemoradiation in locally advanced rectal cancer . Br J Radiol 2016. ; 89 : 20150328 . doi: 10.1259/bjr.20150328

33. Messina C, Bignone R, Bruno A, et al. Diffusion-Weighted Imaging in Oncology: An Update. Cancers (Basel). 2020;12(6):1493. Published 2020 Jun 8. doi:10.3390/cancers12061493

34. Schurink NW, Lambregts DMJ, Beets-Tan RGH. Diffusion-weighted imaging in rectal cancer: current applications and future perspectives. Br J Radiol. 2019;92(1096):20180655. doi:10.1259/bjr.20180655

Chapter 4 Clinical Validation (Qualification) of Novel Quantitative MR Biomarker in Tumour Response Assessment and Prediction as part of whole body MRI assessment- metastatic prostate cancer

CONTRIBUTION:

Primary author responsible for conception (of T1 mapping element), data acquisition/analysis, drafting, revision, final approval and accuracy/integrity.

4.1 RUNNING TITLE

In vivo measurement of tumour T1 relaxation time using whole body multiple flip angle method can assess and predict extra-prostatic disease response to systemic cancer therapy in radiorecurrent metastatic prostate cancer.

4.2 INTRODUCTION

Prostate cancer (PCa) is the leading male cancer in the United Kingdom and a principal cause of cancer-related mortality with around 46,700 cases diagnosed in 2014 [1]. The incidence of PCa has increased markedly over the past 20 years and it is projected to rise by 12% between 2014 and 2035 to 233 per 100,000 males [1]. Common treatments options for localised prostate cancer are external beam radiation therapy (EBRT), brachytherapy (BRT) or radical prostatectomy [2]. However, biochemical recurrence (BCR) following radiotherapy occurs in 25% of treated men within 5 years, manifesting as a rising prostate-specific-antigen (PSA) [3]. This is strictly a biochemical diagnosis, most commonly defined as an increase in serum PSA of 2.0 ng/ mL above the nadir [4]. Once a PSA relapse has been diagnosed, it is important to determine whether the recurrence has developed at local or distant sites and hence BCR following therapy often initiates an imaging based assessment of local and metastatic disease in patients.

Local recurrence of PCa is investigated by prostate multi-parametric magnetic resonance imaging (mp-MRI) [5] whilst metastatic disease is assessed by multimodality imaging. The standard workup to detect PCa metastases usually includes 99mTc bone scan and chest/ abdomen/pelvis (CT-CAP) scan (or where available 18F-choline or 68Ga-PSMA PET-CT) [6]. The workup to exclude distant metastasis is of greatest significance as it is estimated that about half of those who develop radio-recurrent disease have distant metastases (either overt on imaging or micro-metastases) [7]. It is also believed that up to 50% of men who are considered to be free of metastases by conventional staging methods have micro-metastatic disease [8]. This is reflected in prostate cancer recurrence following EBRT or BRT, where biochemical relapse commonly precedes clinical detection of metastases by an average of 7–8 years [9]. Prospective trials examining the use of whole body MRI to stage radiorecurrent disease have been prompted by recognition of the need for an imaging modality that is capable of improved sensitivity for metastatic disease detection than current conventional imaging [10–12]. Previous meta-analysis has reported the sensitivity of 99mTc bone scan to be less than 50% [13]. Similarly, studies have shown the sensitivity of 18F-choline PET to be as low as 50% for the detection of nodal disease when PSA is < 5 ng/ml [14]. Furthermore, CT Scan sensitivity for nodal disease detection in prostate cancer is reported to be much lower with a meta-analysis reporting a sensitivity of 30% [15]. As part of these studies, attention is also being turned to the ability of quantitative parameters at baseline and their change during treatment can predict and/or assess treatment response.

Several options exist for salvage of biochemical recurrence after definitive external beam radiation therapy when disease is confined to the prostate. Because the original source of the disease, the prostate, remains in situ, many salvage therapies continue to be locally directed. Metastatic disease in this setting may be managed by active surveillance and/or cytoreductive local salvage for low risk individuals or systemic therapy in higher risk patients. Several therapeutic approaches have received recent regulatory approval for use in men with mPCa. In addition to androgen deprivation and traditional chemotherapeutic treatments, there are now additional agents with varying mechanisms of action showing survival benefit in this patient population; including agents that target the androgen axis (e.g. enzalutamide and abiraterone), stimulate the immune system (e.g. sipuleucel-T), have a chemotherapeutic effect (e.g. docetaxel and cabazitaxel), and alpha-particle emitter that directly targets bone metastases (radium-223). The optimal use of these therapies remains debatable with the presence, volume, and location of metastases being important determinants.

Systemic oncological therapies are associated with variable efficacy, side-effects and financial cost. The current standard for assessing treatment response of tumours based on imaging is one-dimensional size measurement as defined by the Response Evaluation

Criteria In Solid Tumours (RECIST) 1.1 [16]. Whilst acceptable for certain solid tumour and conventional chemotherapies, the RECIST criteria fail to adequately characterise response of cystic lesions or indeed the response of solid lesions to newer anti-angiogenic treatments [17]. In addition specific metastatic deposits (e.g. bone disease which are common sites for mPCa) are non-measurable. One proposed improvement on the RECIST criteria was the Choi criteria which incorporated density and intensity changes by CT (and MR) imaging, in addition to changes in size [18]. Finally, to address the inadequacies of the RECIST and Choi criteria, the use of a standardized uptake value (SUV) obtained by PET was incorporated into response criteria to better assess tumour metabolism [19]. Thus, the revised RECIST 1.1 criteria endorsed the functional assessment of tumours following treatment as complementary to anatomic evaluations [20]. More accurate assessments of therapy response (including the detection of primary and secondary resistance and heterogeneity of response) could aid in the rationale development of targeted therapies [21]. Two fairly recent reviews have indicated that WB-MRI is suitable for wider deployment in disease detection settings, given its established test performance, potential for wide availability, and multi-organ evaluation capabilities [22,23]; with the latter under the auspices of the European Organisation for Research and Treatment of Cancer position paper concluding that MRI offers a good “one size fits all” solution for assessing therapy effectiveness in this cohort. Prostate cancer frequently metastasizes to lymph nodes (both locoregional and distant) and the bones- both are frequently problematic for conventional imaging to detect and furthermore assess response to treatment. Diffusion weighted MRI (DW-MRI) is sensitive to thermal motion (diffusion) of water molecules. In biologic tissues barriers such as endothelium, cell membranes, components of the extracellular matrix and intracellular organelles restrict diffusion; increase or decrease in these barriers modifies the degree of water diffusion leading to a reduction or retention of MR signal. Tumour foci are visualised as increased signal intensity on DW-MRI images with a corresponding decrease in the measured apparent diffusion coefficient (ADC), which represents the rate of signal loss with increasing diffusion weighting [24]. As the diffusion properties of bone metastases are significantly different to age-matched normal marrow [25], DW-sequences are now almost routinely used as an adjunct to conventional T1-W images. Preclinical studies have provided critical insight into DW-MRI changes with treatment [26]. In a mouse model of prostate cancer treated with docetaxel, increased diffusion on functional diffusion maps (fDM) after treatment corresponded to local loss of cell viability at histology treatment. Lee [27]and Rozel [28] reported a significant increase in ADC values over time in responders; however, in a rat model of bone metastases from breast cancer (cells MDA-MB-231), ADC was not significantly affected in treated rats compared to controls despite significant changes in DCE-MRI parameters [29]. In clinical research, a study in 26 patients demonstrated that mean ADC in the majority of lesions from both responders and non-responders showed a significant increase at 12 weeks, although some lesions in both groups demonstrated a fall in ADC greater than the limits of reproducibility of the technique [25]. Whilst there are multiple single centre predominant studies that support WB-MRI DWI for the evaluation of prostatic bone metastases, this technique is not yet universally recognized, as evidenced in a recent work by Wieder et al. showing the superiority of 11C-choline PET/CT in detecting bone lesions, though no difference was reported in the detection of lymph node metastasis [30]. Other biomarker developments are underway to varying degrees.

Quantitation of T1 may be used to characterise normal and pathological tissue and there is increasing interest in its use in producing robust dynamic contrast enhanced (DCE) quantitative parameters. Whole body (WB) T1 quantitation presents several difficulties (most notably respiratory artefact and large field of view requirement). Several techniques have been explored; the most commonly used being inversion recovery (IR; which remains the reference standard method), multiple flip angle (MFA), and Look-Locker (LL) techniques. Several studies using different T1 estimation methodologies showed that the T1 of experimental tumours was greater than in normal tissue and was considered to reflect predominantly the extracellular space [31-33], with a larger interstitial compartment giving a longer (higher) T1. In untreated tumours, a low tumour T1 has been correlated to increased necrosis [32], low water content, high levels of soluble protein [31], and low proliferation. In four different quantitative T1 studies, only decreases were detected in response to successful therapy with radiation or cytotoxics [34-37]. McSheehy et al (2009) used animal tumour models to show five anticancer drugs (with different mechanisms of action) caused a rapid decrease in T1 that often preceded and was strongly correlated with the change in tumour size suggesting that a fractional change in T1 (ΔT1) was a useful generic early-response marker. Furthermore, ΔT1 did not occur in tumours resistant to that drug, and ΔT1 was positively correlated with proliferation markers (Ki67 and choline), suggesting that ΔT1 indirectly reflected the number of viable cells in a solid tumour [37].

Repeatability of a whole body B1 corrected multiple flip angle (MFA) method has been reported in healthy volunteers with co-efficients of variation across tissue types between 5-10% when selecting two appropriate flip angles [38].

This study evaluated metastatic T1 relaxation times using multiple flip angle method prior to, and during, systemic therapy for prediction and evaluation of response to therapy in patients being worked up for biochemical relapse in radiorecurrent prostate cancer. This was embedded within an overarching trial, "localising occult prostate cancer metastasis with advanced imaging techniques (LOCATE trial)", a prospective cohort, observational diagnostic accuracy trial investigating whole–body magnetic resonance imaging in radio-recurrent prostate cancer which primarily aimed to investigate the diagnostic performance of WB-MRI for detection of nodal and metastatic disease compared to current standard multimodality imaging comprising 18F-choline PET-CT and 99mTc bone scan (+/− CT-CAP) in patients with the radio-recurrent PCa [39].

4.3 MATERIALS AND METHODS

Institutional review board approval was granted and patients' informed consent (written and verbal) was obtained prior to entry into this prospective study (LOCATE trial received UK Research Ethic Committee (REC) approval from the National research ethics service (NRES) Committee London-Chelsea with REC reference 15/LO/0776).

4.3.1 Patient cohort

Men aged above 18 years who had undergone previous EBRT or brachytherapy (BRT) with or without neo-adjuvant/adjuvant hormone therapy and under monitoring were found to have biochemical relapse as determined by increase in PSA (2.0 ng/ mL above the nadir) were invited to participate between 18/09/2015 to 11/01/2018. Exclusion criteria were men unable to have MRI scan, or in whom artefact would significantly reduce quality of MRI, as well as men unable to give informed consent.

In total 132 men were recruited into the LOCATE trial. Within this substudy examining treatment response, 23 men were excluded because they did not undergo a follow up WB-MRI. Of the remaining 109, only patients with either locoregional or metastatic lymph node disease or other extraprostatic metastatic disease (see below) which by expert consensus required systemic therapy were considered (N=48). A further eleven men were excluded due to lack of comprehensive follow-up data, 2 patients were excluded due to poor MR image quality compromising and 3 men were excluded due to MR occult disease at baseline. In total 32 men were accrued in this study (mean age 71.3 years; range 55-82 years) and underwent systemic therapy consisting of predominantly androgen deprivation therapy and/or systemic chemotherapy (e.g. docetaxol).

4.3.2 Multi-parametric whole body magnetic resonance imaging

Recruited patients underwent WB-MRI scan on presentation with BCR and then nominally at 12 months (follow-up). The follow up scan was performed at median 359 days (range 343 to 378 days) for these patients. All subjects were imaged supine with arms by their side. 'Full body' coverage (vertex to mid-thigh) was obtained through a multi-station acquisition of contiguous body regions. Coronal WB-MRI T1 weighted modified Dixon imaging at two flip angles (2.5 and 15 degrees) with coronal dual time to repetition (TR) B1 mapping over the same stations were performed and complimented by axial T2 weighted turbo spin echo (TSE), axial DW WB-MRI (using 2 b-values: b0 and 1000 s/mm2). The coronal mDixon was repeated post intravenous contrast at 15 degree flip angle. Table 1 demonstrates relevant MR parameters.

Parameters	**T2-TSE**	**DWI (b0, 1000)**	**mDixon- 2.5+15 flip angles**	**B1 map**
Plane	Transverse	Transverse	Coronal	Coronal
TE (ms)	80	69	1.15/2.3	2.8
TR (ms)	1228	6300	3.5	10/50 (extension 90)
Field of view	500x300	500x306	500x300	500x300
Voxel size (mm*mm)	1x1	4x4.2	2.1x2.1	2.1x2.1
Number of slices	40	40	120	60
Slice thickness (mm)	5	5	5	10
Acquisition matrix	500x286	124x72	240x240	120x120
Echo train	91	39	2	1

length				
Acceleration factor (SENSE)	2	2.5	2	3
Pixel bandwidth (Hz)	537	3369	1992	3356
Scan time (s)	47	152	17	172
Number of stations	5	5	4 (5)	4(5)
Total FH coverage (mm*mm)	987.75 (10% overlap)	987.75 (10% overlap)	1388 (10% overlap)	1388

Table 4.1 Table showing whole body MRI (WBMR) parameters. T2-TSE: T2-weighted turbo spin echo, mDixon: modified Dixon, DWI diffusion weighted imaging, TE: time of echo, TR: repetition time, SENSE: sensitivity encoding.

4.3.3 Conventional Imaging

Recruited patients had standard imaging investigations according to local and national guidelines and as below:

1. Multi-parametric prostate MRI: A standard prostate mp-MRI protocol as defined by the UK consensus guidelines on prostate MRI [40] was used to locally stage prostate cancer. Conventional T1 and T2-weighted images of the prostate supplemented with DW-MRI (as per hospital site protocol); and +/− dynamic contrast enhanced (DCE) T1-weighted images.
2. 99mTechnetium bone scan: performed using 99mTc labelled diphosphonates administered through intravenous injection. For prostate cancer patients with suspected bone metastases, the standard protocol employed at the hospital site was used. As a guide, whole body imaging was conventionally performed with anterior and posterior views, 256 × 1024 matrix and energy window(s) of 140 KeV.
3. Positron emission tomography-computed tomography (PET-CT): 18F-choline PET-CT was acquired using a dedicated combined PET/64-detector-CT (VCT-XT Discovery, GEHealthcare Technology, Chicago, Illinois), CT was performed (for attenuation correction) using 64 × 3.75mm detectors, a pitch of 1.5 and a 5mm collimation (120

kVp and 10mA in 0.8 s). Maintaining the patient position, a whole-body choline PET emission scan was performed and cover an area identical to that covered by CT [41].

4.3.4 Treatment Outcome Categorisation

Multidisciplinary review of at least two years of clinical, biochemical, radiological, and (where relevant) histopathological follow-up data was attended by at least one of each of the following clinicians: experienced uroradiologist (XX and/or YY), experienced nuclear medicine physician/radiologist (ZZ), treating clinical oncologist (AA and/or BB), and urological surgeon (as well as histopathologist with expertise in prostate cancer where relevant). Categorisation relied upon baseline and follow-up imaging together with biochemical and clinical data according to RECIST 1.1 and Prostate Cancer Working Group modifications thereof incorporating functional changes derived from e.g. BS and PET-CT [42], however ultimate categorisation was by consensus. This approach was preferred over sole reliance on traditional PCWG modified response assessments for several reasons e.g. some of these patients underwent cytoreductive local therapy as well as systemic therapy such that PSA change may not reflect distant lesional treatment response or otherwise, the fact that many positive lymph nodes were not technically measurable or even pathological by size criteria at baseline (>15mm and >9mm respectively; though deemed highly likely involved due to e.g. choline avidity) and the setting of radiorecurrence in this study. Categorisation was achieved in consensus on a per patient level into two groups- 50% of patients demonstrated partial response to therapy (PR; n=16/32) with mean age 70.5 years (range 55-82) whilst the other 50% of patients demonstrated non-response (either stability, mixed response or progressive disease) to therapy (NR; n=16/32) with mean age 72.3 years (range 60-81). Individual lesions were also categorised on the same basis (presented in results).

4.3.5 Image Analysis

Two experienced radiologists (with six and fifteen years’ experience) in consensus identified extra-prostatic disease sites on the WB-MRI in each patient with the benefit of multidisciplinary review and the baseline and post treatment conventional imaging (though blinded to subsequent outcome information). Volumetric T1 maps (Fig 1) were generated using dual flip angle (2.5 and 15 degrees) 2-point modified (m)Dixon 'in phase' images by a linear fitting algorithm [43] with MATLAB (v7.13) incorporating a separately acquired dual

TR B1 map [44]. Apparent diffusion coefficient (ADC) maps were calculated by mono-exponential curve fitting of mean signal intensities of all b-values from diffusion weighted sequences.

A third radiologist (9 years' experience; blinded for review) aware of the location of each lesion though unaware of the parametric maps or follow-up data referenced the anatomical images to volumetrically contour the pre-defined lesions on T1 weighted modified Dixon in phase images (flip angle 15) and diffusion weighted (b value 50ms) images using dedicated software (Jim 5.0, Xinapse systems, Thorpe, Waterville, UK) excluding areas of necrosis. These segmented volumes were then transferred to T1 and monoexponential ADC parametric maps respectively. Short axis diameter and maximal diameter for bone lesions were also recorded.

4.3.6 Statistical Analysis

All analyses compared the two outcome groups (i.e. PR versus NR) to assess for significant pretreatment differences between histographic T1 and ADC metrics and were performed on an all lesion basis performed separately for lymph node and bone disease (though these could co-exist and respond differently in the same patient). Additionally, differences in the 12 month T1 and ADC values and percentage change from baseline in cancerous lymph nodes and bone lesions were analysed. Lesional histographic T1 and ADC differences were analysed using a linear mixed model (Stata v13, StataCorp LLC, College Station, Texas, USA) to account for multiple samples per patient. Fitting was performed using response as the fixed factor and patient as the random factor. Means and 95% confidence intervals (CI) were estimated from the model for each patient response group. Where the data was right skewed the log measurement was used. Where the data was right skewed including negative values, a small constant was added to all values before the log transformation was applied with estimated means and 95% CIs back-transformed. Additionally, per lesion differences in the 12 month T1 and ADC values and percentage change from baseline were analysed using univariate accuracy for prediction of partial response determined by receiver operating characteristic (ROC) area under curve (AUC) analysis and 95% confidence intervals.

4.4 RESULTS

4.4.1 Patient cohort and lesion distribution

Summarised in Table 2, of the 32 men examined, on a per patient level, half demonstrated overall partial response to therapy (n=16/32) with the remaining 50% non-response (n=16/32; stable disease, progressive disease or mixed response). Twenty patients had lymph node only disease (62.5%), seven men had bone disease only (21.9%) and five had both lymph node and bone disease (15.6%).

Of the 25 men with lymph node disease, 36 diseased lymph nodes were contoured and followed up (median 1 node/patient; range 1-3 nodes). Of these 36 nodes, 58.3% (n=21/36) demonstrated partial response to therapy and the remaining 41.7% (n=15/36) showed non-response. The majority of involved nodes, 89% (n=32/36) were pelvic, with only four nodes outside of the pelvis; three retroperitoneal (two responding and one stable disease on follow up) and one subcarinal (demonstrating stable disease on follow-up).

Patient group	**Total Number Patients (%)**	**Mean age (yrs) (range)**	**Per patient disease type overall and by response**	**Per lesion type overall and response by type (either LN or bone)**
Total	32 (100%)	54.4 (55-82)	LN only- 62.5% (n=20/32) Bone disease only- 21.9% (n=7/32) Bone and LN disease- 15.6% (n=5/32)	LN- 36 nodes in 25 pts (pelvic 88.9% or 32/36) Bone- 27 lesions in 12 pts (Axial 55.6% or 15/27 Appendicular 44.6% or 12/27)
Partial Response	16 (50%)	70.5 (55-82)	LN only- 75% (n=12/16) Bone disease only- 18.8% (n=3/16) Bone and LN disease- 6.2% (n=1/16)	LN- 58.3% (21 LN of 36 total) Bone- 22.2% (6 bone lesions of 27 total)
Non-response	16 (50%)	72.3 (60-81)	LN only- 50% (n=8/16) Bone disease only- 25% (n=4/16) Bone and LN disease- 25% (n=4/16)	LN- 41.7% (15 LN of 36 total) Bone- 77.8% (21 bone lesions of 27 total)

Table 4.2 demonstrating patient demographics as well as per patient and per lesion disease distribution overall and by response.

Of the 12 men with bone disease, 27 bone deposits were contoured and followed up (median 2 bone deposits/patient; range 1-4 sites). Of these 27 bone lesions, only 22.2% (n=6/27) demonstrated partial response on follow up. Axial skeletal deposits accounted for 55.6% (n=15/27, two of which, or 13.3%, demonstrated partial response) of these with the remaining 44.4% being appendicular (n=12/27, four of which, or 33.3%, subsequently demonstrated partial response).

4.4.2 Lymph node disease

Histographic lymph node parameter changes between baseline and 12 months:
Table 3 summarises histographic values (median, skewness and kurtosis) and ranges of baseline and post 12 months T1 and ADC values and percentage changes post 12 months in both partial responding and non-responding lymph nodes. Percentage change in T1 value after 12 months of systemic therapy in partial responding lymph nodes was significantly lower at -24% versus +9.1% in non-responders (p<0.0001) and was the best predictor of response to treatment. ADC percentage change post 12 months of systemic therapy was significantly higher at +43% in partially responding lesions versus +11% in non-responding lesions (p=0.015). The mean post 12 month T1 value was slightly lower in partial responding lesions at 2096ms compared with 2142ms in lesions demonstrating no response, though this was not significant (p=0.84). The absolute ADC value post 12 months' systemic therapy was also higher at 1480 x10-6mm2/s in partial responding vs 1216 x10-6mm2/s in non-responding lesions which approached significance (p=0.066). These findings are demonstrated in Figure 1. All other absolute 12 month T1 and ADC histographic measures as well as percentage changes of these from baseline were not significantly different between the two groups. Of those which were significant, percentage change in lymph node T1 values after 12 months systemic therapy was the single best predictor of partial response in lymph nodes with ROC-AUC 0.92. This was followed by the percentage change in ADC at 12 months from baseline with ROC-AUC 0.75. These are shown in Figure 2.

Parameter	All Patients	Partial Response (21 nodes)	Non-response (15 nodes)	P value (linear mixed)
Baseline short axis diameter (mm)	10 (5mm to 25mm)	10 (6mm to 25mm)	11 (5mm to 25mm)	0.96
Median % change size	-15.9% (-75% to 87.5%)	-44% (-75% to -22%)	24% (-20% to 87.5%)	<0.0001
Baseline mean of median T1 (ms)	2426 (1056-3178)	2727 (1888-3098)	2003 (1056-3178)	0.012
Baseline mean of T1 kurtosis	0.28	0.21	0.30	0.67
Baseline mean of T1 skewness	0.55	0.66	0.52	0.13
Baseline mean of median ADC (x10-6mm2/s)	1092 (726-1874)	1081 (726-1500)	1108 (862-1874)	0.93
Baseline mean of ADC kurtosis	0.79	0.64	0.85	0.42
Baseline mean of ADC skewness	0.71	0.88	0.52	0.55
Post 12 months mean of median T1 (ms)	2115 (1087-3382)	2096 (1100-3191)	2142 (1087-3382)	0.84
Post 12 months mean of T1 kurtosis	0.77	0.71	0.59	0.66
Post 12 months mean of T1 skewness	0.45	0.54	0.27	0.12
Post 12 months mean of median ADC (x10-6mm2/s)	1370 (761-2534)	1480 (869-2534)	1216 (761-1808)	0.066
Post 12 months mean of ADC kurtosis	0.35	0.29	0.51	0.93
Post 12 months mean of ADC skewness	0.77	0.55	0.89	0.68
Post 12 months median % change T1	-9.7% (-62% to 64%)	-24% (-62% to 2.1%)	9.1% (-25% to 64%)	<0.0001
Post 12 months median % change ADC	30% (-23% to 189%)	43% (-23% to 189%)	11% (-12% to 55%)	0.015

Table 4.3 Summarising lymph node histographic T1 and ADC values at baseline, after 12 months of systemic therapy and percentage change compared to baseline (parentheses range) as well as short axis size change denoting response status. Panes coloured yellow

denote parameters demonstrating significant differences between 'partial' responding and non-responding lesions (p<0.05)

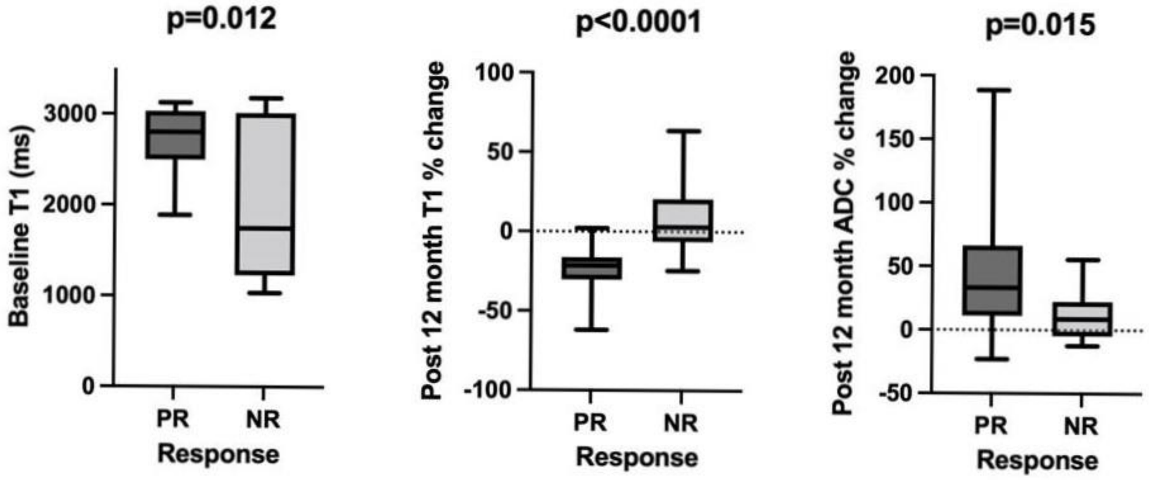


Figure 4-1 Box plots showing differences between responding and non-responding lymph nodes (a) Baseline T1 value, (b) Percentage change in T1 value post 12 months, (c) Percentage change ADC value post 12 months. Box indicates interquartile range, line median and whiskers most deviated range. Linear mixed method p-values shown.

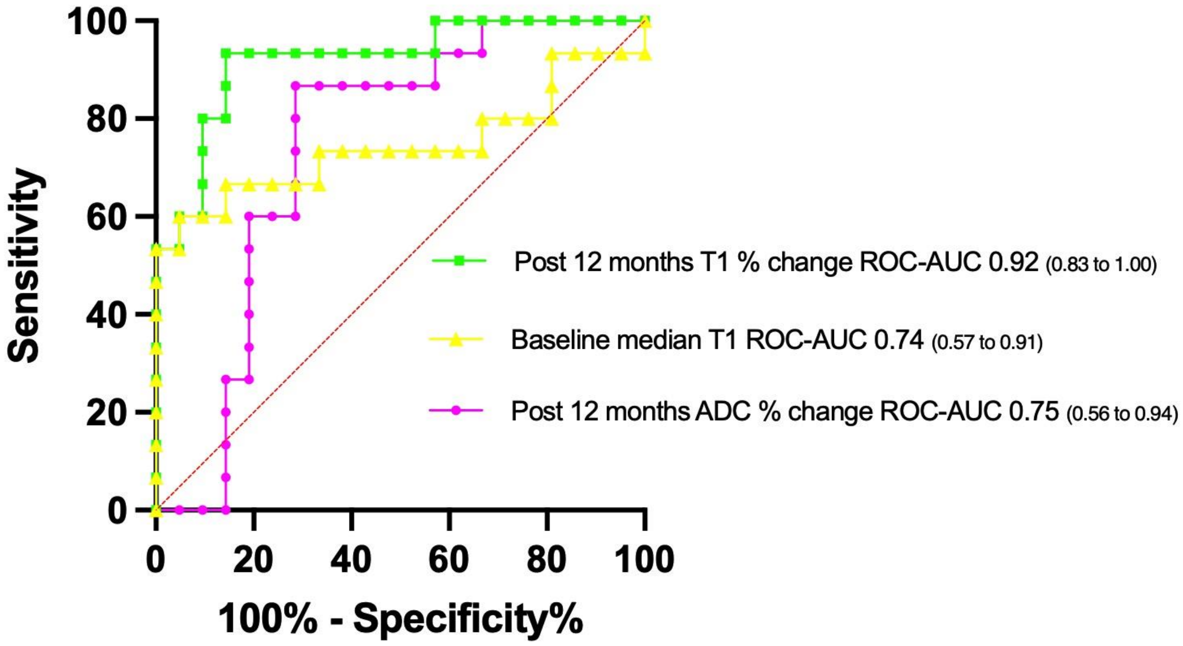


Figure 4-2 Receiver operating characteristics of lymph node baseline T1 as well as post 12 month percentage changes in T1 and ADC values for prediction of response with area under curve (AUC) values as shown (parentheses indicate 95% confidence intervals).

Histographic lymph node parameter baseline prediction:

In terms of prediction of response by comparison of baseline T1 and ADC values, partial responding lesions had significantly higher median T1 values with mean of 2727ms versus 2003ms in non-responding lesions (p=0.012). Baseline median ADC values were slightly lower in lesions going on to demonstrate partial response 1081x10-6mm2/s compared with 1181x10-6mm2/s in non-responding lesions but this was clearly not significant (p=0.93). Baseline lymph node T1 values were the best baseline predictor of subsequent partial response to systemic therapy in lymph nodes with ROC-AUC 0.74 (Figure 2). Other histographic parameters for both T1 and ADC at baseline were not significantly different and did not allow meaningful prediction of eventual response. Also of note, the baseline nodes were not significantly different in size and were below threshold for 'measurable' target lesion status by RECIST response criteria which mandates nodes must be greater than or equal to 15mm short axis diameter.

4.4.3 Bone disease

Histographic bone lesion parameter changes between baseline and 12 months:

Table 4 summarises histographic values (median, skewness and kurtosis) and ranges of baseline and post 12 months T1 and ADC values and percentage changes post 12 months in both partial responding and non-responding bone deposits. Percentage change in T1 value after 12 months of systemic therapy in partial responding bone lesions was significantly lower at -19% versus -6.4% in non-responders (p=0.013) and was the only significant predictor of response to systemic treatment (Figure 3) though this parametric change also outperformed size decrease with respect to categorising responding and non-responding bone lesions (p=0.035). ADC percentage change post 12 months of systemic therapy was marginally higher at +14% in partially responding bone lesions versus +13% in non-responding lesions but this was not significant (p=0.93). The mean post 12 month T1 value was slightly lower in partial responding lesions at 1671ms compared with 1777ms in lesions demonstrating no response, though this was not significant (p=0.58). The absolute ADC value post 12 months' systemic therapy was slightly lower at 1061x10-6mm2/s in partial responding vs 1141 x10-6mm2/s in non-responding lesions again this was not significant (p=0.28). All other absolute 12 month T1 and ADC histographic measures, as well as percentage changes of these from baseline, were not significantly different between the two groups. Percentage change in bone lesion T1 values after 12 months systemic therapy was the single predictor of partial response in bone disease lesions with ROC-AUC 0.83 and is shown in Figure 4.

Parameter	**All Patients**	**Partial Response (6 bone lesions)**	**Non-response (21 bone lesions)**	**P value (linear mixed)**
Baseline short axis diameter (mm)	18 (7mm to 50mm)	23 (10 to 50mm)	17 (7mm to 44mm)	0.14
Median % change size	18% (-40% to 178%)	-11% (-40% to 0%)	26% (-11% to 178%)	0.035
Baseline mean of median T1 (ms)	1948 (1350-2750)	2072 (1645-2750)	1913 (1350-2466)	0.58
Baseline mean of T1 kurtosis	0.87	0.96	0.55	0.56
Baseline mean of T1 skewness	0.12	0.05	0.23	0.74
Baseline mean of median ADC (x10-6mm2/s)	998 (750-1359)	936 (764-1085)	1016 (750-1359)	0.42
Baseline mean of ADC kurtosis	0.56	0.42	0.71	0.11
Baseline mean of ADC skewness	0.67	0.88	0.57	0.23
Post 12 months mean of median T1 (ms)	1754 (1136-2250)	1671 (1239-2150)	1777 (1136-2250)	0.58
Post 12 months mean of T1 kurtosis	0.24	0.15	0.38	0.09
Post 12 months mean of T1 skewness	0.56	0.74	0.49	0.56
Post 12 months mean of median ADC (x10-6mm2/s)	1123 (913-1501)	1061 (935-1217)	1141 (913-1501)	0.28
Post 12 months mean of ADC kurtosis	0.66	0.79	0.45	0.33
Post 12 months mean of ADC skewness	0.31	0.22	0.46	0.55
Post 12 months median % change T1	-9.3% (-45% to 13%)	-19% (-31% to -7.2%)	-6.4% (-45% to 13%)	0.013
Post 12 months median % change ADC	14%	14.2%	13.7%	0.93

	(-8.3% to 41%)	(-1.1% to 35%)	(-8.3% to 41%)	

Table 4.4 Summarising bone deposit histographic T1 and ADC values at baseline, after 12 months of systemic therapy and percentage change compared to baseline (parentheses range) as well as largest diameter size change. Panes coloured yellow denote parameters demonstrating significant differences between 'partial' responding and non-responding lesions (p<0.05)

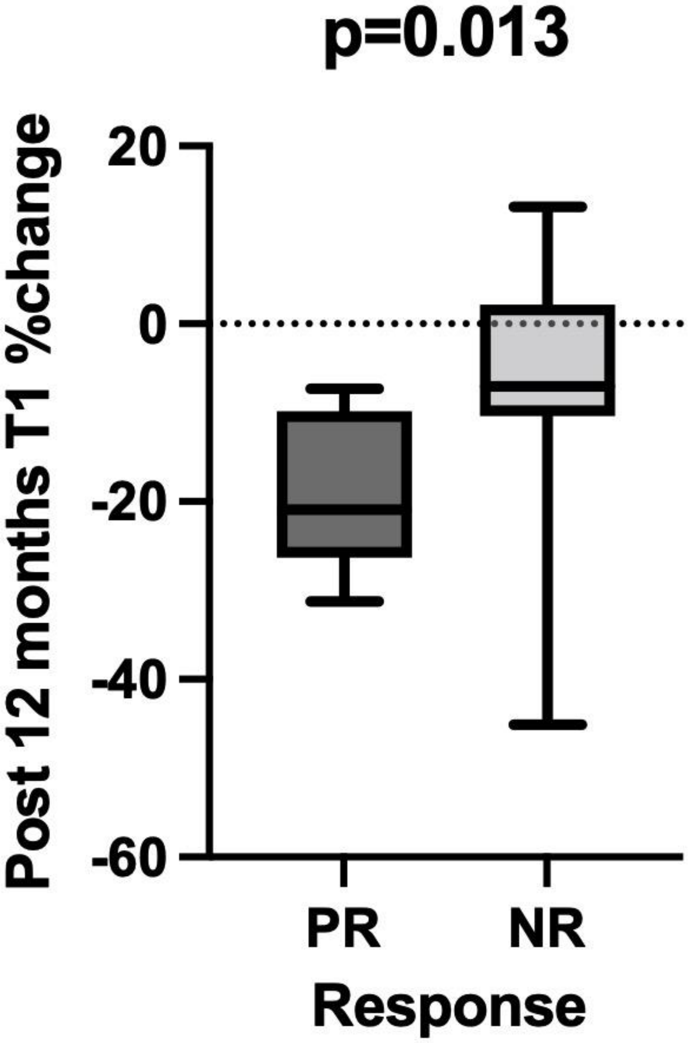


Figure 4-3 Box plots showing differences between responding and non-responding bone lesions percentage change in T1 value post 12 months. Box indicates interquartile range, line median and whiskers most deviated range. Linear mixed method p-values shown.

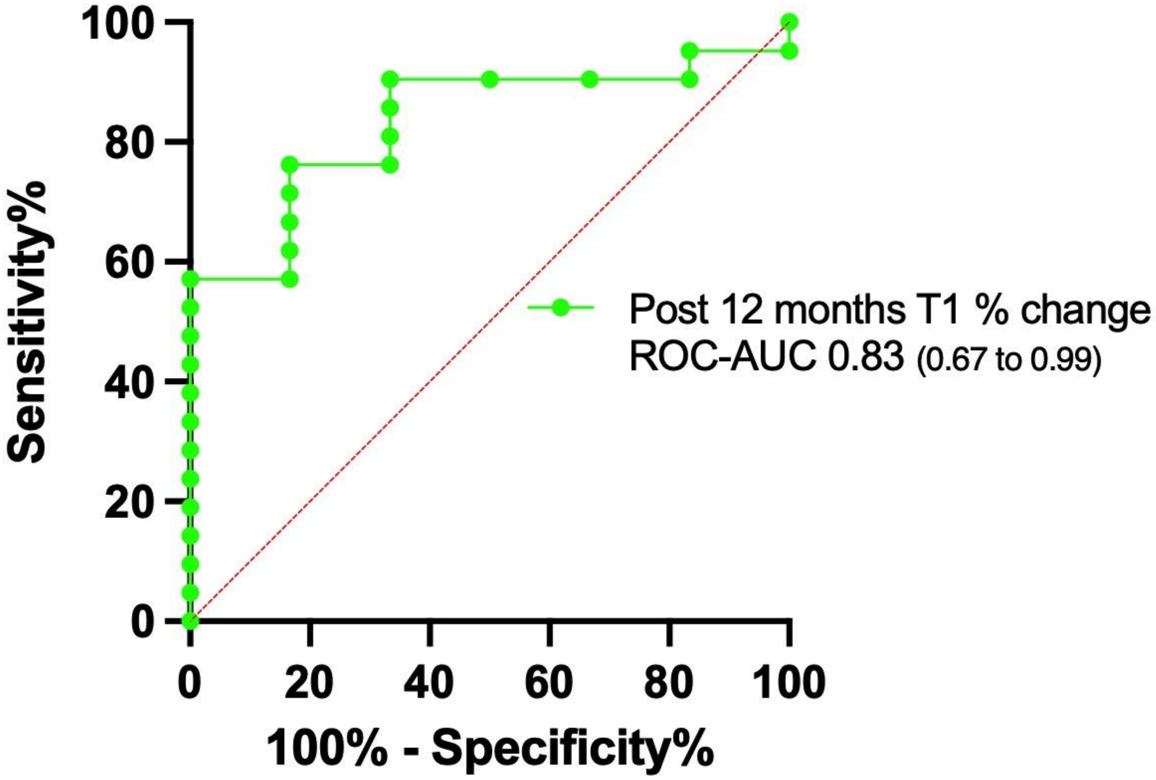


Figure 4-4 Receiver operating characteristics of the post 12 months percentage change of bone lesion T1 for prediction of response with area under curve (AUC) values as shown (parentheses indicate 95% confidence intervals)

Histographic bone lesion parameter baseline prediction:

Bone disease individual lesion sizes were not significantly different between the two response groups at baseline with mean partial responding bone lesions averaging 23mm compared with 17mm in the non-responding group (p=0.14). Partial responding bone lesions had slightly higher median T1 values with mean of 2072ms versus 1913ms in non-responding lesions but this was not a significant difference (p=0.58). Baseline median ADC values were slightly lower in bone lesions going on to demonstrate partial response 936x10-6mm2/s compared with 1016x10-6mm2/s in non-responding lesions but this was also not significant (p=0.42). Other histographic parameters for both T1 and ADC at baseline were not significantly different and therefore we did not demonstrate any baseline MR parameters that facilitated meaningful prediction of eventual response in bone lesions.

4.5 DISCUSSION

This study has demonstrated that for both pathological lymph nodes and bone deposits involved in spread of radiorecurrent prostate cancer, T1 value measurement and change derived from clinically feasible whole body MRI scans can be used to discriminate response to systemic therapy. To our knowledge this is the first study evaluating the utility of in vivo T1 mapping in predicting and assessing response to systemic therapy in metastatic prostate cancer.

With respect to cancerous mPCa lymph nodes, this study found percentage change in T1 values between baseline and post 12 months of systemic therapy was the single best MR biomarker for differentiating partially responding and non-responding lymph nodes with a significant decrease of around 20% in T1 values between the baseline and follow-up 12 month scan in responding nodes. Of note whilst these nodes were also significantly decreased in size, their mean baseline short axis diameters were only on the borderline of pathology and most were not ‘measurable’ under conventional response criteria such that they would not be formally be defined as responding nodes under these assessments. Baseline nodal T1 value was also the only baseline predictor of subsequent response to therapy in these involved nodes, with responding lesions demonstrating a significantly higher baseline T1 value compared with those which would go on to demonstrate non-response. The percentage change in T1 at 12 months outperformed percentage change in ADC at 12 months (a more widely studied and characterised imaging biomarker for response) in discriminating response. The finding that those lymph nodes which demonstrate partial response to chemotherapy by the end of treatment show decreases in T1 of this order supports preclinical murine tumour model work [45]. This effect has been postulated to reflect a decrease in remaining viable/proliferating tumour cells due to cell destruction or inhibition and/or consequent release of proteins and/or metals (causing T1 relaxation) [46]. The demonstration of higher baseline T1 being correlated with subsequent response to systemic treatment is novel and will also require further work to understand.

With respect to mPCa bone deposits, percentage change in T1 values between baseline and 12 month follow-up scan was the only tested discriminator of partially responding and non-responding lesions. We can postulate that the mechanisms underpinning this finding are commensurate with those preclinical indications for soft tissue tumours. This is particularly important when considering the difficulties bone lesions pose for imaging response assessment described previously as is perhaps demonstrated by the fact that percentage T1 change actually outperformed size change in differentiating response. In contrast with lymph nodes however, actual baseline bone lesion T1 values were not sufficiently different between those which would go on to partially respond compared with non-responders to allow meaningful prediction of response. ADC assessment either at baseline or in terms of change between baseline and 12 months did not allow prediction or discrimination of responding versus non-responding bone lesions. This is perhaps not unsurprising given that although there are reports of this metric being useful in assessing response of bone lesions, there are a similar number reporting to the contrary.

To our knowledge, this is the first study to globally assess tumour response on a whole body basis in humans with the aim of assessing baseline and change in metastatic prostate cancer T1 values to differentiate and potentially predict response to systemic therapy. In line with discussed biomarker validation theory, this contributes to the qualification process of T1 tumour value as a biomarker for response assessment. Prior work (contributing to analytical validation) has demonstrated that the method used in our study is adequately reproducible and repeatable across a range of T1 values to detect changes in T1 which might predict tumour response to systemic treatment before conventional size assessment based on these findings [38].

This study has several limitations including relatively small sample size; particularly e.g. partially responding bone lesions (however we were keen to examine these separately due to their potentially different biological setting as well as the fact they are often problematic to assess in conventional imaging follow-up). Furthermore we accounted for multiple lesions per patient in our analysis to compensate for multiple samples. The systemic therapy in these patients was not homogenous and could also confound the observed quantitative metrics however as noted in the introduction is not standardised and is often dependent on a range of factors. Independent verification of disease presence was not confirmed on a lesional basis e.g. with histology (though clearly this was not ethically or logistically possible given clinical guidelines). In addition for the purposes of this analysis we opted for binary categorisation of response as opposed to stratification of response as described by traditional tumour response criteria.

In conclusion, this study demonstrates that there are significant differences in T1 tumoral value changes between responding versus non-responding mPCa disease sites which may aid response assessment and in cancerous lymph nodes baseline T1 measurement could feasibly be exploited in risk stratification prior to systemic therapy given that traditional qualitative radiological classifiers of nodal pathology and DWI parameters at baseline were poor discriminators of subsequent response. Future work is, however, required to understand the mechanisms underpinning these observed differences in metastatic prostate cancer.

4.6 REFERENCES

1. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer#heading-Zero- Accessed 30th of June 2021.

2. Naik M, Marta GN, Abdel-Wahab M. Reirradiation of locally recurrent prostate cancer after primary radiotherapy. J Radiat Oncol. 2015;4(2):149–56.

3. Shilkrut M, McLaughlin PW, Merrick GS, Vainshtein JM, Feng FY, Hamstra DA. Interval to biochemical failure predicts clinical outcomes in patients with high-risk prostate cancer treated by combined-modality radiation therapy. Int J Radiat Oncol Biol Phys. 2013;86(4):721–8.

4. Roach M 3rd, Hanks G, Thames H Jr, Schellhammer P, Shipley WU, Sokol GH, Sandler H. Defining biochemical failure following radiotherapy with or without hormonal therapy in men with clinically localized prostate cancer: recommendations of the RTOG-ASTRO Phoenix consensus conference. Int J Radiat Oncol Biol Phys. 2006;65(4):965–74.

5. Turkbey B, Brown AM, Sankineni S, Wood BJ, Pinto PA, Choyke PL. Multiparametric prostate magnetic resonance imaging in the evaluation of prostate Cancer. CA Cancer J Clin. 2016;66(4):326–36.

6. Testa C, Pultrone C, Manners DN, Schiavina R, Lodi R. Metabolic imaging in prostate Cancer: where we are. Front Oncol. 2016;6:225.

7. Kanthabalan A, Shah T, Arya M, Punwani S, Bomanji J, Haroon A, Illing RO, Latifoltojar A, Freeman A, Jameson C, van der Meulen J, Charman S, Emberton M, Ahmed HU. The FORECAST study — focal recurrent assessment and salvage treatment for radiorecurrent prostate cancer. Contemp Clin Trials. 2015;44:175–86.

8. Edelstein RA, Zietman AL, De Las MA, Krane RJ, Babayan RK, Dallow KC, Moreland RB. Implications of prostate micro-metastases in pelvic lymph nodes: an archival tissue study. Urology. 1996;47(3):370–5.

9. Schiavina R, Ceci F, Borghesi M, Brunocilla E, Vagnoni V, Gacci M, Castellucci P, Nanni C, Martorana G, Fanti S. The dilemma of localizing disease relapse after radical treatment for prostate Cancer: which is the value of the actual imaging techniques? Curr Radiopharm. 2013;6(2):92–5.

10. Briganti A, Abdollah F, Nini A, Suardi N, Gallina A, Capitanio U, Bianchi M, Tutolo M, Passoni NM, Salonia A, Colombo R, Freschi M, Rigatti P, Montorsi F. Performance

characteristics of computed tomography in detecting lymph node metastases in contemporary patients with prostate cancer treated with extended pelvic lymph node dissection. Eur Urol. 2012;61(6):1132–8.

11. Talab SS, Preston MA, Elmi A, Tabatabaei S. Prostate cancer imaging: what the urologist wants to know. Radiol Clin N Am. 2012;50(6):1015–41.

12. Sartor O, Eisenberger M, Kattan MW, Tombal B, Lecouvet F. Unmet needs in the prediction and detection of metastases in prostate cancer. Oncologist. 2013;18(5):549–57.

13. Suh CH, Shinagare AB, Westenfield AM, Ramaiya NH, Van den Abbeele AB, Kim KW. Yield of bone scintigraphy for the detection of metastatic disease in treatment-naïve prostate cancer: a systemic review and meta-analysis. Clin Radiol. 2018;73(2):158–67.

14. Rinnab L, Mottaghy FM, Simon J, Volkmer BG, De Petriconi R, Hautman RE, et al. 11C choline PET/CT for targeted salvage lymph node dissection in patients with biochemical recurrence after primary curative therapy for prostate cancer: preliminary results of a prospective study. Urol Int. 2008;81:191–7.

15. Abuzallouf S, Dayes I, Lukka H. Baseline staging of newly diagnosed prostate cancer: a summary of the literature. J Urol. 2004;171:2122–7.

16. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009 45(2): 228-47.

17. A.R. Padhani. Functional MRI for anticancer therapy assessment. European Journal of Cancer 2002 38: 2116-2127.

18. Stacchiotti S, Collini P, Messina A, Morosi C, Barisella M, Bertulli R, Piovesan C, Dileo P, Torri V, Gronchi A, Casali PG. High-grade soft-tissue sarcomas: tumor response assessment--pilot study to assess the correlation between radiologic and pathologic response by using RECIST and Choi criteria. Radiology 2009 251(2):447-56.

19. Choi H. Response evaluation of gastrointestinal stromal tumors. Oncologist 2008 13(Suppl 2):4-7.

20. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. J Nucl Med. 2009 May;50 Suppl 1(Suppl 1):122S-50S. doi: 10.2967/jnumed.108.057307. PMID: 19403881; PMCID: PMC2755245.

21. Scher H.I., Morris M.J., Stadler W.M. Trial design and objectives for castration-resistant prostate cancer: Updated recommendations from the Prostate Cancer Clinical Trials Working Group 3. J Clin Oncol. 2016:1–38.

22. Padhani AR, Lecouvet FE, Tunariu N, Koh DM, De Keyzer F, Collins DJ, Sala E, Fanti S, Vargas HA, Petralia G, et al. Eur Urol Focus. 2017 Apr; 3(2-3):223-239.

23. Lecouvet F.E., Talbot J.N., Messiou C. Monitoring the response of bone metastases to treatment with Magnetic Resonance Imaging and nuclear medicine techniques: A review and position statement by the European Organisation for Research and Treatment of Cancer imaging group. Eur J Cancer. 2014;50:2519–2531.

24. Thoeny, H.C. and Ross, B.D. (2010), Predicting and monitoring cancer treatment response with diffusion-weighted MRI. J. Magn. Reson. Imaging, 32: 2-16.

25. Messiou, C., Collins, D.J., Giles, S. et al. Assessing response in bone metastases in prostate cancer with diffusion weighted MRI. Eur Radiol 21, 2169–2177 (2011).

26. Graham TJ, Box G, Tunariu N, Crespo M, Spinks TJ, MirandaS, et al. Preclinical evaluation of imaging biomarkers forprostate cancer bone metastasis and response to cabozantinib.J Natl Cancer Inst 2014;106:dju033.

27. Lee KC, Sud S, Meyer CR, et al. An imaging biomarker of earlytreatment response in prostate cancer that has metastasized tothe bone. Cancer Res 2007;67:3524–8.

28. Rozel S, Galban CJ, Nicolay K, et al. Synergy between anti-CCL2 and docetaxel as determined by DW-MRI in a metastaticbone cancer model. J Cell Biochem 2009;107:58–64.

29. Bauerle T, Bartling S, Berger M, et al. Imaging anti-angiogenictreatment response with DCE-VCT, DCE-MRI and DWI in ananimal model of breast cancer bone metastasis. Eur J Radiol2010;73:280–7.

30. Wieder H., Beer A.J. 11C-choline PET/CT and whole-body MRI including diffusion-weighted imaging for patients with recurrent prostate cancer. Oncotarget. 2017;8:66516–66527.

31. Jakobsen I, Kaalhus O, Lyng H, Rofstad EK. Detection of necrosis in human tumour xenografts by proton magnetic resonance imaging. Br J Cancer 1995 71: 456-61.

32. Gambarota G, Veltien A, van Laarhoven H, et al. Measurements of T1 and T2 relaxation times of colon cancer metastases in rat liver at 7 T. Magn Res Mat Phys Biol Med 2004 17: 281-7.

33. Thomsen C, Sørensen PG, Karle H, Christoffersen P, Henriksen O. Prolonged bone marrow T1-relaxation in acute leukaemia. In vivo tissue characterization by magnetic resonance imaging. Magn Reson Imag 1987 5: 251-7.

34. Jensen KE, Sørensen PG, Thomsen C, et al. Magnetic resonance imaging of the bone marrow in patients with acute leukemia during and after chemotherapy. Acta Radiologica 1990 31:361-9.

35. Santoni R, Bucciolini M, Chiostrini C, Cionini L, Renzi R. Quantitative magnetic resonance imaging in cervical carcinoma: a report on 30 cases. Br J Radiol 1991 64: 498-504.

36. Smith SR, Roberts N, Percy DF, Edwards RH. Detection of bone marrow abnormalities in patients with Hodgkin's disease by T1 mapping of MR images of lumbar vertebral bone marrow. Br J Cancer 1992 65: 246-51.

37. McSheehy PM1, Weidensteiner C, Cannet C, Ferretti S, Laurent D, Ruetz S, Stumm M, Allegrini PR. Quantified tumor t1 is a generic early-response imaging biomarker for chemotherapy reflecting cell viability. Clin Cancer Res. 2010 16(1): 212-25.

38. Barnes A et al. ISMRM 2014 #3241.

39. Adeleke S, Latifoltojar A, Sidhu H, et al. Localising occult prostate cancer metastasis with advanced imaging techniques (LOCATE trial): a prospective cohort, observational diagnostic accuracy trial investigating whole-body magnetic resonance imaging in radio-recurrent prostate cancer. BMC Med Imaging. 2019;19(1):90. Published 2019 Nov 15. doi:10.1186/s12880-019-0380-y

40. Kirkham APS, Haslam P, Keanie JY, McCafferty I, Padhani AR, Punwani S, Richenberg J, Rottenberg G, Sohaib A, Thompson P, Turnbull LW, Sahdev A, Clements R, Carey BM, Allen C. Prostate MRI: Who, when, and how? Report from a UK consensus meeting. Clin Radiol. 2013;68(10):1016–23.

41. Johnston EW, Latifoltojar A, Sidhu HS, Ramachandran N, Sokolska M, Bainbridge A, et al. Multiparametric whole-body 3.0-T MRI in newly diagnosed intermediate- and high-risk prostate cancer: diagnostic accuracy and interobserver agreement for nodal and metastatic staging. Eur Radiol. 2019;29(6):3159–69.

42. Scher HI, Morris MJ, Stadler WM, et al. Trial Design and Objectives for Castration-Resistant Prostate Cancer: Updated Recommendations From the Prostate Cancer Clinical Trials Working Group 3. J Clin Oncol. 2016;34(12):1402-1418. doi:10.1200/JCO.2015.64.2702

43. Wang D, Shi L, Wang YX, Yuan J, et al. Concatenated and parallel optimization for the estimation of T1 map in FLASH MRI with multiple flip angles. Magn Reson Med 2010 63(5): 1431-6.

44. Yarnykh VL. Optimal radiofrequency and gradient spoiling for improved accuracy of T1 and B1 measurements using fast steady-state techniques. Magn Reson Med 2010 63(6): 1610-26.

45. McSheehy PM, Weidensteiner C, Cannet C, Ferretti S, Laurent D, Ruetz S, Stumm M, Allegrini PR. Quantified tumor t1 is a generic early-response imaging biomarker for chemotherapy reflecting cell viability. Clin Cancer Res. 2010 16(1): 212-25.

46. Weidensteiner C, Allegrini PR, Sticker-Jantscheff M, et al. Tumour T1 changes in vivo are highly predictive of response to chemotherapy and reflect the number of viable tumour cells--a preclinical MR study in mice. BMC Cancer. 2014 Feb 14;14:88.

SECTION C: CLINICAL EVALUATION OF ESTABLISHED MR BIOMARKERS IN HEAD AND NECK CANCER

Chapter 5 Clinical Validation (Qualification) of Existing Quantitative MR Biomarker in Tumour Response Prediction- BOLD imaging in Head and Neck cancer

CONTRIBUTION:

Primary author responsible for data analysis, drafting, revision, final approval and accuracy/integrity.

5.1 RUNNING TITLE:

Oxygen-enhanced R2* weighted MRI and diffusion weighted MRI of head and neck squamous cell cancer lymph nodes in prediction of 2-year outcome following chemoradiotherapy.

5.2 INTRODUCTION

Head and neck squamous cell carcinoma (HNSCC) is common, occurring with an estimated incidence of greater than 900,000 cases per year, with approximately 60% of patients presenting with locally advanced non-metastatic disease, and responsible for more than 350,000 deaths per year [1]. The presence of cervical lymph node metastases is an important adverse prognostic factor [2]. Over the past decade, organ preservation strategies employing chemoradiotherapy (CRT) have become an accepted alternative to surgery [3] with 5-year overall survival rates of 30 to 50% [4] though associated with significant toxicity [5]. Locoregional failure is the predominant pattern of post-CRT relapse occurring in 25-30% [6] and pre-treatment identification of resistant tumour could facilitate treatment modification and/or tailored monitoring. Current prognostication strategies include accurate staging, clinical features [7] and tumour biology markers (e.g. preceding oncogenic human papillomavirus infection [8]). MRI is established in diagnosis, staging and radiotherapy planning though ability to discriminate involved lymph remains an area for development [9]. Short axis diameter and use of morphological criteria (e.g. border regularity) add value in detection of nodal metastatic disease [10] and has not been applied as a prognostic tool.

Tumour hypoxia, defined as a mismatch between cellular oxygen demand and supply, triggers a cellular response in individual cells after only several hours [11] and often occurs when the distance from a cell to the nearest vasculature is too large for adequate cellular oxygenation [12]. Successful radiotherapy oxygen for free radical formation induce DNA cell death [13]. Invasive direct measurement of tumour oxygenation using invasive Eppendorf computerized histographic oxygen electrode system [14] has confirmed HNSCC tumours are hypoxic relative to normal tissue [15] and that increasing tumour hypoxia is associated with poorer post-CRT outcomes [16]. However, invasive oxygenation assessment is not practical and potentially suffers for sampling error as the needles samples only the very local environment at its tip and ignores the heterogeneity that exists across a tumour which may account for the observation that directly measured tumor oxygenation is not necessarily an independent predictor of treatment outcome [17].

Nonetheless, other studies have suggested marked tumoral heterogeneity in oxygen distribution with directly measured hypoxia not necessarily being an independent predictor of treatment outcomes [17]. Non-invasive MRI measurement of transverse relaxation time (T2*) has been proposed as a potential marker of tissue oxygenation status [18, 19] offering measurement of both hypoxic extent and spatial distribution [20-22]. Elevated levels of deoxyhaemoglobin shorten T2*, thereby generating image contrast and giving an indication of tissue hypoxia [23], (though quantitative application has yet to be established [24]) with the potential to detect clinically relevant changes in tumour oxygenation reported in HNSCC [25]. The inverse of the transverse relaxation time, the transverse relaxation rate (1/T2*=R2*), is often used as a more convenient marker.

However, R2* is also effected by static magnetic field gradients that are found at tissue-tissue and tissue-air boundaries. The problem is compounded in the head and neck region due to the complex shape and anatomy to be imaged. Hence, R2* differences between measurements performed on air and then breathing hyperoxic gas, could theoretically reflect differences in oxygenation status whilst removing confounding static susceptibility gradients effects. The magnitude of the difference has previously been shown to reflect the tumor hypoxic fraction as determined by histologic pimonidazole labelling [26] though are influenced by tumour biology/type [27].

Diffusion-weighted imaging (DWI) characterises tissue based on movement of water within a volume, which depends on microstructural features, such as cellular density, and may be quantified by the apparent diffusion coefficient (ADC). Adverse biological tumour characteristics may decrease diffusion restriction of water molecules in HNSCC, and lead to higher pretreatment ADC. Thus far, however, evidence for the use of pretreatment tumoral ADC in prediction of sustained response to therapy in HNSCC is mixed [28].

Within this study we evaluate baseline lymph node (LN) T2* relaxation times on breathing air and 100%-oxygen to predict chemoradiotherapeutic locoregional response at 2 years in head and neck squamous cell cancer (HNSCC) and compare against nodal ADC measurement, short axis diameter and established qualitative radiological features [10].

5.3MATERIALS AND METHODS

Institutional review board approval was granted and patients' informed consent (written and verbal) was obtained prior to entry into this prospective study (R&D No: 09/0327).

5.3.1 Patient cohort

Patients aged over 17 years; with histologically confirmed HNSCC staged N2/N3 (AJCC TNM Cancer Staging Manual 7th edition [29]) and due for primary chemoradiotherapy between 10th February 2010 to 31st July 2017 inclusive were eligible (n=105). Patients who declined trial entry, received incomplete CRT, were unable to undergo MRI, were pregnant or with prior malignancy were excluded (n=33). Furthermore, those with incomplete two year clinical, imaging and histopathological follow-up were excluded for subsequent analysis (n=14), as were those with inadequate/incomplete baseline MR datasets (n=4). In total, 54 patients were accrued (mean age 57.1 years; range 25-79 years) of whom forty-three were male (mean age 57.7 years; range 33-79 years) and eleven were female (mean age 54.8 years; range 25-74 years). Each patient received CRT according to departmental protocol involving intensity modulated radiotherapy (IMRT) with minimal dose of 60-70Gy with concurrent cisplatin (or cetuximab where platin contra-indicated) over 6-7 weeks.

5.3.2 Multi-parametric magnetic resonance imaging

Subjects underwent multiparametric MRI using a 1.5T static magnet (Avanto, Siemens, Erlangen, Germany) with carotid coils in supine position with median interval to CRT of 22 days. Anatomic axial T2 weighted sequences covered base of skull to upper thorax. Diffusion weighted imaging was then performed using six b-values (0, 50, 100, 300, 600, and 1000 s/mm^2). Axial T2* weighted imaging was undertaken using a multiple gradient echo sequence (echo times 12, 24, 36, 48ms). T2* images were initially acquired with patient breathing room air and then repeated after 100%-oxygen inhalation at 15L/min for four minutes via non-rebreather facemask prior to scanning which continued for the duration of these sequences. Full MRI parameters are given in Table 1.

	T2w TSE	**T2* GRE**	**STIR- EPI DWI**
Orientation	Axial	Axial	Axial

Repetition time (ms)	6670	1450	9400
Echo time (ms)	93	12, 24, 36, 48	93
Flip angle (degrees)	120	25	90
Bandwidth (Hz/Px)	250	140	1502
Field of view (mm)	180 x 180	195x250	206 x 206
Acquired Matrix	256x256	512x400	128x128
Slice thickness (mm)	3	5	4
Slice gap (mm)	0.3	2.5	0.4
Averages	1	1	4
Phase encoding direction	Anteroposterior	Anteroposterior	Anteroposterior
Fat suppression	No	No	STIR
Base matrix	256	256x100	128
Number of acquisitions	1	1	1
b-values (s.mm^{-2})	n/a	n/a	0, 50, 100, 300, 600, 1000
Total acquisition time (min)	2m54s	2mins53s	2m31s

Table 5.1 Multi-parametric MRI sequence parameters used for study. TSE- turbo spin echo, DWI- diffusion weighted imaging, STIR- short tau inversion recovery, GRE- gradient recall echo.

5.3.3 Image Analysis

Multiparametric MR images were evaluated by two experienced head and neck radiologists (TB and SM; 16 and 9 years' experience respectively) in consensus to identify pathological involved head and neck nodes with reference to all prior imaging and cytology/histology though blinded to 2-year outcome. The short axis diameters of each involved node, the largest short axis diameter node for each patient and qualitative morphological parameters for each lymph node as binary descriptors of nodal contour (ovoid/round), margins (smooth/irregular), enhancement pattern (diffuse/heterogenous), and necrosis (present/absent) were recorded. T2* maps for both ‘air’ and ‘100%-oxygen’ were produced from the multiple gradient echo sequences by previously described numerical fitting algorithm [30] using Matlab (version 7.13, MathWorks Inc., Natick, MA, USA). ADC maps were produced using monoexponential least-squares fitting incorporating all b-values.

A third radiologist (HS with 7 years experience; blinded for review) aware of the location of each pathological node though unaware of the parametric maps or follow-up data referenced the anatomical images to volumetrically contour the pre-defined nodes on both ‘air’ and ‘100%-oxygen’ 12ms echo T2* weighted images using dedicated software (Jim 5.0, Xinapse systems, Thorpe, Waterville, UK) excluding areas of necrosis. These segmented volumes were then transferred to the air and 100%-oxygen T2* parametric maps respectively (Figure 1). Secondly, the radiologist volumetrically contoured the identified nodes on b300 diffusion images, and these segmented volumes were transferred to the monoexponential ADC parametric maps for each patient.

In total 170 nodal volumes of interest (median 3/patient; range 1-13 nodes) were contoured for each of the ‘air’ and ‘100%-oxygen’ T2* maps and the ADC maps. Median, skewness, kurtosis were derived for each nodal volume from both of the T2* and the ADC maps.

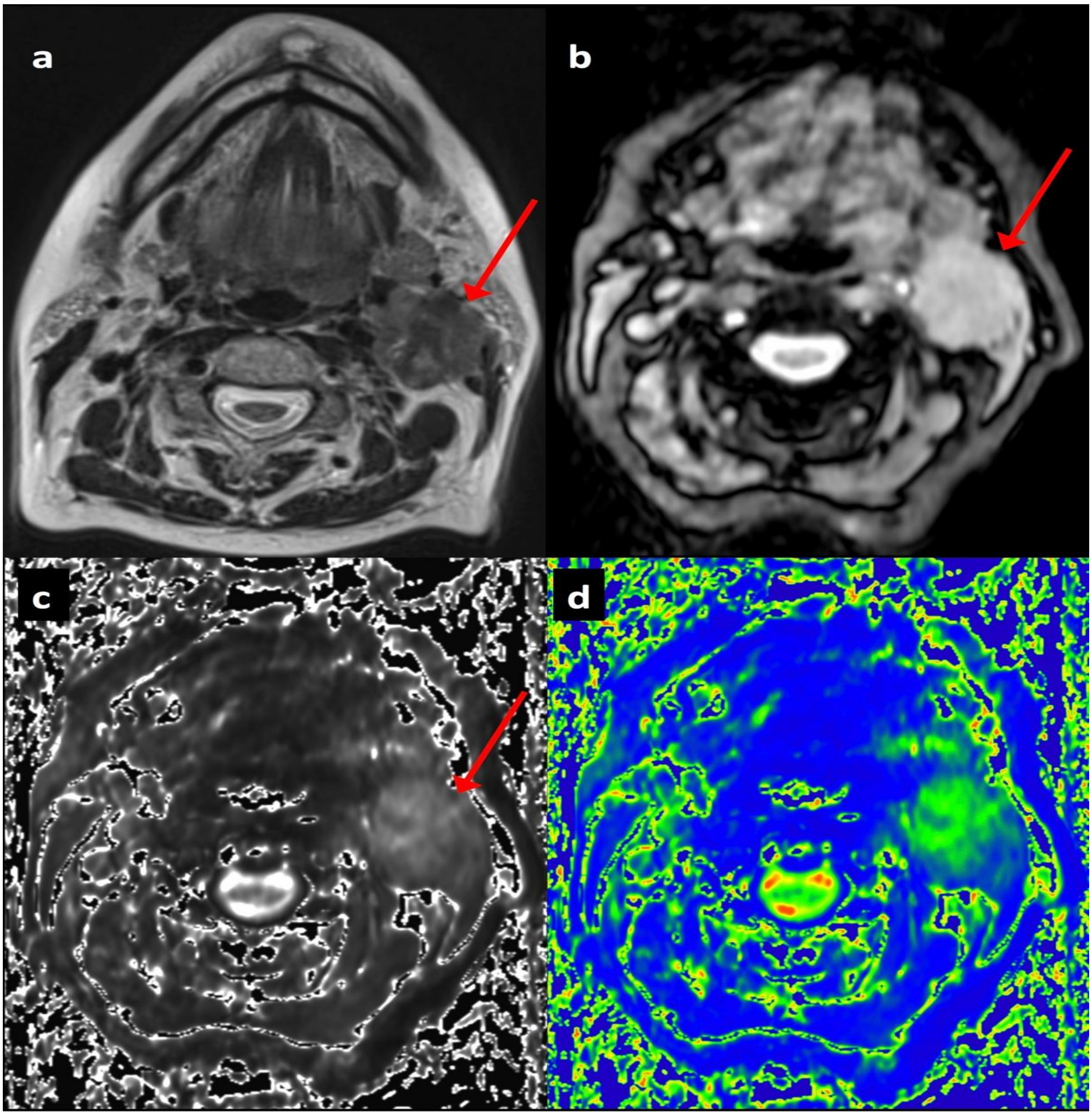


Figure 5-1 Seventy four year old female patient with T2 N2b left tongue base squamous cell carcinoma showing left upper deep cervical nodal mass: (a) T2-TSE axial, (b) T2* gradient echo (12ms) on air, (c) T2* parametric map on air, (d) T2* parametric map on air with colour look up table applied

5.3.4 Treatment Outcome Categorisation

Multidisciplinary review of at least two years of clinical, radiological, and histopathological follow-up data was attended by at least one of each of the following clinicians: experienced head and neck radiologist (XX and/or YY), treating clinical oncologist (AA and/or BB), histopathologist with expertise in head and neck oncology and ENT surgeon. Categorisation was achieved by consensus into two groups- 59% of patients demonstrated sustained post-CRT complete local response (CR; n=32/54; total 104/170 nodes) with mean age 60.5 years (range 39 to 79 years) whilst 41% of patients developed local nodal disease relapse (RD; n=22/56; total 66/170 nodes) with mean age 57.2 years (range 25 to 75 years).

5.3.5 Statistical Analysis

All analyses compared the two outcome groups (i.e. CR versus RD) to assess for significant pretreatment differences between short axis diameter, qualitative descriptors and histographic R2* and ADC metrics and were performed on both a largest node per patient (LNPP) and all nodes (AN) bases.

LNPP analysis compared short axis diameters using two-tailed Mann Whitney U test and qualitative classifiers using Fishers exact test. LNPP pairwise parameter differences between air and 100%-oxygen R2* datasets were calculated using Wilcoxon signed rank test. Histographic LNPP R2* and ADC parameters differences were compared between outcome groups using Mann Whitney U test. Statistical significance was assigned at p values below 0.05 performed using SPSS statistics for Windows (version 16; IBM, Armonk, NY, USA).

For the AN analysis, baseline morphological characteristic differences were assessed using Fisher's exact test. Nodal short axis diameter and histographic R2* and ADC differences were analysed using a linear mixed model (Stata v13, StataCorp LLC, College Station, Texas, USA) to account for multiple samples per patient. Fitting was performed using response as the fixed factor and patient as the random factor. Means and 95% confidence intervals (CI) were estimated from the model for each patient response group. Where the data was right skewed the log measurement was used. Where the data was right skewed including negative values, a small constant was added to all values before the log transformation was applied with estimated means and 95% CIs back-transformed.

For the analysis of paired air and 100%-oxygen measurements within patient groups, a linear mixed model was fitted using measurement type (air or 100%-oxygen) as the fixed factor with patient and node as the random factors. Means and 95% confidence intervals were estimated from the model for each measurement type in the same manner.

5.4 RESULTS

5.4.1 Largest Node Per Patient Analysis

The results of this analysis are summarised in Table 2.

PARAMETER		**TREATMENT OUTCOME GROUP**		**p-value**
		Complete Response **(CR)**	Residual/ recurrent nodal disease **(RD)**	
Number patients (percentage)		32 (59%)	22 (41%)	
Qualitative Descriptors				
				Fisher's exact
Contour	Round	16 nodes	9 nodes	0.55
	Ovoid	16 nodes	13 nodes	
Margins	Irregular	20 nodes	10 nodes	0.17
	Regular	12 nodes	112 nodes	
Enhancement	Heterogenous	20 nodes	14 nodes	0.39
	Diffuse	11 nodes	8 nodes	
Necrosis	Present	19 nodes	17 nodes	0.32
	Absent	13 nodes	5 nodes	
Quantitative Parameters				
				Mann Whitney
Mean short axis diameter (mm)		18.5	22.6	0.21
Mean of ADC median values (10^{-3} mm^2/s)		0.91	0.89	0.99
Mean of ADC skewness		0.59	0.67	0.63
Mean of ADC kurtosis		1.04	1.2	0.72
Mean of R2* median values (s^{-1})	Air	25.6	20.2	0.04
	100% O2	27.6	22.5	0.05
	Wilcoxon p-value	0.012	0.055	
Mean of R2* skewness	Air	0.757	0.483	0.39
	100% O2	0.582	0.777	0.32
	Wilcoxon p-value	0.597	0.356	
Mean of R2* kurtosis	Air	1.57	0.489	0.64
	100% O2	0.667	1.37	0.44
	Wilcoxon p-value	0.36	0.33	

Table 5.2 Table summarising comparison between complete responding (CR) and residual/recurrent lymph node disease (RD) groups on a 'largest node per patient' (LNPP) basis. Yellow panes indicate those differences which are significant.

Short Axis Diameter and Qualitative Descriptors

Mean short axis diameter was slightly larger for the RD group though this difference was not significant (19mm for CR group and 23mm for RD group; p=0.21). There were no significant differences in the binary qualitative radiological descriptors between CR and RD groups (p values between 0.17 to 0.55).

R2* Parameters

Pairwise comparison revealed a significant increase in median R2* values when breathing 100%-oxygen compared to breathing air in CR group (mean $R2^*_{air}$=25.6ms, $R2^*_{O2}$=27.6ms; p=0.012), in RD group whilst there was also a trend to lengthen median T2* this was not significant (p=0.055) and the changes are illustrated in line graph for each largest node per patient in Figure 5.2. There were no significant pairwise differences for histogram R2* skewness or kurtosis (p values between 0.33 to 0.60).

When comparing absolute baseline median R2* values between the two groups, the CR group largest nodes were significantly longer in R2* times compared to RD group when breathing air (p=0.04) and close to significant when breathing 100%-oxygen (p=0.05), illustrated in box plot in Figure 3. No significant differences were observed in R2* histogram skewness or kurtosis on air or 100%-oxygen between the two cohorts (p values between 0.32 to 0.64).

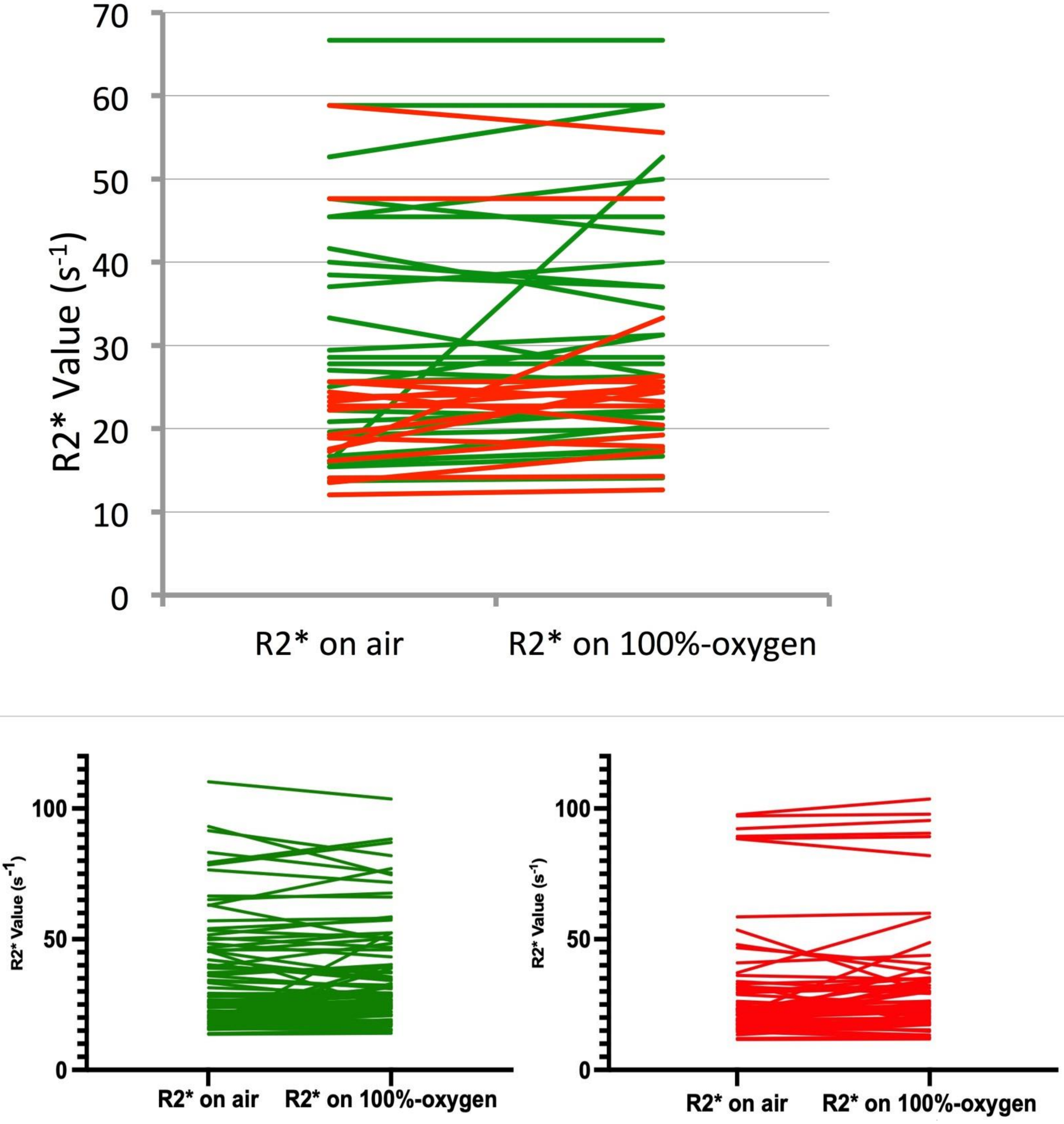


Figure 5-2 Line chart demonstrating change in R2* from air to 100%-oxygen in the largest node per patient for each CR patient (represented by a single green line) and RD patient (represented by red line) top pane and all nodes bottom.

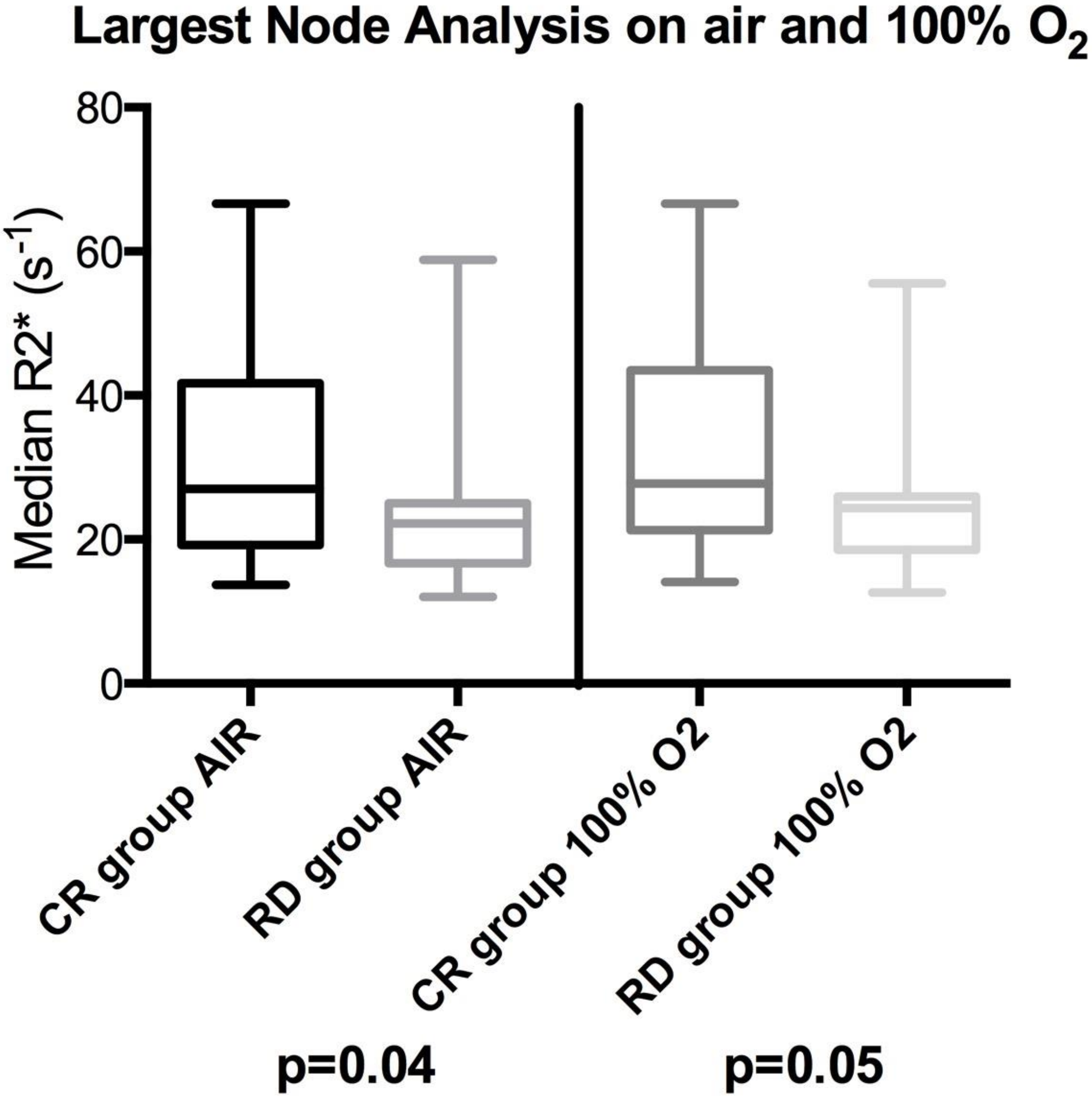


Figure 5-3 Box plots showing differences between complete responding pathological lymph node (CR) and residual/recurrent lymph node disease relapse (RD) groups median R2* distributions on largest node per patient basis. Left pane shows largest node node analysis on air and right pane on 100%-oxygen. Box indicates interquartile range, line median and whiskers most deviated range. Mann Whitney U p values are shown beneath each plot.

Diffusion weighted imaging Parameters

There were no significant differences between the largest nodal pretreatment median ADC values between the CR (ADC=0.91 x $10^{-3}mm^2/s$) and RD (ADC=0.89 x $10^{-3}mm^2/s$) groups (p=0.99). Furthermore, histogram skewness and kurtosis were also not significantly different between the two groups (p=0.63 and 0.72 respectively).

5.4.2 All Nodes Analysis

The results for this linear mixed model analysis are summarised in Table 3.

PARAMETER		**TREATMENT OUTCOME GROUP**		**p-value**
		Complete Response **(CR; n=32 pts/59%)**	Residual/ recurrent nodal disease **(RD; n=22 pts/41%)**	
Number of nodes (percentage)		104 (61.2%)	66 (38.9%)	
Qualitative Descriptors				
				Fisher's exact
Contour	Round	50 nodes	23 nodes	0.11
	Ovoid	54 nodes	43 nodes	
Margins	Irregular	50 nodes	38 nodes	0.27
	Regular	54 nodes	28 nodes	
Enhancement	Heterogenous	52 nodes	35 nodes	0.64
	Diffuse	52 nodes	31 nodes	
Necrosis	Present	48 nodes	38 nodes	0.16
	Absent	56 nodes	28 nodes	
Quantitative Parameters				
				linear mixed
Mean short axis diameter (mm)		12.0 (11.0;13.1)	13.5 (12.1;15.0)	0.21
Mean of ADC median values^{a} (10^{-3} mm^{2}/s)		0.86 (0.81;0.93)	0.88 (0.83;0.96)	0.91
Mean of ADC skewness		0.72 (0.55;0.93)	0.58 (0.38;0.79)	0.26
Mean of ADC kurtosis^{a}		3.74 (3.29;4.13)	3.71 (3.25;4.23)	0.97
Mean of R2* median values (s^{-1})	Air	26.4 (23.6;29.9)	22.1 (19.5;25.4)	0.049
	100% O2	28.1 (25.1;31.9)	23.6 (20.8;27.4)	0.07
	pairwise p-value	0.0006	0.14	
Mean of R2* skewness	Air^{b}	0.5 (0.4;0.7)	0.4 (0.3;0.6)	0.15
	100% O2^{b}	0.5 (0.4;0.7)	0.6 (0.4;0.7)	0.91
	pairwise p-value	>0.99	0.05	
Mean of R2* kurtosis	Air^{c}	0.2 (-0.1;0.6)	0.1 (-0.3;0.5)	0.51
	100% O2^{d}	0.5 (0.2;1.0)	0.2 (-0.2;0.7)	0.26
	pairwise p-value	0.36	0.68	

Table 5.3 Table summarising comparison between complete responding (CR) and residual/recurrent lymph node disease (RD) groups on an 'all nodes' (AN) basis using linear mixed models. Data in parentheses indicate 95% confidence intervals: [a]log(x) measurement used in linear mixed model, [b]log(x+1) measurement analysed in linear mixed model, [c]log(x+1.5) measurement analysed in linear mixed model, [d]log(x+2) measurement analysed in linear mixed model. Yellow panes indicate those differences which were significant.

Short Axis Diameter and Qualitative Descriptors

Mean short axis diameter of all pathological nodes was slightly larger for the RD group though this difference was not significant (12.0mm for 104 nodes in the CR group and 13.5mm for 66 nodes in the RD group; p=0.10). There were no significant differences in the binary qualitative radiological descriptors between CR and RD groups (p values between 0.11 to 0.64).

R2 Parameters*

All node pairwise comparison again revealed a significant increase in median R2* values when breathing 100%-oxygen compared to breathing air in CR group (mean R2*air=26.4s^{-1}, R2*O2=28.1s^{-1}; p=0.0006), and in RD group whilst there was also a trend to lengthen median R2* times, again this was not significant (p=0.14) and is also depicted in Fig 5.2. There were no significant pairwise differences for histogram R2* skewness or kurtosis (p values between 0.05 to >0.99).

When comparing absolute all node baseline median R2* values between the two groups, the CR group nodes were significant lower in R2* value compared to RD group when breathing air (p=0.049) though this difference was not significant on 100%-oxygen (p=0.07). No significant differences were observed in all node R2* skewness or kurtosis on air or 100%-oxygen between the two cohorts (p values between 0.15 to 0.91).

Diffusion weighted imaging Parameters

All node DWI comparison revealed there were no significant differences (p=0.91) between the pretreatment median ADC values between the CR group (ADC=0.86 x 10^{-3}mm^2/s) and the RD group (ADC=0.88 x 10^{-3}mm^2/s). Furthermore, ADC histogram skewness and kurtosis were also not significantly different between the two groups (p=0.26 and 0.97 respectively).

5.5 DISCUSSION

This study has demonstrated that cancerous lymph nodes in patients which subsequently respond to CRT demonstrate significant lengthening of R2* relaxation times when switching from breathing air to 100%-oxygen whilst nodes in patients who relapse within two years do not suggesting a paradoxical increase in deoxyhaemoglobin on breathing 100%-oxygen in nodes which respond favourably to CRT. To our knowledge, this is the first study to examine R2* changes exclusively within HNSCC lymph nodes in response to hyperoxic gas compared with air correlated with CRT response. The second finding is that responding nodes had significantly longer baseline R2* times on air compared with relapsing nodes which may imply more pretreatment hypoxia in responding nodes. Both of these findings are potentially counterintuitive given prior direct measurement hypoxia studies as well as several preclinical and early clinical studies evaluating oncological R2* imaging.

It is know that whilst the R2* of the vascular space is dependent on fractional blood oxygenation, it is also affected by other factors and is a quadratic function of haematocrit levels and magnetic field strength [31]. Whilst direct inference of tissue hypoxia based on quantitative R2* measurements may be problematic (though a moderate correlation with directly measured oxygen tension has been observed in prior validation studies [21]), differential response to breathing hyperoxic gas circumvents many of these additional confounding factors by evaluating the significance of relative R2* changes in the same sitting and thus will likely be the more important finding. Differences in R2* contrast in cancerous LNs to hyperoxic challenge between prior studies and the current findings may potentially be explained by the vascular properties of these nodes (most studies focus or combine with primary tumours) as well as the use of 100% oxygen without hypercapnia. The magnitude of R2* changes within tumours on breathing hyperoxic gas have been shown to mirror tumor hypoxic fraction changes as determined by pimonidazole labelling [26] though in preclinical studies the changes were not directly proportional to the absolute measured tissue oxygenation [32]. Kotas et al [23] examined changes in T2* values (1/R2*) in HNSCC tumours in response to breathing 2%-CO2/98%-oxygen gas and 100%-oxygen and demonstrated no significant tumoral T2* differences when breathing either hyperoxic gas mixture compared to air. However, this study was on a relatively small cohort (13 patients) and combined assessment of primary HNSCC tumours with involved lymph nodes (only six lymph nodes were examined). Other studies have examined the effect of breathing hyperoxic hypercapnic gas (2%-CO2/98%-oxygen) on T2* times in HNSCC [33] though examined primary tumours rather than lymph nodes and demonstrated lengthening in T2* times (shortening of R2*) with this gas mixture in eleven primary HNSCC tumours inferring increasing oxygenation on hyperoxic gas.

Unlike these hypercapnic challenges, 100%-oxygen as used in our study will cause different physiological effects on respiration, oxygenation and blood flow. It has been noted that breathing hyperoxic gas results in heterogenous effects on tumour blood flow [34] which may result in a 'steal' effect whereby blood flow may be increased at the expense of an adjacent location [35]. T2* measurement using multiple gradient echoes should be relatively robust to blood flow variation [36] however, perturbation due to vasomodulation can alter the effective concentration of blood in tissues [34] and a decrease in R2* rate will therefore only occur in hypoxic tumours with functional vasculature that respond in a certain manner to hyperoxic challenge [37]. A preclinical study by McPhail and Robinson [38] correlated R2* measurement in a breast tumour murine model validated with pimonidazole and other quantitative histological markers of hypoxia.

This demonstrated that tumours with a faster baseline R2* and a larger decrease in R2* with carbogen- breathing were more vascularized and had a greater functional blood volume than tumors with a slower baseline R2* and negligible response to carbogen- functional vasculature was central to the observed R2* contrast changes and was said to be dominated by blood volume. A clinical study has further reported a significant inverse correlation between breast tumor R2* and grade [39]. It was also noted that both benign and necrotic tumors were omitted from this clinical study, as necrosis can cause a paradoxical decrease in R2* owing to the lack of erythrocyte delivery [40].

A recent human prostate cancer study [41] has demonstrated that tumours exhibiting low R2* times stained positive whilst those with high R2* were negative to pimonidazole staining which was attributed to tumour regions occurring in tissues which were highly vascularised. It is increasingly clear that, in the context of hypoxia, R2* measurement can vary widely, and may be highly dependent on the tumor examined and its vasculature and further evaluation of the tissue bases of these contrast mechanisms is required.

Nonetheless, this first demonstration of a significant difference in R2* response to hyperoxic gas (and potentially in the baseline R2* characteristics) of responding nodes compared with those which relapse within two years of CRT is in itself important. We posit that the underlying the observed R2* differences between the outcome groups may thus relate to lymph node specific blood flow/vascular mechanisms as well as hypoxia which would potentially be useful for exploitation in biomarker development.

The requirement for predictive HNSCC biomarker development is reinforced by the demonstration that pre-treatment nodal short axis diameter and various conventionally employed qualitative morphological descriptors of malignancy were, perhaps unsurprisingly, not useful as discriminators of subsequent successful response to CRT.

Furthermore, this study also demonstrated that baseline diffusion weighted imaging parameters derived from monoexponential ADC were not significantly different between those nodes demonstrating sustained local complete post CRT response and those that subsequently relapse. Again, this is perhaps not unexpected as there have been conflicting results in other studies. A previous study specifically examining nodal HNSCC failure after CRT demonstrated significantly higher mean pre-treatment ADCs than in relapsing nodes compared with those that were controlled [42].

Other studies have found that, though there were trends towards higher pretreatment ADC predicting treatment failure, the relationship was not significant [43] whilst another study demonstrated the opposite relationship based on two-year disease free survival [44]. Additionally it should be highlighted that there is a lack of generalizability of DWI data given variable b value selection and indeed variability of ADC values across MRI systems and sequences [45].

This study has some limitations that should be considered. The sample size is relatively small though the analysis of multiple nodes in both groups per patient (with linear mixed modelling to account for sampling) does recapitulate the largest node per patient differences between CR and RD groups. Independent verification of nodal oxygenation was not undertaken (e.g. using invasive polarographic electrode measurement or histological verification) though would probably not be feasible in this cohort given the number of nodes examined. It is also difficult to account for true physiological oxygenation fluctuation in blood flow within tumour capillary network that can lead to well described transient or cyclical hypoxia [46]. As noted previously, there is considerable heterogeneity within tumour volumes particularly with respect to R2* and under hyperoxic breathing [23]- the use of volumes (rather than single axial regions) of interest was employed to mitigate this effect as far as possible. Previous clinical R2* studies have employed between 4 to 16 echo times though it has been demonstrated that the signal acquired with GRE sequences can be dominated by noise for longer echo times [25]. Lastly, this study used room air rather than medical air from cylinder.

In conclusion, this study demonstrates that the significant differential response to 100%-oxygen and higher baseline R2* measures between responding versus non-responding HNSCC lymph nodes could be exploited in risk stratification prior to CRT given that traditional qualitative radiological classifiers of nodal pathology and DWI parameters at baseline were poor discriminators of subsequent local response. Future work is, however, required to understand the contrast mechanisms of R2* imaging underpinning these observed differences in the context of hypoxia and would benefit from more formal evaluation of confounding factors (e.g. blood volume fraction and macroscopic field homogeneity) as well as direct histological verification in the HNSCC setting.

5.6 REFERENCES

1. Global Burden of Disease Cancer C: Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability- adjusted life-years for 32 cancer groups, 1990 to 2015: A systematic analysis for the global burden of disease study. JAMA Oncol. 2016.

2. Leemans CR, Tiwari R, Nauta JJ et al. Recurrence at the primary site in head and neck cancer and the significance of neck lymph node metastases as a prognostic factor. Cancer 1994; 73:187–190.

3. Lefebvre JL, Chevalier D, Luboinski B, Kirkpatrick A, Collette L, Sahmoud T. Larynx preservation in pyriform sinus cancer: preliminary results of a European Organization for Research and Treatment of Cancer phase III trial. EORTC Head and Neck Cancer Cooperative Group. J Natl Cancer Inst 1996; 88: 890-899.

4. Blanchard P, Baujat B, Holostenco V, Bourredjem A, Baey C, Bourhis J, Pignon J-P. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): A comprehensive analysis by tumour site. Radiother Oncol. 2011; 100(1):33–40.

5. Chen AM, Daly ME, Farwell DG, Vazquez E, Courquin J, Lau DH, Purdy JA. Quality of life among long-term survivors of head and neck cancer treated by intensity-modulated radiotherapy. JAMA Otolaryngol Head Neck Surg. 2014; 140(2):129–33.

6. MacKenzie RG, Franssen E, Balogh JM, Gilbert RW, Birt D, Davidson J. Comparing treatment outcomes of radiotherapy and surgery in locally advanced carcinoma of the

larynx: a comparison limited to patients eligible for surgery. Int J Radiat Oncol Biol Phys 2000; 47:65-71.

7. Leoncini E, Vukovic V, Cadoni G, Pastorino R, Arzani D, Bosetti C, Canova C, Garavello W, La Vecchia C, Maule M, et al. Clinical features and prognostic factors in patients with head and neck cancer: Results from a multicentric study. Cancer Epidemiol. 2015; 39(3):367–74.

8. Chakravarthy A, Henderson S, Thirdborough SM, Ottensmeier CH, Su X, Lechner M, Feber A, Thomas GJ, Fenton TR. Human Papillomavirus Drives Tumor Development Throughout the Head and Neck: Improved Prognosis Is Associated With an Immune Response Largely Restricted to the Oropharynx. J Clin Oncol. 2016; 0(0):JCO682955.

9. de Bondt RB, Nelemans PJ, Hofman PA et al. Detection of lymph node metastases in head and neck cancer: A meta-analysis comparing US, USgFNAC, CT and MR imaging. Eur J Radiol 2007; 64(2): 266-72.

10. de Bondt RB, Nelemans PJ, Bakers F, Casselman JW, Peutz-Kootstra C, Kremer B, Hofman PA, Beets-Tan RG. Morphological MRI criteria improve the detection of lymph node metastases in head and neck squamous cell carcinoma: multivariate logistic regression analysis of MRI features of cervical lymph nodes. Eur Radiol. 2009; 19(3):626-33.

11. Taabazuing, C. Y., J. A. Hangasky, and M. J. Knapp. Oxygen sensing strategies in mammals and bacteria. J. Inorg. Biochem. 2014; 133:63–72.

12. Nguyen, L. K., M. A. Cavadas, C. C. Scholz, S. F. Fitzpatrick, U. Bruning, E. P. Cummins, et al. A dynamic model of the hypoxia-inducible factor 1 (HIF-1) network. J. Cell Sci. 2013; 126:1454–1463.

13. Powers WE, Tolmach LJ. Demonstration of an anoxic component in mouse tumour cell population by in vivo assay of survival following irradiation. Radiology 1964; 83:328 –336.

14. Fleckenstein W, Weiss C, Heinrich R, Schomerus H, Kersting T. A new method for the bed-side recording of tissue pO2 histograms. Verh Dtsch Ges Inn Med 1984;90:439–443.

15. Terris DJ, Dunphy EP. Oxygen tension measurements of head and neck cancers. Arch Otolaryngol Head Neck Surg 1994;120(3):283–287.

16. Nordsmark M, Bentzen SM, Rudat V, Brizel D, Lartigau E, Stadler P, Becker A, Adam M, Molls M, Dunst J, Terris DJ, Overgaard J. Prognostic value of tumor oxygenation in 397 head and neck tumors after primary radiation therapy. An international multi-center study. Radiother Oncol. 2005 Oct; 77(1):18-24.

17. Terris DJ. Head and neck cancer: the importance of oxygen. Laryngoscope. 2000 May; 110(5 Pt 1):697-707.
18. Brown JM. Tumor hypoxia in cancer therapy. Methods Enzymol 2007; 435:297–321.
19. Tadamura E, Hatabu H, Li W, Prasad PV, Edelman RR. Effect of oxygen inhalation on relaxation times in various tissues. J Magn Reson Imaging 1997; 7:220–225.
20. Kim CK, Park SY, Park BK, Park W, Huh SJ. Blood oxygenation level- dependent MR imaging as a predictor of therapeutic response to concurrent chemoradiotherapy in cervical cancer: a preliminary experience. Eur Radiol. 2014; 24(7):1514–20.
21. Kotas M, Schmitt P, Jakob PM, Flentje M. Monitoring of tumor oxygenation changes in head-and-neck carcinoma patients breathing a hyperoxic hypercapnic gas mixture with a noninvasive MRI technique. Strahlenther Onkol 2009;185:19–26.
22. Li XS, Fan HX, Fang H, Song YL, Zhou CW. Value of R2* obtained from T2*-weighted imaging in predicting the prognosis of advanced cervical squamous carcinoma treated with concurrent chemoradiotherapy. J Magn Reson Imaging. 2015; 42(3):681-8.
23. Krishna MC, Subramanian S, Kuppusamy P, Mitchell JB. Magnetic resonance imaging for in vivo assessment of tissue oxygen concentration. Semin Radiat Oncol. 2001; 11(1):58–69.
24. Christen T, Bolar DS, Zaharchuk G. Imaging brain oxygenation with MRI using blood oxygenation approaches: methods, validation, and clinical applications. AJNR Am J Neuroradiol 2013;34:1113–1123.
25. Panek R, Welsh L, Dunlop A, Wong KH, Riddell AM, Koh D-M, Schmidt MA, Doran S, McQuaid D, Hopkinson G, Richardson C, Nutting CM, Bhide SA, Harrington KJ, Robinson SP, Newbold KL, Leach MO. Repeatability and sensitivity of T2* measurements in patients with head and neck squamous cell carcinoma at 3T. J Magn Reson Imaging 2016; 44(1): 72–80.
26. Jiang L, Weatherall PT, McColl RW, Tripathy D, Mason RP. Blood oxy- genation level-dependent (BOLD) contrast magnetic resonance imaging (MRI) for prediction of breast cancer chemotherapy response: a pilot study. J Magn Reson Imaging 2013;37:1083–1092.
27. Padhani AR. Science to practice: what does MR oxygenation imaging tell us about human breast cancer hypoxia? Radiology 2010; 254:1–3.
28. King AD, Thoeny HC. Functional MRI for the prediction of treatment response in head and neck squamous cell carcinoma: potential and limitations. Cancer Imaging 2016; 16(1):23.

29. Edge SB, Byrd DR, Carducci MA, et al, eds. American Joint Committee on Cancer (AJCC) Cancer Staging Manual. 7th ed. New York: Springer; 2009.

30. Hagberg GE, Indovina I, Sanes JN, Posse S. Real-time quantification of T2* changes using multiecho planar imaging and numerical methods. Magn Reson Med. 2002; 48(5):877-82.

31. Zhao JM, Clingman CS, Narvainen MJ, Kauppinen RA, van Zijl PC. Oxygenation and hematocrit dependence of transverse relaxation rates of blood at 3T. Magn Reson Med. 2007; 58:592–597.

32. Baudelet C, Ansiaux R, Jordan BF, Havaux X, Macq B, Gallez B. Physiological noise in murine solid tumors using T2*-weighted gradient- echo imaging: a marker of tumor acute hypoxia? Phys Med Biol. 2004; 49:3389–3411.

33. Rijpkema M, Kaanders JH, Joosten FB, van der Kogel AJ, Heerschap A. Effects of breathing a hyperoxic hypercapnic gas mixture on blood oxygenation and vascularity of head-and-neck tumors as measured by magnetic resonance imaging. Int J Radiat Oncol Biol Phys. 2002; 53: 1185–1191.

34. Neeman M, Dafni H, Bukhari O, et al. In vivo BOLD contrast MRI mapping of subcutaneous vascular function and matura- tion: Validation by intravital microscopy. Magn Reson Med. 2001; 45:887– 898.

35. Al Hallaq HA, River JN, Zamora M, et al. Correlation of magnetic resonance and oxygen microelectrode measurements of carbogen-induced changes in tumour oxygenation. Int J Radiat Oncol Biol Phys 1998; 41:151–159.

36. Lebon V, Carlier PG, Brillault-Salvat C, et al. Simultaneous measurement of perfusion and oxygenation changes using a multiple gradient-echo sequence: Application to human muscle study. Magn Reson Imaging 1998; 16:721–729.

37. Howe FA, Robinson SP, McIntyre DJO, Stubbs M, Griffiths JR. Issues in flow and oxygenation dependent contrast (FLOOD) imaging of tumours. NMR Biomed 2001; 14:497–506.

38. McPhail LD, Robinson SP. Intrinsic susceptibility MR imaging of chemically-induced rat mammary tumors: relationship to histological assessment of hypoxia and fibrosis. Radiology 2009; 254(1):110–118.

39. Padhani AR, Ah-See MLW, Taylor NJ, et al. An investigation of histological and DC-MRI correlates of intrinsic susceptibility contrast relaxivity (R2*) in human breast cancer [abstr]. In: Proceedings of the Thirteenth Meeting of the International Society for Magnetic

Resonance in Medicine. Berkeley, Calif: International Society for Magnetic Resonance in Medicine, 2005; 1846.

40. Alonzi R, Padhani AR, Maxwell RJ, et al. Carbogen breathing increases prostate cancer oxygenation: a translational MRI study in murine xenografts and humans. Br J Cancer 2009;100(4):644-648.

41. Hoskin PJ, Carnell DM, Taylor NJ, et al. Hypoxia in prostate cancer: Correlation of BOLD-MRI with pimonidazole immuno- histochemistry: initial observations. Int J Radiat Oncol Biol Phys 2007;68(4):1065– 1071.

42. Kim S, Loevner L, Quon H, Sherman E, Weinstein G, Kilger A, et al. Diffusion-weighted magnetic resonance imaging for predicting and detecting early response to chemoradiation therapy of squamous cell carcinomas of the head and neck. Clin Cancer Res. 2009; 15:986–94.

43. King AD, Chow KK, Yu KH, Mo FK, Yeung DK, Yuan J, et al. Head and neck squamous cell carcinoma: diagnostic performance of diffusion-weighted MR imaging for the prediction of treatment response. Radiology 2013; 266:531–8.

44. Nakajo M, Nakajo M, Kajiya Y, Tani A, Kamiyama T, Yonekura R, et al. FDG PET/CT and diffusion-weighted imaging of head and neck squamous cell carcinoma: comparison of prognostic significance between primary tumor standardized uptake value and apparent diffusion coefficient. Clin Nucl Med. 2012;37:475–80.

45. Kolff-Gart AS, Pouwels PJ, Noij DP, Ljumanovic R, Vandecaveye V, de Keyzer F, et al. Diffusion-weighted imaging of the head and neck in healthy subjects: reproducibility of ADC values in different MRI systems and repeat sessions. AJNR Am J Neuroradiol. 2015; 36:384-90.

46. Kimura H, Braun RD, Ong ET, et al. Fluctuations in red cell flux in tumor microvessels can lead to transient hypoxia and reoxygenation in tumor parenchyma. Cancer Res. 1996; 56:5522-5528.

Chapter 6 Clinical Validation (Qualification) of Existing Quantitative MR Biomarker in Tumour Response Prediction- diffusion weighted modelling in Head and Neck cancer

CONTRIBUTION:

Primary co-author responsible for data analysis (barring extended DWI modelling which was performed with post-doctoral colleague), drafting, revision, shared final approval and co-responsible for accuracy/integrity.

6.1 RUNNING TITLE

Comparison of diffusion MRI characteristics of cervical lymph nodes using exponential diffusion models in patients with head and neck squamous cell carcinoma

6.2 INTRODUCTION

Computed tomography (CT) and magnetic resonance imaging (MRI) have similar performance for detection of cervical nodal metastasis with a reported accuracy 78% and 82% respectively [1]. Local disease recurrence following therapy can occur in over 50% of patients with head and neck squamous cell carcinoma (HNSCC) who present with nodal metastases at initial staging [2]. One possible explanation for local recurrence maybe the inaccuracy of detecting all disease sites using conventional imaging, resulting in undertreatment of undetected disease sites.

In oropharyngeal HNSCC, the identification of the human papilloma virus (HPV) status is an important prognostic feature [3]. Although there is no consensus in terms of the best method for HPV detection, p16 immunohistochemistry (IHC) is considered as a sufficient surrogate marker [4], supporting a more favourable prognosis with p16 overexpression [5]. However, in non-oropharyngeal HNSCC, the prognostic value of p16 is still uncertain [6]. Whilst distribution of disease provides some prognostic information, by itself it is unable to differentiate between patients that do and do not demonstrate recurrence following chemoradiotherapy (CRT).

Diffusion weighted (DW) MRI and the assessment of apparent diffusion coefficient (ADC) are considered powerful imaging tools in oncology [7]. The last decade has seen increased exploration of ADC as an imaging biomarker for both the detection [8-12] and treatment response prediction [9-15] in the management of HNSCC. However, studies report varying degrees of success of using DW-MRI (with simple calculation of mono-exponential ADC) for these purposes [16, 17]. More complex diffusion models [18, 19] and analysis protocols exist and can potentially provide more specific information of individual diffusion components, such as histogram analysis of ADC distribution to provide inter-voxel heterogeneity assessment or a stretched exponential diffusion fitting model [20] which can provide information about intra-voxel diffusion heterogeneity. Imaging heterogeneity has previously been correlated with a number of cancer types [21, 22].

Diffusion heterogeneity assessed by histogram analysis in HNSCC patients has previously been reported as helpful in classifying primary tumour sites [23-26] and predicting treatment response [27]. To date there is little reported on the application of diffusion heterogeneity to the classification of HNSCC lymph nodes or the prediction of local nodal disease recurrence following chemo-radiotherapy.

The aim of this study was therefore to determine whether pre-treatment nodal microstructural heterogeneity, as reflected through diffusion MRI characteristics, can classify diseased from normal nodes and/or predict local disease recurrence of patients with HNSCC.

6.3 MATERIALS AND METHODS

6.3.1 Patient population

Institutional review board approval was achieved and written informed consent obtained from all patients and healthy volunteers for this prospective study. From March 2010 to July 2017, 79 patients with histologically confirmed head and neck SCC with unilateral cervical nodal N2/N3 metastatic disease [28] were recruited prior to chemo-radiotherapy or radiotherapy only. 5 patients declined trial entry and 2 had neck dissection without any residual disease. Subsequently, 72 recruited patients underwent up to two years (2yr) of clinical, imaging and histopathological follow-up. Exclusion criteria were related to MR imaging, i.e incomplete 2yr MR dataset, claustrophobia, pacemaker, metallic implant, protocol deviations, poor image quality. Patients that were recruited and died, prior to completion of 2yr follow-up, of causes unrelated to HNSCC were excluded from analysis. 14/72 patients were excluded from analysis; 8 due to incomplete 2yr follow-up, 2 due to death from causes unrelated to HNSCC within the 2yr follow up period; and 4 due to inappropriate MR imaging. Patient selection is illustrated in Figure 1. 58/72 patients (mean age 57.1 years; range 25 to 79 years), 46 males (mean age 57.7 years; range 33 to 79 years) and 12 females (mean age 54.75 years; range 25 to 74 years), completed 2yr follow-up. To compare the microstructural heterogeneity between diseased and healthy nodes, 8 healthy volunteers were also recruited (mean age 43.4 years; range 19 to 65 years), 3 male (mean age 58 years; range 50 to 65 years) and 5 female (mean age 39.4 years; range 19 to 56 years), from March 2010 to March 2012. For the volunteers' scans, inclusion criteria were no previous malignancy and infective disease and the exclusion criteria were contraindications to MR imaging.

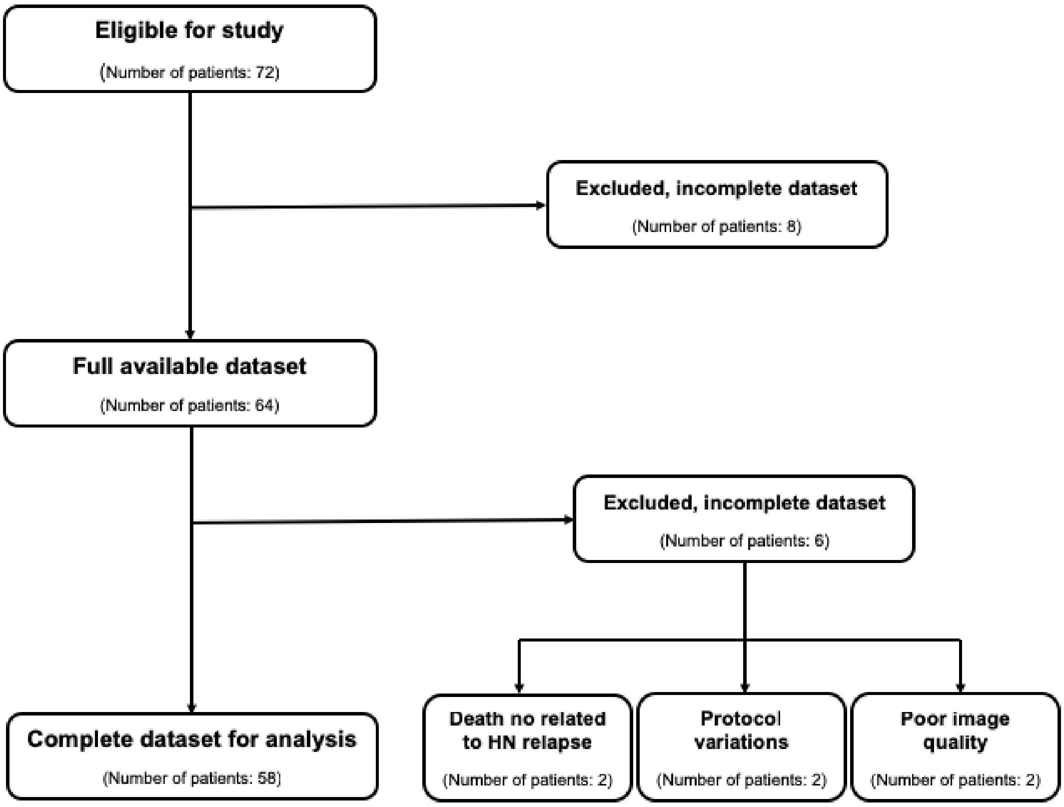


Figure 6-1 Trial summary flow chart illustrating the patient selection

6.3.2 Nodal staging and p16 assessment

Two experienced head and neck radiologists (with 18 and 10 years of experience respectively) reviewed all CT and anatomical MRI and performed neck ultrasound (US) in all patients, as per local standard of care. Cervical nodes were assessed as per the Union for International Cancer (UICC):TNM Classification of Malignant Tumour [29]. The largest node within the neck was sampled by US guided fine needle aspiration (FNA) and N2a/N2b or N3status confirmed.

In 33/58 patients, fluorescence in situ hybridization (FISH) was performed for the assessment of p16 expression. Not all the oropharyngeal patients underwent p16 testing as this became standard of care at our institution during the course of this study.

6.3.3 Treatment outcome

Each patient received a 2yr follow up, including clinical, radiological and histopathology examinations approximately every 3 months. Based on the multidisciplinary consensus review of this follow up, the patients were categorised into post-therapy local complete response (CR) and nodal recurrent disease (RD) depending on the presence or absence of any local disease recurrence.

6.3.4 Research MRI protocol

Patients pre-treatment and healthy volunteers were scanned on a 1.5T MR scanner (MAGNETOM Avanto, Siemens AG, Erlargen, Germany) using the carotid coils in the supine position. Axial T2 weighted (Periodically Rotated Overlapping ParallEL Lines with Enhanced Reconstruction (PROPELLER)) and DW images were acquired covering the neck, base of skull to upper thorax (Table 1). For the DW acquisition, images with 6 different b-values were acquired. The low b-values (b= 0, 50 ,100 s/mm2) images were obtained for reflecting the signal loss mainly due to micro-capillary perfusion and the higher b-values (b= 300, 600 ,1000 s/mm2) images for reflecting the signal loss mainly of the diffusion component. This approach provided a clinically acceptable acquisition time of 4 to 5 minutes.

Parameter	T2 weighted PROPELLER	Diffusion weighted
Sequence	Turbo spin echo	Single shot echo planar imaging
Slice orientation	Axial	Axial
Field of view [mm (read) x mm (phase)]	180 × 180	206 × 206
Acquired matrix (read)	256	128
Reconstructed matrix (read)	256	256

Number of signal averages	1	4
Slice thickness (mm)	3	4
Slice gap (mm)	0.3	0.4
Number of slices	60	40 to 46
Parallel imaging reduction factor	2	2
Echo time (ms)	107	88
Repetition time	6310	8700
Fat suppression	n/a	Short tau inversion recovery
Diffusion weightings (s/mm^2)	n/a	0, 50, 100, 300, 600 and 1000

PROPELLER – Periodically Rotated Overlapping ParallEL Lines with Enhanced Reconstruction

Table 6.1 T2 and diffusion weighted pulse sequence parameters.

6.3.5 Diffusion parameter extraction

Mono-exponential (Eq. 1) and stretched exponential (Eq. 2) models were fitted to all pixels of the trace DW images of each slice using a least-squares fit for all b-values (MATLAB 2016; MathWorks Inc., Natick, MA). A radiologist with 5 years of experience, aware of the location of sampled nodes and nodal status at conventional staging, reviewed the T2- weighted PROPELLER and trace DW images. Volumes of interest (ROI) were drawn using Jim 5.0 software (Xinapse Systems, Thorpe, Waterville, UK). For the patient groups, the radiologist contoured the ROIs on DW images of b=300s/mm2 on each imaging slice containing the metastatic node, encompassing all solid nodal tissue while avoiding any macroscopic necrotic and cystic area (Figure 2). For the healthy volunteers, the radiologist contoured one or more normal node ROIs per volunteer on DW images of b=300s/mm2 on each imaging slice containing the specified node taking care to only include nodal tissue whereby each selected node was clearly visible and suitable to draw around. For both cases, the ROIs from the DW images were electronically transferred to the corresponding calculated maps and a list of voxel by voxel diffusion parameter values were derived for each node (Figure 3). For each patient and volunteer node: the histogram distribution of ADC and DDC values for the entire nodal volume and the median α value were estimated.

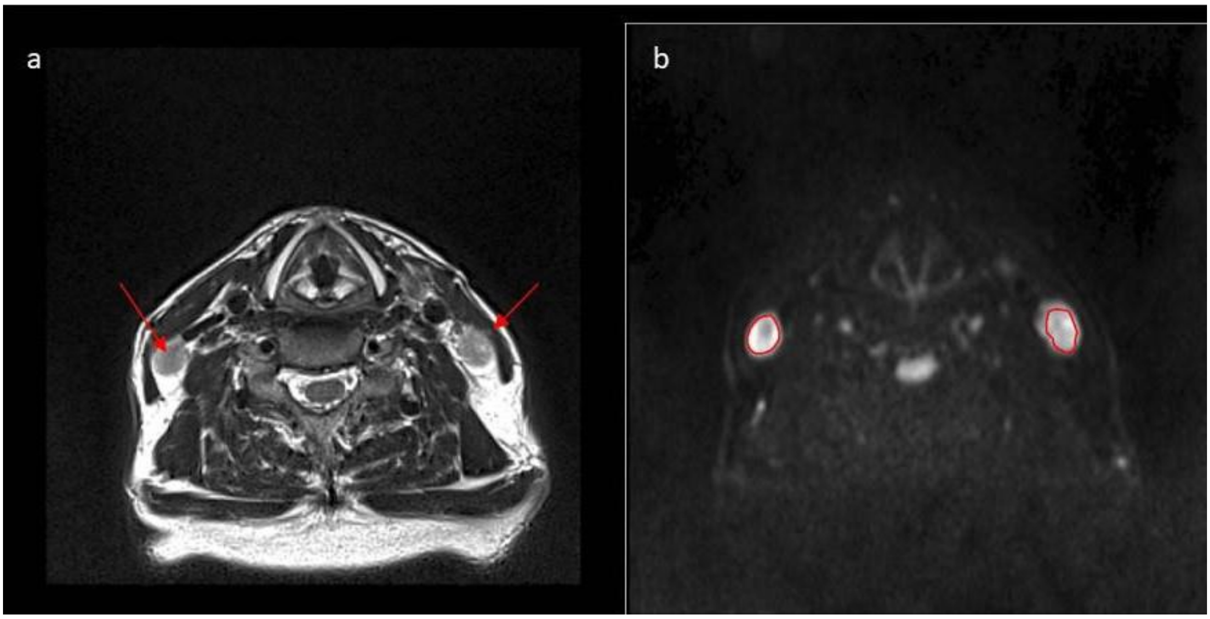


Figure 6-2 (a) Axial Periodically Rotated Overlapping ParallEL Lines with Enhanced Reconstruction (PROPELLER) T2-weighted image demonstrating two diseased nodes, arrows denoted the diseased nodes) in 56-year old man. (b) Axial diffusion weighted trace (DW) image (b=300 s/mm2) with the contoured regions of interest in 56-year old man.

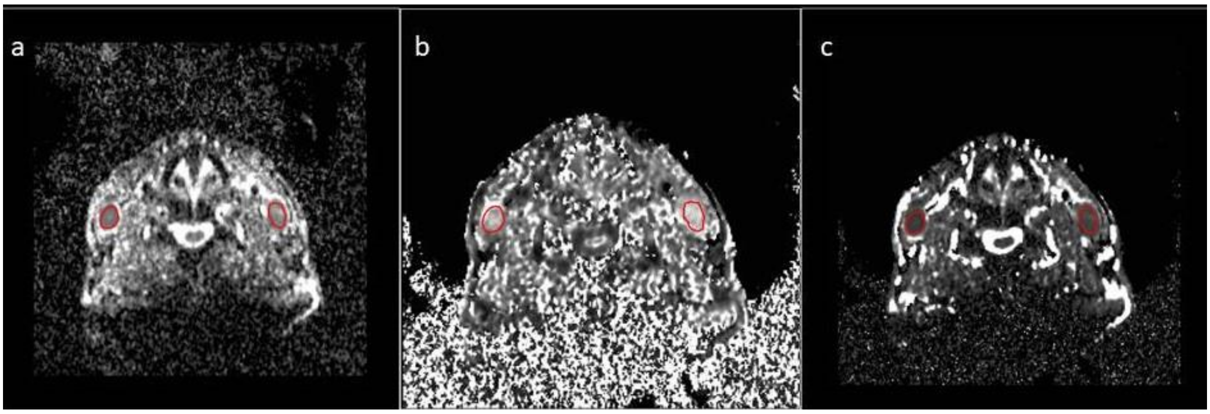


Figure 6-3 Calculated parametric diffusion maps (a) apparent diffusion coefficient (ADC) map, (b) alpha (⍰) value map and (c) distributed diffusion coefficient (DDC) map with the contoured ROIs in 56-year old man.

6.3.6 Statistical analysis

To assess if there are significant differences between the histogram metrics of ADC and DDC values and the median α values of the patients' groups and the healthy volunteers, each variable was compared between the groups. For the statistical analysis, a linear multilevel (mixed) model was used to account for patient's contributing more than one node and for the inter-patient variability (StataCorp 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP). Two-level models were used with measurements from individual nodes nested within patients. The majority of the diffusion parameters were analysed on the original scale of each measurement. The exceptions were the kurtosis values, which were found to have highly positively skewed distributions, and were thus analysed on the logarithmic scale. A significance level of 0.05 was used for all comparisons.

The difference in p16 status between CR and RD patients was assessed, in the subgroup where this information was measured. Due to the relatively small numbers of patients with the p16 assessment, this analysis was performed using Fishers exact test.

6.4 Results

In this study, 205 cancerous nodes were analysed from 58 patients and 40 normal nodes from the 8 healthy volunteers. Patient demographics, primary tumour site and TNM stage are shown in Table 2. After the 2yr follow-up, 32 patients (55%) had confirmed local complete response (109 nodes from CR patients) and 26 patients (45%) demonstrated local nodal recurrent disease (96 nodes from RD patients).

The pre-treatment ADC histogram derived metrics of patients and healthy volunteers and their statistical comparisons are presented in Table 2. Tables 3 summarises the pre-treatment histogram and statistical analysis of the DDC values assessing the inter-voxel heterogeneity, as well as the median α values assessing the intra-voxel heterogeneity utilising the stretched exponential model of each patient group and healthy volunteers and their comparisons. The p16 status of CR and RD patient groups are summarised in Table 4.

	Patients (CR and RD) group (58 patients, 205 nodes)	**CR patient group** (32 patients, 109 nodes)	**RD patient group** (26 patients, 96 nodes)	**Healthy Volunteers** (8 volunteers, 40 nodes)
ADC (10^{-3} mm^2/s)				
Minimum	0.46 ± 0.20	0.46 ± 0.18	0.47 ± 0.22	0.46 ± 0.17
25th percentile	0.74 ± 0.17	0.74 ± 0.11	0.75 ± 0.21	0.67 ± 0.12
75th percentile	0.96 ± 0.21	0.95 ± 0.16	0.98 ± 0.25	0.88 ± 0.16

Maximum	1.48 ± 0.35	1.48 ± 0.36	1.49 ± 0.33	1.14 ± 0.23
Median	0.85 ± 0.18	0.84 ± 0.13	0.86 ± 0.23	0.78 ± 0.13
Mean	0.86 ± 0.18	0.85 ± 0.13	0.87 ± 0.22	0.78 ± 0.13
Skewness	0.63 ± 0.71	0.66 ± 0.76	0.59 ± 0.64	0.16 ± 0.51
Kurtosis	3.54 [2.87, 4.51]	3.70 [3.00, 4.57]	3.32 [2.77, 4.48]	3.05 [2.35, 3.52]
Comparisons				
	Patients (CR and RD) group versus healthy volunteers	**CR group versus healthy volunteers**	**RD group versus healthy volunteers**	**CR group versus RD group**
		P value		
Minimum	0.82	1.00	1.00	1.00
25th percentile	0.16	0.73	0.41	1.00
75th percentile	0.16	1.00	0.42	1.00
Maximum	0.22	0.003	0.006	1.00
Median	0.23	1.00	0.47	1.00
Mean	0.14	0.69	0.31	1.00
Skewness	0.001	0.003	0.02	0.95
Kurtosis	0.001	<0.001	0.04	0.47

Table 6.2 ADC histogram results from patients and healthy volunteers (including mean ± standard deviation for all the metrics, and median and inter-quartile range for kurtosis), and statistical results (p values) derived from the different comparisons (ADC: apparent diffusion coefficient; CR: Post-therapy local complete response patients; RD: Post-therapy nodal recurrent disease patients).

	Patients group	CR patient	RD patient	Healthy

	(58 patients, 205 nodes)	**group** (32 patients, 109 nodes)	**group** (26 patients, 96 nodes)	**Volunteers** (8 volunteers, 40 nodes)
DDC (10^{-3} mm^2/s)				
Minimum	0.40 ± 0.23	0.38 ± 0.21	0.41 ± 0.26	0.36 ± 0.20
25th percentile	0.76 ± 0.20	0.76 ± 0.16	0.77 ± 0.23	0.68 ± 0.16
75th percentile	1.07 ± 0.28	1.06 ± 0.27	1.08 ± 0.29	0.97 ± 0.24
Maximum	2.42 ± 0.96	2.40 ± 1.01	2.44 ± 0.90	1.85 ± 0.92
Median	0.90 ± 0.22	0.89 ± 0.20	0.90 ± 0.24	0.81 ± 0.17
Mean	0.95 ± 0.24	0.94 ± 0.23	0.95 ± 0.25	0.84 ± 0.19
Skewness	1.43 ± 1.14	1.40 ± 1.16	1.46 ± 1.11	1.03 ± 1.28
Kurtosis	5.96 [3.95, 10.03]	6.08 [4.28, 10.21]	5.74 [3.71, 9.89]	4.20 [2.60, 9.61]
Comparisons				
	Patients group (CR and RD) versus healthy volunteers	**CR group versus healthy volunteers**	**RD group versus healthy volunteers**	**CR group versus RD group**
		P value		
Minimum	0.86	1.00	1.00	1.00
25th percentile	0.15	0.65	0.40	1.00
75th percentile	0.29	1.00	0.71	1.00
Maximum	0.008	0.05	0.03	1.00
Median	0.31	0.80	0.50	1.00
Mean	0.15	0.65	0.40	1.00
Skewness	0.06	0.28	0.18	1.00

Kurtosis	0.03	0.12	0.12	1.00
	Patients group (58 patients, 205 nodes)	**CR patient group** (32 patients, 109 nodes)	**RD patient group** (26 patients, 96 nodes)	**Healthy Volunteers** (8 volunteers, 40 nodes)
α				
Median	0.76 ± 0.09	0.75 ± 0.07	0.78 ± 0.10	0.73 ± 0.08
Comparisons				
	Patients group (CR and RD) versus healthy volunteers	**CR group versus healthy volunteers**	**RD group versus healthy volunteers**	**CR group versus RD group**
	P value			
α				
Median	0.31	1.00	0.54	0.92

Table 6.3 DDC histogram results and alpha (α) values from each patient group and from the healthy volunteers (including mean ± standard deviation for all the metrics, and median and inter-quartile range for kurtosis), and statistical results (p values) derived from the different comparisons (DDC: distributed diffusion coefficient; α: alpha value; CR: Post-therapy local complete response patients; RD: Post-therapy nodal recurrent disease patients).

	Number of CR patients (%)	**Number of RD patients (%)**	**P value**
p16 status			
p16 positive (p16+)	19 (86%)	6 (55%)	0.08
p16 negative (p16-)	3 (14%)	5 (45%)	

Table 6.4 p16 status of the patients and comparison (p value) between the two patient groups (CR: Post-therapy local complete response patients; RD: Post-therapy nodal recurrent disease patients).

6.4.1 Healthy volunteers vs. Patients analysis

The ADC nodal histogram distribution was more skewed (0.63, 0.66, 0.59, 0.16 for all patients, CR group, RD group and healthy volunteers respectively) and more “peaked” (kurtotic) (3.54, 3.70, 3.32, 3.05 for all patients, CR group, RD group and healthy volunteers respectively) in patient groups compared with healthy volunteers (p-value≤0.001 for skewness and p-value: from 0.001 to 0.003 for kurtosis). The median nodal ADC value of the healthy volunteers was slightly lower compared to the patient groups (0.85 x 10-3 mm2/s, 0.84 x 10-3 mm2/s, 0.86 x 10-3 mm2/s, 0.78 x 10-3 mm2/s for all patients, CR group, RD group and healthy volunteers respectively), but this did not reach statistical significance (p-value from 0.23 to 1.00) .

The DDC nodal histogram distribution demonstrated significant increased kurtosis (5.96, 6.08, 5.74, 4.2 for all patients, CR group, RD group and healthy volunteers respectively) in the overall patient group compared with healthy volunteers (p-value=0.03 between patients and volunteers). However, this did not reach statistical significance. The median nodal DDC value of healthy volunteers was lower compared to the patient groups (0.90 x 10-3 mm2/s, 0.89 ⍰ 10-3 mm2/s, 0.90 x 10-3 mm2/s, 0.81 x 10-3 mm2/s for all patients, CR group, RD group and healthy volunteers respectively), however differences did not reach statistical significance (p-value from 0.31 to 1.00).The median nodal stretched exponential derived α value of healthy volunteers was lower compared to patient groups (nodal DDC value of the healthy volunteers was lower compared to the patient groups (0.85, 0.84, 0.86, 0.78 for all patients, CR group, RD group and healthy volunteers respectively), but not significantly different (p-value from 0.31 to 1.00).

6.4.2 Complete response vs Recurrent disease Patient analysis

CR patients were more likely to be p16-positive (p16+) with 86% of this group having this status, compared to only 55% of RD patients (Table 5). The comparisons of ADC and DDC histogram analysis didn’t show any statistically significant difference between CR and RD patient groups (for ADC comparisons: p-value ranged from 0.47 to 1.00 and for DDC comparisons: p-value=1.00). Lastly, no statistically significant difference in median α values (αCR=0.75, αRD=0.78) were shown between the two patient groups (p-value=0.92).

6.5 Discussion

Our study aimed to determine whether pre-treatment nodal microstructural heterogeneity, as reflected through diffusion MRI characteristics, could classify diseased from normal nodes and/or predict local disease recurrence of patients with HNSCC. We utilised two exponential models for the histogram analysis, namely the well-established mono-exponential model for the ADC assessment, and the stretched-exponential model for the assessment of DDC and α value derivation; the latter allows interrogation of intra-voxel diffusion heterogeneity [30-32]. Histogram analysis was utilised for ADC and DDC assessment to provide a measures of inter-voxel heterogeneity [20, 23, 33].

Our main finding was that nodal disease could be identified by assessing inter-voxel diffusion heterogeneity, through ADC histogram analysis. The significantly increased kurtosis in diseased patient nodes suggests a decrease in heterogeneity for the larger part of the diseased nodal volume with voxels with a narrower range of ADC values compared with the broader distribution of healthy volunteer nodes. The increased skewness in patients resulted from distributions with longer right tail, which may be due to micronecrosis [34], resulting in very high ADC values and in asymmetric ADC distributions. In HNSCC primary tumour, the value of the ADC histogram analysis has been applied to assess differentiation of tumour grades [25] and reflect tumour heterogeneity [33]. Our study suggests that a similar approach is valid for classification of lymph node disease status in HNSCC patients.

Our second, unexpected, finding was that the mean or median ADC values across the whole volume of the lymph nodes was not significantly different between the patients and the healthy volunteers. Other researchers have reported significant differences between benign and malignant cervical lymph nodes. However, in these studies [2, 8, 35-37] the patient population was smaller (ranged from 16 to 32 patients) than in our study (58 patients), no healthy lymph nodes were included and a mean ADC value was used for the ADC comparison. To our knowledge, there are no other studies utilising ADC histogram analysis of lymph nodes in patients with HNSCC. In the current study the ADC distribution of each node was assessed taking into account the contribution of all the voxels within that node.

The intra-voxel (α value) and the intervoxel (DDC histogram distribution) MRI diffusion heterogeneity assessed by the stretched exponential couldn’t classify the diseased nodes from the healthy volunteer nodes, and only a significant higher kurtosis in the DDC distributions of the diseased nodes was observed. Reported comparisons of the diagnostic performance of different exponential models at various cancers have revealed the potential value of DDC and α values in tumour discrimination in ovarian cancer [38], cervical carcinoma [39], grading of gliomas [40], and in prostate cancer [30, 41, 42]. However, in these studies mean and median of and DDC values were utilised only in the primary tumours. In our study, we not only assessed the median of α values of lymph nodes in HNSCC patients, but also the DDC histogram distribution. These results suggest that the stretched exponential model analysis was not able to identify nodal disease.

Our study didn't show any significant difference in any metric of the ADC histogram analysis between the two patient groups. There is an increased interest in investigating the correlation between the pre-treatment ADC and the treatment outcome in HNSCC tumours [9, 10, 43, 44]. In predicting treatment response through the nodal MRI diffusion characteristics assessment in HNSCC patients [12, 16, 34], lower ADC values have been reported in responders than in non-responders, whereas in these studies the mean and median ADC values were across the primary tumour only. In the current study, all the ADC values from all the voxels through the node were included for the ADC nodal histogram analysis. Similarly, the DDC histogram analysis and the median α value didn't reveal any significant differences between the patient groups. α value is a stretching parameter taking into account the distribution of the diffusion coefficients within a voxel [19]. In breast cancer [32] and in nasopharyngeal carcinoma staging [45], the mean DDC values and the α values were able to predict treatment response. However, in these studies DDC values were estimated across the tumour, while we estimated the DDC histogram analysis across the whole metastatic lymph nodes.

Recently HPV status is recognised as a prognostic marker in oropharyngeal HNSCC; patients with positive HPV responded better to treatment and demonstrated better overall survival as compared to the patients with negative HPV [5]. In our study, the assessment of HPV status, by p16 expression in a subgroup of patients, showed that CR group was more likely to be p16+ than the RD group, albeit without reaching statistically significance. Different groups have investigated whether there is a relationship between MRI diffusion characteristics and HPV status, reported either statistically significant correlation between them [4, 33] or not [46]. This type of analysis was outside the scope of the current study.

Our study had some limitations. The sample size was reduced, because a number of the enrolled patients didn't manage to complete the 2y follow-up, but the fact that our data were significant suggest they are valid. DW-MRI in head and neck is challenging due to the introduced magnetic field inhomogeneities across this area, resulting in DW images with low signal intensity and distortions [47]. In this study, DW images of six b-values were acquired, to ensure sufficient signal intensity covering the multiple exponential effects and resulting in a clinically acceptable scan time of 4 to 5 minutes. More accurate quantitation may be possible using a larger number of b-values [48] but would limit clinical acceptability. Lastly, not all the patient samples underwent a p16 assessment, because at the early stage of the study p16 assessment was not routinely performed at our institution. However, since over recent years p16 assessment is highly recommended by clinicians for oropharyngeal carcinomas we included the results at least in a subgroup of the patient cohort.

In summary, our quantitative analysis of pre-treatment DW-MRI suggest that the inter-voxel MR diffusion heterogeneity assessed by ADC histogram analysis can classify diseased from normal lymph nodes, but it cannot be used as a prognostic factor. The lymph nodes intra-voxel and inter-voxel MR diffusion heterogeneities assessed by the stretched exponential model, didn't classify normal from diseased nodes or predict treatment response in this study. Future studies are required to further investigate the utility of the intra-voxel diffusion heterogeneity, derived by α value, as an imaging biomarker for disease classification and for treatment response.

6.6 REFERENCES

1. van den Brekel MWM, Castelijns JA, Stel HV, Golding RP, Meyer CJ, Snow GB. Modern imaging techniques and ultrasound-guided aspiration cytology for the assessment of neck node metastases: a prospective comparative study. Eur Arch Otohinolaryngol. 1993;250(1):1027-32.

2. Pryor DI, Solomon B, Porceddu SV. The emerging era of personalized therapy in squamous cell carcinoma of the head and neck. Asia Pac J Clin Oncol. 2011;7(3):236-51.

3. Robinson M, Schache A, Sloan P, Thavaraj S. HPV specific testing: a requirement for oropharyngeal squamous cell carcinoma patients. Head Neck Pathol. 2012;6 Suppl 1:S83-90.

4. Nakahira M, Saito N, Yamaguchi H, Kuba K, Sugasawa M. Use of quantitative diffusion-weighted magnetic resonance imaging to predict human papilloma virus status in patients with oropharyngeal squamous cell carcinoma. Eur Arch Otorhinolaryngol. 2014;271(5):1219-25.

5. Albers AE, Qian X, Kaufmann AM, Coordes A. Meta-analysis: HPV and p16 pattern determines survival in patients with HNSCC and identifies potential new biologic subtype. Sci Rep. 2017;7(1):16715.

6. Chung CH, Zhang Q, Kong CS, Harris J, Fertig EJ, Harari PM, et al. p16 protein expression and human papillomavirus status as prognostic biomarkers of nonoropharyngeal head and neck squamous cell carcinoma. J Clin Oncol. 2014;32(35):3930-8.

7. Padhani AR, Liu G, Mu-Koh D, Chenevert TL, Thoeny HC, Takahara T, et al. Diffusion-Weighted Magnetic Resonance Imaging as a Cancer Biomarker: Consensus and Recommendations. Neoplasia. 2009;11(2):102-25.

8. de Bondt RB, Hoeberigs MC, Nelemans PJ, Deserno WM, Peutz-Kootstra C, Kremer B, et al. Diagnostic accuracy and additional value of diffusion-weighted imaging for discrimination of malignant cervical lymph nodes in head and neck squamous cell carcinoma. Neuroradiology. 2009;51(3):183-92.

9. Kato H, Kanematsu M, Tanaka O, Mizuta K, Aoki M, Shibata T, et al. Head and neck squamous cell carcinoma: usefulness of diffusion-weighted MR imaging in the prediction of a neoadjuvant therapeutic effect. Eur Radiol. 2009;19(1):103-9.

10. AD King KC, KH Yu, FK Fai Mo, DKW Yeung, J Yuan, KS Bhatia, AC Vlantis, AT Ahuja. Head and Neck Squamous Cell Carcinoma: diagnostic performance of diffusion-weighted MR Imaging for the prediction of treatment response. Radiology. 2013;266(2):531-8.

11. Hatakenaka M, Nakamura K, Yabuuchi H, Shioyama Y, Matsuo Y, Ohnishi K, et al. Pretreatment apparent diffusion coefficient of the primary lesion correlates with local failure in head-and-neck cancer treated with chemoradiotherapy or radiotherapy. Int J Radiat Oncol Biol Phys. 2011;81(2):339-45.

12. Zhang GY, Wang YJ, Liu JP, Zhou XH, Xu ZF, Chen XP, et al. Pretreatment Diffusion-Weighted MRI Can Predict the Response to Neoadjuvant Chemotherapy in Patients with Nasopharyngeal Carcinoma. Biomed Res Int. 2015;2015:307943.

13. Kim S, Loevner L, Quon H, Sherman E, Weinstein G, Kilger A, et al. Diffusion-weighted magnetic resonance imaging for predicting and detecting early response to chemoradiation therapy of squamous cell carcinomas of the head and neck. Clin Cancer Res. 2009;15(3):986-94.

14. King AD, Mo FK, Yu KH, Yeung DK, Zhou H, Bhatia KS, et al. Squamous cell carcinoma of the head and neck: diffusion-weighted MR imaging for prediction and monitoring of treatment response. Eur Radiol. 2010;20(9):2213-20.

15. Vaid S, Chandorkar A, Atre A, Shah D, Vaid N. Differentiating recurrent tumours from post-treatment changes in head and neck cancers: does diffusion-weighted MRI solve the eternal dilemma? Clin Radiol. 2017;72(1):74-83.

16. Chawla S, Kim S, Dougherty L, Wang S, Loevner LA, Quon H, et al. Pretreatment diffusion-weighted and dynamic contrast-enhanced MRI for prediction of local treatment response in squamous cell carcinomas of the head and neck. AJR Am J Roentgenol. 2013;200(1):35-43.

17. Zhou M, Lu B, Lv G, Tang Q, Zhu J, Li J, et al. Differential diagnosis between metastatic and non-metastatic lymph nodes using DW-MRI: a meta-analysis of diagnostic accuracy studies. J Cancer Res Clin Oncol. 2015;141(6):1119-30.

18. D Le Bihan EB, D Lallemand, ML Aubin , J Vignaud, M Laval-Jeantet Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology. 1988;168(2):497-505.

19. Bennett KM, Schmainda KM, Bennett RT, Rowe DB, Lu H, Hyde JS. Characterization of continuously distributed cortical water diffusion rates with a stretched-exponential model. Magn Reson Med. 2003;50(4):727-34.

20. Kwee TC, Galban CJ, Tsien C, Junck L, Sundgren PC, Ivancevic MK, et al. Comparison of apparent diffusion coefficients and distributed diffusion coefficients in high-grade gliomas. J Magn Reson Imaging. 2010;31(3):531-7.

21. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. Nature. 2013;501(7467):338-45.

22. Ganeshan B, Miles KA. Quantifying tumour heterogeneity with CT. Cancer Imaging. 2013;13:140-9.

23. Tozer DJ, Jager HR, Danchaivijitr N, Benton CE, Tofts PS, Rees JH, et al. Apparent diffusion coefficient histograms may predict low-grade glioma subtype. NMR Biomed. 2007;20(1):49-57.

24. Khayal IS, Nelson SJ. Characterization of low-grade gliomas using RGB color maps derived from ADC histograms. J Magn Reson Imaging. 2009;30(1):209-13.

25. Ahn SJ, Choi SH, Kim YJ, Kim KG, Sohn CH, Han MH, et al. Histogram analysis of apparent diffusion coefficient map of standard and high B-value diffusion MR imaging in head and neck squamous cell carcinoma: a correlation study with histological grade. Acad Radiol. 2012;19(10):1233-40.
26. Swartz JE, Driessen JP, van Kempen PMW, de Bree R, Janssen LM, Pameijer FA, et al. Influence of tumor and microenvironment characteristics on diffusion-weighted imaging in oropharyngeal carcinoma: A pilot study. Oral Oncol. 2018;77:9-15.
27. Ren JL, Yuan Y, Li XX, Shi YQ, Tao XF. Histogram analysis of apparent diffusion coefficient maps in the prognosis of patients with locally advanced head and neck squamous cell carcinoma: Comparison of different region of interest selection methods. Eur J Radiol. 2018;106:7-13.
28. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surg Oncol. 2010;17(6):1471-4.
29. Sobin LH. TNM classification of malignant tumours. 7th ed. Chichester, West Sussex, UK ;: Hoboken, NJ: John Wiley and Sons; 2009.
30. Mazaheri Y AA, Rowe DB, Lu Y, Shukla-Dave A, Grover J. Diffusion-weighted magnetic resonance imaging of the prostate: improved robustness with stretched exponential modeling. Journal of computer assisted tomography. 2012;36(6):695-703.
31. Winfield JM, deSouza NM, Priest AN, Wakefield JC, Hodgkin C, Freeman S, et al. Modelling DW-MRI data from primary and metastatic ovarian tumours. Eur Radiol. 2015;25(7):2033-40.
32. Bedair R, Priest AN, Patterson AJ, McLean MA, Graves MJ, Manavaki R, et al. Assessment of early treatment response to neoadjuvant chemotherapy in breast cancer using non-mono-exponential diffusion models: a feasibility study comparing the baseline and mid-treatment MRI examinations. Eur Radiol. 2017;27(7):2726-36.
33. de Perrot T, Lenoir V, Domingo Ayllon M, Dulguerov N, Pusztaszeri M, Becker M. Apparent Diffusion Coefficient Histograms of Human Papillomavirus-Positive and Human Papillomavirus-Negative Head and Neck Squamous Cell Carcinoma: Assessment of Tumor Heterogeneity and Comparison with Histopathology. AJNR Am J Neuroradiol. 2017;38(11):2153-60.

34. Scalco E, Marzi S, Sanguineti G, Vidiri A, Rizzo G. Characterization of cervical lymph-nodes using a multi-parametric and multi-modal approach for an early prediction of tumor response to chemo-radiotherapy. Phys Med. 2016;32(12):1672-80.

35. Vandecaveye V, De Keyzer, F., Vander Poorten, V., Dirix, P., Verbeken, E., Nuyts, S. and Hermans, R. Head and neck squamous cell carcinoma: value of diffusion-weighted MR imaging for nodal staging. Radiology. 2009;251(1):134-46

36. Perrone A, Guerrisi P, Izzo L, D'Angeli I, Sassi S, Mele LL, et al. Diffusion-weighted MRI in cervical lymph nodes: differentiation between benign and malignant lesions. Eur J Radiol. 2011;77(2):281-6.

37. Si J, Huang S, Shi H, Liu Z, Hu Q, Wang G, et al. Usefulness of 3T diffusion-weighted MRI for discrimination of reactive and metastatic cervical lymph nodes in patients with oral squamous cell carcinoma: a pilot study. Dentomaxillofacial Radiology. 2014;43(3).

38. Winfield JM, Orton MR, Collins DJ, Ind TE, Attygalle A, Hazell S, et al. Separation of type and grade in cervical tumours using non-mono-exponential models of diffusion-weighted MRI. Eur Radiol. 2017;27(2):627-36.

39. Lin M, Yu X, Chen Y, Ouyang H, Wu B, Zheng D, et al. Contribution of mono-exponential, bi-exponential and stretched exponential model-based diffusion-weighted MR imaging in the diagnosis and differentiation of uterine cervical carcinoma. Eur Radiol. 2017;27(6):2400-10.

40. Bai Y LY, Tian J, Shi D, Cheng J, Haacke EM, Hong X, Ma B, Zhou J, Wang M. Grading of gliomas by using monoexponential, biexponential, and stretched exponential diffusion-weighted MR imaging and diffusion kurtosis MR imaging. Radiology. 2016;278(2):496-504.

41. Liu X, Zhou L, Peng W, Wang H, Zhang Y. Comparison of stretched-Exponential and monoexponential model diffusion-Weighted imaging in prostate cancer and normal tissues. J Magn Reson Imaging. 2015;42(4):1078-85.

42. Mazaheri Y, Hotker AM, Shukla-Dave A, Akin O, Hricak H. Model selection for high b-value diffusion-weighted MRI of the prostate. Magn Reson Imaging. 2018;46:21-7.

43. Hatakenaka M, Nakamura K, Yabuuchi H, Shioyama Y, Matsuo Y, Kamitani T, et al. Apparent diffusion coefficient is a prognostic factor of head and neck squamous cell carcinoma treated with radiotherapy. Jpn J Radiol. 2014;32(2):80-9.

44. Lombardi M, Cascone T, Guenzi E, Stecco A, Buemi F, Krengli M, et al. Predictive value of pre-treatment apparent diffusion coefficient (ADC) in radio-chemiotherapy treated head and neck squamous cell carcinoma. Radiol Med. 2017;122(5):345-52.
45. Lai V, Lee VH, Lam KO, Sze HC, Chan Q, Khong PL. Intravoxel water diffusion heterogeneity MR imaging of nasopharyngeal carcinoma using stretched exponential diffusion model. Eur Radiol. 2015;25(6):1708-13.
46. Schouten CS, de Graaf P, Bloemena E, Witte BI, Braakhuis BJ, Brakenhoff RH, et al. Quantitative diffusion-weighted MRI parameters and human papillomavirus status in oropharyngeal squamous cell carcinoma. AJNR Am J Neuroradiol. 2015;36(4):763-7.
47. Thoeny HC DKF, King AD. Diffusion-weighted MR imaging in the head and neck. Radiology. 2012;263(1):19-32.
48. Fujima N ST, Homma A, Shimizu Y, Yoshida A, Harada T, Tha K, Kudo K, Shirato H. Advanced diffusion models in head and neck squamous cell carcinoma patients: Goodness of fit, relationships among diffusion parameters and compariosn with dynamic contrast-enhanced perfusion. Magn Reson Med. 2016;36:16-23.

SECTION D: BIOMARKERS, TEXTURAL ANALYSIS MODELS DERIVED FROM IMAGE ANALYSIS

Chapter 7 Quantitative MR Biomarkers generated from Post Processing Textural Analysis- Tumour Assessment in the Prostate Gland

CONTRIBUTION:

Primary author responsible for data acquisition/analysis (though was assisted in textural analysis application and interpretation), drafting, revision, final approval and accuracy/integrity.

7.1 RUNNING TITLE

Textural analysis of multiparametric MRI detects transition zone prostate cancer.

7.2 INTRODUCTION

Management of early prostate cancer has been revolutionised by the use of multi-parametric MRI (mpMRI; using T2 and T1 weighted, diffusion weighted and contrast enhanced imaging) [1]. Nonetheless, transition zone (TZ) tumors remain more difficult to appreciate on mpMRI studies [2]; with reported sensitivity/specificity for detection of 0.53/0.83 compared with 0.80/0.97 respectively for peripheral zone (PZ) tumors [1]. TZ tumor signal homogeneity has been advocated as a discriminator of significant grade disease at mpMRI and has been incorporated into the recently revised ‘Pi-RADS 2’ guidelines [3]; endorsed by the European Society of Urogenital Radiology and the American College of Radiology.

Textural analysis is an image-processing technique that can assess image signal heterogeneity (both at and beyond that appreciated by the human eye) by quantifying the coarseness and regularity of the spatial distribution of pixel grey level values within normal and pathological tissue. Macroscopic heterogeneity in medical images may reflect microscopic heterogeneity at the histopathological level, particularly in oncological imaging with recent demonstrations of utility in tumor detection/grading, prognosis and treatment response [4, 5]. Compared to CT, MRI offers the advantages of improved soft tissue contrast resolution and of a wealth of imaging data afforded by a multi-parametric approach. Recent studies have used MR textural analysis (MRTA) for lesion detection, classification, treatment response-evaluation and prediction for example in breast, brain, and rectal cancer [6-8]. A number of approaches to texture analysis exist, with one approach being quantification of features through histogram analysis [9].

It is possible the additional tissue heterogeneity data provided by textural analysis could augment the diagnostic accuracy of radiologists in detecting TZ tumors, where lesions tend to be subtle and difficult to differentiate from adjacent benign nodular tissue. In this study, image analysis was performed on a whole TZ basis with the aim of obviating the need for radiological pre-identification of tumor thus increasing potential utility in computer-aided diagnosis (CAD) and to minimise inherent difficulties of small lesion contouring. The purpose of this study was to evaluate multiparametric MRI (mpMRI) derived histogram textural analysis parameters [9] for detection of transition zone (TZ) prostatic tumor.

7.3 MATERIALS AND METHODS

Our local institutional review board approved the study and waived the requirement for individual consent for this retrospective study of consecutive patient data acquired as part of routine clinical care (R&D No: 12/0195).

7.3.1 Patient cohort

The cohort comprised men with clinically suspected prostate cancer undergoing prostatic mpMRI prior to ‘20 zone’ template prostate mapping (TPM) biopsies (within 12 months) between 1st January 2010 to 31st December 2012 (n=210). Men who had peripheral zone tumor, undergone biopsy within 6 months prior to mpMRI, received previous treatment for prostate cancer, had metallic hip prostheses, or had incomplete mpMRI and/or TPM data sets were excluded (n=143). In total sixty-seven men with mean age 63.4 years (45–80 years), mean PSA 9.2 ng/ml (0.2–39.0 ng/ml) and mean gland volume of 42.9 ml (15–101 ml) were accrued. Of these 29/67 (43%) had no cancer, 26/67 (39%) had ‘significant’ TZ cancer and 12/67 (18%) had ‘insignificant’ TZ cancer (see below). Table 1 summarizes cohort demographics.

Cancer significance	**Number (%)**	**Mean Age years (range)**	**Mean PSA ng/dL (range)**	**Mean prostate volume ml (range)**	**Median time interval mpMRI to TPM days (range)**	**Mean area TZ ROI cm^2 (Total; range)**	**Mean area TZ tumor cm^2 (Total; range)**	**Percentage TZ tumor area/TZ area (Total; range)**
Benign OR Insignificant (<4mm	41 (61%)	64 (45-79)	8.6 (0.2-39)	47.0 (20-101)	54 (6-214)	40.0 (n=41; 10.0-	0.76 (n=12; 0.52-	2.4 (n=12; 1.1-3.7)

MCCL AND ≤G3+3)						68.9)	1.22)	
Significant (≥4mm MCCL OR ≥G3+4)	26 (39%)	63 (52-80)	9.9 (0.3-35)	37.2 (15-78)	56 (16-145)	29.9 (n=26; 9.3-61.2)	1.45 (n=26; 0.65-3.80)	5.3 (n=26; 1.7- 14.0)

Table 7.1 Summary of demographic and ROI areas for recruited patients categorised by benign/insignificant transition (TZ) pathology and significant TZ tumor (MCCL= maximum cancer core length, G= Gleason grade, PSA= prostate specific antigen serum concentration, TPM= template mapping biopsy).

7.3.2 Multi-parametric magnetic resonance imaging

Subjects underwent 1.5T magnet mpMRI (Avanto, Siemens, Erlangen, Germany) with pelvic-phased array coil, following intravenous spasmolytic (Buscopan; Boehringer Ingelheim, Germany) 0.2 mg/kg (maximum 20mg) to minimize bowel peristalsis. Full mpMRI parameters are given in Table 2. Figure 1 shows an example mpMRI with significant TZ tumor.

	T2w TSE axial/coronal	**EPI DWI**	**T1 3D FLASH***
Repetition time (ms)	5170 / 5240	2100	5.61
Echo time (ms)	92 / 104	98	2.5
Flip angle (degrees)	180 / 150	90	15
Echo train length	22 / 24	172	n/a
Bandwidth (Hz/Px)	190 / 190	968	300
Field of view (mm)	180 / 180	260	260
Phase FoV %	100 / 100	100	100
Slice thickness (mm)	3 / 3	5	3
Slice gap (mm)	0.3 / 0.3	0	0.6
Averages	2 / 2	16	1
Phase encoding direction	A>P / R>L	A>p	A>P
Fat saturation	No / No	Yes	Yes
Base matrix	256 / 256	172	192
Matrix phase %	95 / 95	100	100
b-values (s.mm^{-2})	n/a	0, 300, 500, 1000	n/a
Number of acquisitions	1 / 1	1	35
Temporal resolution (s)	n/a	n/a	16
Total acquisition time (min)	3m54s / 4m18s	3m39s	10m

Table 7.2 Multi-parametric MRI sequence parameters used for study; *dynamic contrast enhanced MRI – 0.2 ml/kg intravenous gadolinium contrast agent injected at the beginning of acquisition 6 at 3 ml/s followed by a saline flush of 20 ml; T2w TSE – T2 weighted turbo spin echo; EPI-DWI – echo planar imaging - diffusion weighted imaging; FLASH – fast low angle shot. Note coronal T2 acquisition was not used for image analysis (though remains part of the clinical scan). Apparent diffusion co-efficient (ADC) map automatically generated from the four b-values (mono-exponential). 'Early post contrast T1' image refers to the second image temporally from the point at which contrast first appears in the prostate gland (imaging every 16 seconds).

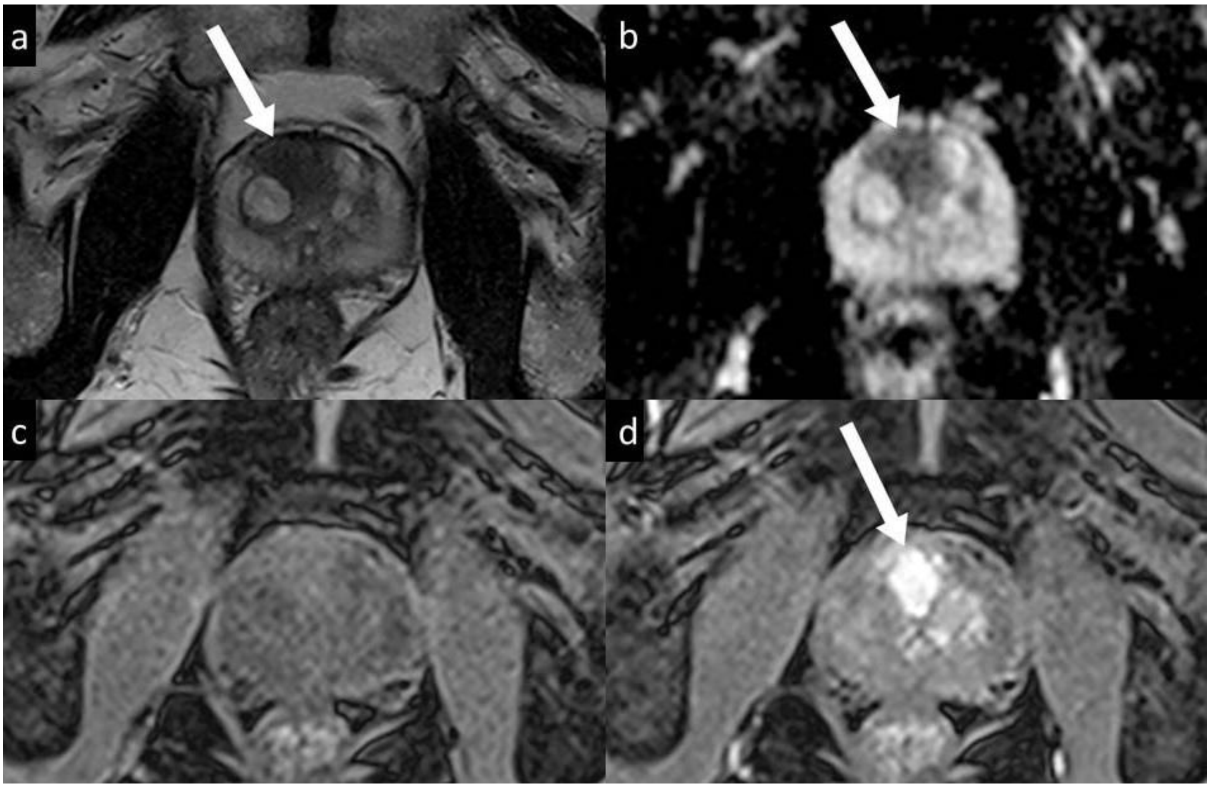


Figure 7-1 Demonstrates single slice axial images of significant tumor (arrow): (a) T2 weighted; (b) ADC map; (c) pre-contrast T1 weighted; and (d) early post contrast T1 weighted images in a 74 year old patient with anterior right transition zone tumor (Gleason 3+4; maximum cancer core length 11mm) prior to transition zone contouring and histogram analysis

7.3.3 Transperineal template-prostate-mapping biopsy

TPM followed mpMRI with median interval of 56 days (2 to 214 days); method as previously described [10, 11]. In brief, systematic biopsy of the whole gland was performed through a brachytherapy template-grid and 5-mm sampling frame, giving a uniform sampling density of approximately 1 core/cc of prostatic tissue. Biopsies were grouped and potted into 20 zones, modified from the technique reported by Barzell et al. [12].

7.3.4 Histopathology Review

Subjects were grouped according to a previously used definition of cancer significance [13] whereby TPM maximum cancer core length (MCCL) values were used to infer volume, following demonstration that a TPM MCCL of ≥4mm approximates to a tumor volume of ≥0.2 ml [14].

Clinically significant disease was defined as ≥ Gleason 3+4 OR ≥4mm MCCL; while other disease (i.e. 'low risk' <4mm MCCL AND ≤Gleason 3+3) was classified as non-significant and grouped with patients demonstrating benign TZ histology [15].

7.3.5 Histology-MRI Matching

Two radiologists in consensus ('X', 'Y'; with 8 years' and 4 years' experience respectively of mpMRI interpretation), aware of histopathological findings, reviewed each dataset using Osirix (version 3.5.1; Geneva, Switzerland) and matched the single axial slice depicting the TZ focus most suspicious for disease to location of disease confirmed according to histopathology. If no tumor was present, the radiologists selected the slice with the largest TZ anteroposterior diameter. Where biopsy-positive significant tumor was present, the radiologists also contoured significant tumor on each ADC, T2 and T1 weighted TZ single slice image to analyse the effect of initially including and then excluding significant tumor from the single-slice whole TZ ROI drawn by third radiologist (below). TZ tumor contoured on T2 images was used for tumor area estimation (relative to total TZ area).

A third radiologist ('Z'; with 2 years' experience of mpMRI interpretation), blinded to biopsy data and earlier tumor region of interest (ROI) placement, contoured solely the entire TZ on the selected slice, on matched ADC, T2 and early post-contrast T1 weighted images for each patient for subsequent MRTA.

The percentage of TZ replaced by tumor was quantified by [(tumor area / TZ area)*100].

7.3.6 MR Textural Analysis (MRTA)

The ADC map, T2 and early post-contrast T1 weighted segmented TZ regions from the selected slice (containing and initially including tumor where present) underwent MRTA ('A'; with 9 years of experience in texture analysis) using proprietary TexRAD research software (version 3.3, TexRAD Ltd, Feedback Plc, Cambridge UK) with automated texture parameter extraction. A subsequent separate analysis examined single-slice whole TZ ROIs excluding significant tumor to assess the effect on textural parameters.

MRTA comprised image histogram analysis to quantify first-order statistics of entropy, skewness, kurtosis of the TZ ROIs. These parameters reflect, to varying degrees, the number, intensity and variability of areas of high and low signal intensity within the TZ [9]. Absolute T2 and T1 weighted signal intensities are not comparable across patients without standardisation, and therefore T1 and T2 mean pixel, mean positive pixel and standard deviation values were not analysed further (unlike mean ADC). By contrast, measurement of entropy, kurtosis and skewness rely on the shape of the histogram i.e. relationships between pixel intensities and not on the absolute pixel intensity values.

7.3.7 Statistical Analysis

For each individual textural variable, the significance of difference between non-significant/benign TZ and significant tumor containing TZ were assessed using the two tailed Mann Whitney U test (statistical significance assigned at $p<0.05$). These analyses were repeated after exclusion of significant tumor area from the single-slice whole TZ ROI to determine if the observed differences in textural parameters could be directly attributed to inclusion of significant tumor within the analyzed TZ ROI. Receiver-operating characteristics (ROC) analyses characterised the performance of TZ textural features extracted from each of ADC, T1 and T2 weighted TZ ROI to predict significant TZ prostate cancer.

The area under the ROC curve (ROC AUC) and 95% confidence intervals for each parameter identified the best performing individual parameters and in combination using multivariate ROC-AUC analysis [16]. Leave-one-out (LOO) analysis [17] validated classification performance for best performing univariate and combined parameters.

All statistical analyses were performed using SPSS statistics for Windows (version 16; IBM, Armonk, NY) and MedCalc for Windows (version 9.2.0.0; MedCalc software, Mariakerke, Belgium).

7.4 RESULTS

7.4.1 Patient cohort

Mean tumor area was 1.45cm^2 (0.65 to 3.8) for patients with significant tumor at histology. Mean TZ area was 42.4cm^2 (10.0 to 68.9) for patients with benign histology; 33.9cm^2 (18.7 to 58.8) for patients with non-significant cancer, and 29.9cm^2 (9.3 to 61.2) for patients with significant cancer. Mean proportion of TZ replaced by tumor was 5.3% (range 1.7 to 14.0%) for patients with significant tumor (Table 1). Table 3 denotes the median values and interquartile ranges of whole TZ MR textural parameters for patients with non-significant/benign and for patients with significant tumor both including and excluding the significant tumor. P-values denote the statistical significance of differences between whole TZ for patients with non-significant/benign histology and patients with significant tumor for TZ ROIs (again including and excluding the significant tumor). The ROC-AUC for whole TZ textural parameters for each of the three MRI sequences, for classification of a TZ image as containing significant tumor, are also detailed.

Cohort/ Sequence	**Non-significant/ benign TZ** median (interquartile range)	**Significant cancer TZ** *ROI includes tumour* median (interquartile range)	**Significant cancer TZ** *ROI excludes tumour* median (interquartile range)	**p-value** *ROI includes significant tumour*	**p-value** *ROI excludes significant tumour*	**Receiver operating characteristics** area under curve (95% confidence intervals)
Kurtosis						
ADC	**0.09 (-0.14 to 0.59)**	**-0.51 (-0.69 to 0.01)**	**0.10 (-0.31 to 0.45)**	**<0.001**	**0.23**	**0.80 (0.69;0.91)**
Early T1	-0.12 (-0.34 to 0.26)	-0.05 (-0.56 to 0.40)	0.06 (-0.47 to 0.37)	0.96	0.84	0.50 (0.35;0.65)
T2	1.51 (0.65 to 2.73)	1.18 (0.40 to 2.86)	1.00 (0.33 to 2.50)	0.78	0.35	0.52 (0.37;0.67)
Entropy						
ADC	5.90 (5.68 to 6.06)	5.67 (5.42 to 5.83)	5.57 (5.38 to 5.91)	0.005	<0.001	0.69 (0.58;0.84)
Early T1	**5.51 (5.41 to 5.62)**	**5.37 (5.09-5.54)**	**5.32 (5.12 to 5.53)**	**0.004**	**0.004**	**0.70 (0.57;0.84)**
T2	**5.12 (4.94 to 5.24)**	**5.04 (4.82 to 5.14)**	**5.02 (4.90 to 5.18)**	**0.13**	**0.34**	**0.61 (0.47;0.75)**
Skewness						
ADC	-0.04 (-0.31 to 0.20)	0.00 (-0.13 to 0.25)	-0.12 (-0.44 to 0.05)	0.31	0.27	0.58 (0.44;0.71)
Early T1	0.16 (-0.24 to 0.31)	0.08 (-0.08 to 0.45)	0.25 (0.01 to 0.46)	0.49	0.06	0.55 (0.41;0.70)
T2	0.60 (0.30 to 1.00)	0.74 (0.57 to 1.03)	0.72 (0.50 to 1.13)	0.35	0.47	0.57 (0.43;0.71)

Mean						
ADC	0.99 (0.88 to 1.00)	0.80 (0.77 to 1.00)	0.92 (0.80 to 1.05)	0.004	0.13	0.71 (0.58;0.85)
Bivariate Model						
ADC Kurtosis + T1 Entropy						**0.86 (0.77;0.95)**

Table 7.3 Median values (and interquartile range 25%-75%) for ADC map, early post-contrast T1, and T2 weighted TZ ROIs for kurtosis, entropy and skewness derived from first order histogram analysis both including and excluding significant tumor where present. P-values have been calculated using two-tailed Mann Whitney U test. Those values in bold indicate most significant difference in values (i.e. between non-significant/benign and significant tumor TZ) for each sequence where applicable. Right hand column compares receiver operator characteristics (ROC) area under curve values for differentiating TZ ROIs containing significant TZ tumor from non-significant/benign histology for the MRTA parameters (values in parentheses indicate 95% confidence intervals).

7.4.2 TZ textural metrics

All analyses were performed on a whole TZ basis (initially incorporating significant tumor where present). Mean and median TZ ADC were significantly higher (0.99 and 0.97 x 10-3 mm2/s respectively; range 0.67 to 1.77) for patients with non-significant/benign histology than those with significant tumor (0.79 and 0.77 x 10-3 mm2/s respectively; range 0.07 to 1.22) (p=0.004).

Kurtosis

Median ADC, T1 and T2 kurtosis were 0.09, -0.12 and 1.51 respectively for non-significant/benign histology versus -0.51, -0.05 and 1.18 respectively for TZ containing significant tumor. Median ADC kurtosis was significantly lower for patients with significant tumor (p<0.001); whereas median T1 and T2 kurtosis were not significantly different between the groups (p=0.96 and 0.78 respectively).

Entropy

Median ADC, T1 and T2 entropy were 5.90, 5.51 and 5.12 respectively for non-significant/benign histology and 5.67, 5.37 and 5.04 respectively for TZ containing significant tumor. Median ADC and T1 entropy were lower for patients with significant tumor (p=0.005 and 0.004 respectively) whilst T2 entropy did not reach significance (p=0.13).

Skewness

Median ADC, T1 and T2 skewness were -0.04, 0.16 and 0.60 respectively for non-significant/benign histology versus 0.00, 0.08 and 0.74 respectively for TZ containing significant tumor. There were no significant differences between the two groups (p=0.31 to 0.49).

7.4.3 Effect of excluding significant tumor on whole TZ textural metrics

Following exclusion of significant tumor from the single-slice whole TZ ROI, there was no longer a significant difference between median ADC value (p=0.13) when patients with those significant tumors were compared with those with non-significant/benign histology.

Kurtosis

Following exclusion of significant tumor from the single-slice whole TZ ROI, there was no longer a significant difference between median ADC kurtosis when patients with significant tumor were compared with those with non-significant/benign histology (p=0.23). Median T1 and T2 kurtosis differences remained non-significant (p=0.84 and 0.35 respectively; Table 3).

Entropy

Significant differences between T1 and ADC entropy remained when patients with significant tumor and those with non-significant/benign histology were compared following exclusion of significant tumor from the former ROI (p <0.01; Table 3).

Skewness

Median ADC, T1 and T2 skewness did not demonstrate any consistent difference between patients with significant tumor and those with non-significant/benign histology following exclusion of significant tumor from the TZ ROI.

Box-and-whiskers plot of the best performing textural parameters are illustrated in Figure 2.

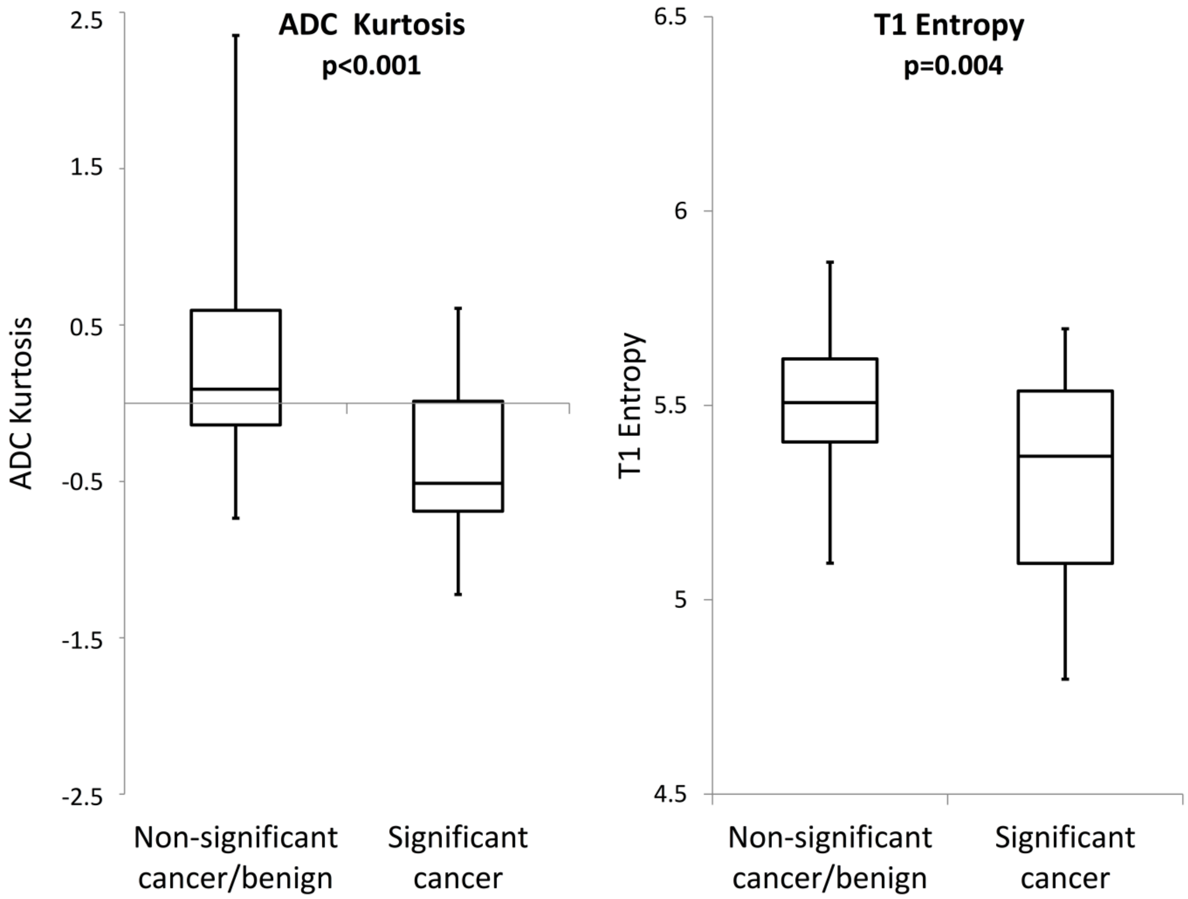


Figure 7-2 Box plots showing best performing textural discriminators of TZ ROIs containing significant and non-significant using ADC kurtosis and early post-contrast T1. In each box plot- box indicates interquartile range; line indicates median and whiskers indicate most deviated data points/range. Two tailed Mann Whitney U p-values are also given for each parameter.

7.4.4 Disease classification by univariate textural metrics

The best performing classifier for ADC was kurtosis (0.80; 95% CI 0.69 to 0.91). Entropy yielded the highest AUC for early T1 post-contrast image (0.70; 95% CI 0.57 to 0.84). This analysis did not show significant differences on T2 weighted whole TZ textural analysis with the best performing T2 parameter being T2 entropy (AUC of 0.61; 95% CI 0.47 to 0.75). LOO validation demonstrated ROC-AUC 0.78 (95% CI 0.66 to 0.90) for ADC kurtosis and 0.66 (95% CI 0.52 to 0.80) for T1 entropy.

7.4.5 Disease classification by multivariate textural metrics

ROC-AUC for the best performing significant univariate parameters and bivariate combination of these parameters are shown in Figure 3. The two best performing parameters for ADC (kurtosis) and T1 (entropy) combined in bivariate model gave ROC-AUC of 0.86 (95% CI 0.77 to 0.95). LOO analysis of this bivariate model yielded ROC-AUC 0.83 (95% CI 0.74 to 0.93) and is shown in Figure 4.

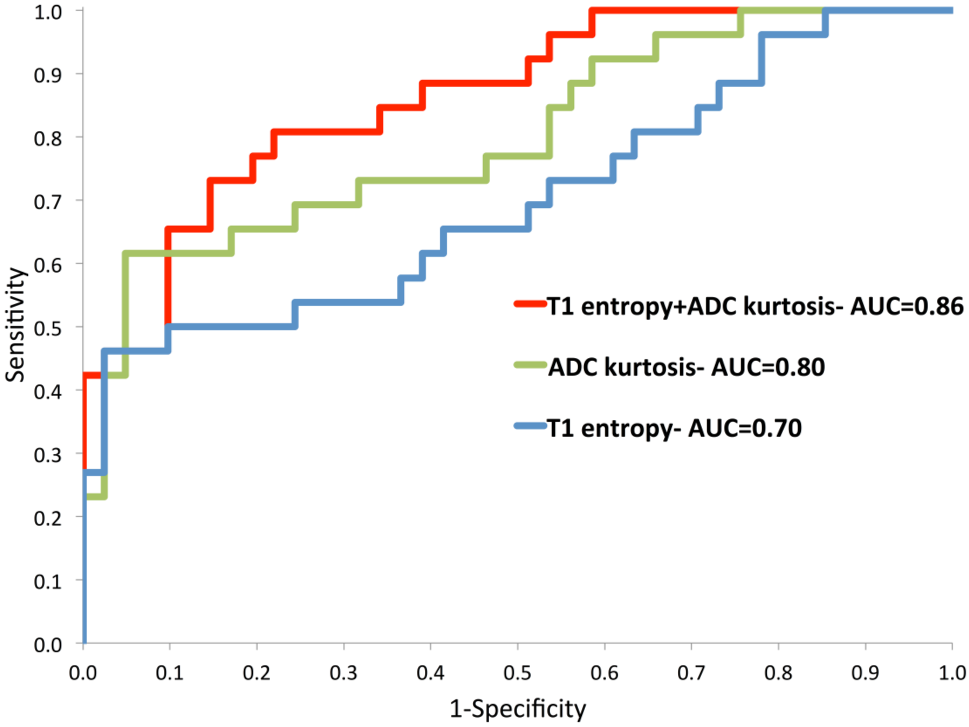


Figure 7-3 Receiver operating characteristic (ROC) curves of the two best performing textural features and bivariate combination for discrimination of transition zone ROIs containing significant prostatic tumor from non-significant TZ with area under curve (AUC) values as shown.

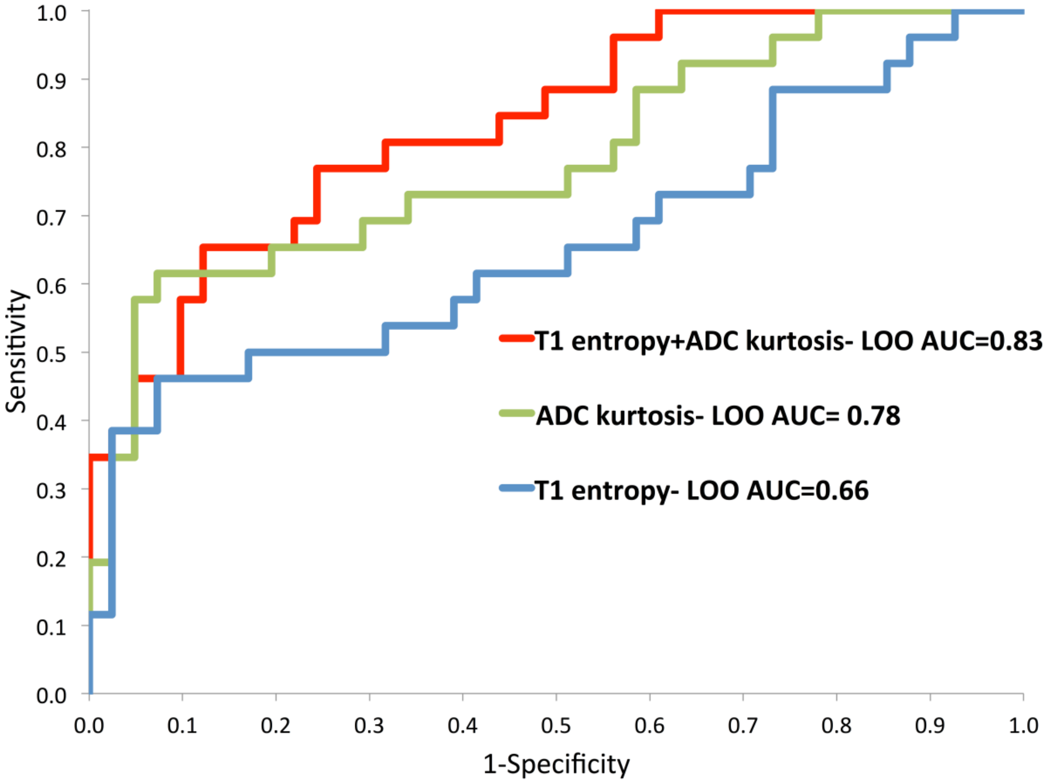


Figure 7-4 Receiver operating characteristic (ROC) curves of the two best performing textural features and bivariate combination for discrimination of TZ ROIs containing significant prostatic tumor from non-significant TZ ROIs after leave one out (LOO) analysis with area under curve (AUC) values as shown.

7.5 DISCUSSION

This study evaluated the diagnostic accuracy of textural parameters, derived from clinical prostate mpMRI, for detection of TZ cancer. Previous work has confirmed that quantitative mpMRI parameters (e.g. ADC) can differ between benign and cancerous TZ regions [18]. Unlike previous studies, here we derived single-slice whole TZ textural parameters (including cancer pixels where cancer was present) and evaluated differences in the histographic pixel distribution between patients with and without significant cancer.

We found that textural features from an image of the entire TZ are altered significantly when containing even a small proportion of significant cancer which no longer holds true (for best performing textural parameter) when the same tumor is excluded from the analysis. Overall, classification of TZ tumor-containing slices by best performing single textural parameter and/or bivariate combination (ROC-AUC 0.80 to 0.86) was comparable with previously reported visual detection of TZ tumor by radiologists (ROC-AUC 0.73 to 0.84) [19, 20].

Kurtosis is a measure of histogram "peakedness". Positive kurtosis indicates a more peaked distribution of pixel signal intensities. We found reduced ADC kurtosis was the best univariate classifying textural feature (ROC-AUC 0.80 on ADC images) on a whole TZ basis and demonstrated a higher ROC-AUC than the non-textural parameter of ADC mean (ROC-AUC 0.71 on ADC images). A larger cohort is required to test the statistical significance of this difference. We expect textural measures, based on the relationship of pixels in a given ROI rather than absolute pixel intensities, are likely more robust to variations between individuals and scanners. Whilst kurtosis measures from T2 and early post-contrast T1 weighted images performed less well for detection of cancer. In spite of the relatively modest ratio of significant tumor compared to remaining TZ area (~5%), we have found that exclusion of this radiologically visible tumor from the single-slice whole TZ ROI results in loss of significance in the difference in ADC kurtosis between the two groups. We posit this to be the result of elimination of a second differing signal intensity ‘population’ (i.e. low ADC significant tumor) from the TZ ROI resulting in normalisation of kurtosis when compared to non-significant/benign cohort reflected in more ‘peaked’ values. Wibner et al [21] have previously demonstrated the utility of ADC derived textural parameters in TZ tumor detection and differentiating grade based on lesion analysis. The textural analysis approach may obviate previously described efforts to normalise by using ADC ratios (normalized to non-tumorous tissues) [22], perform whole lesion measurement and/or generate histogram analysis including median and low percentile ADC values [23].

Entropy is a measure of image 'busyness/irregularity' and as such, the observation of significantly lower mean TZ entropy values in patients with significant cancer would reflect increased overall signal homogeneity. Post-contrast T1 weighted image entropy was the second best univariate classifier, in keeping with previously observed homogenous enhancement features of TZ tumor [20]. However, the observation that the significant difference in T1 entropy between TZ containing significant tumor and non-significant/benign TZ persists even after the exclusion of radiologically visible tumor in the former group reflects more homogeneous enhancement throughout the whole TZ in these cases. The mechanism for this is unclear, though growing evidence implicates chronic inflammation as a contributor to prostate cancer development and progression [24] and present in benign prostate tissue associated with high-grade prostate cancer [25]. The finding of more homogenous enhancement in the 'non-malignant' parts of TZ harbouring significant tumor may reflect these aspects of prostate cancer ontogeny. ADC entropy was also significantly lower in single-slice TZ ROIs containing tumor and this relationship persisted following removal of cancer from the TZ ROI possibly due to the same/similar mechanism. Additionally at the time these data were acquired, as demonstrated in Table 2, the lowest b-value used for calculation of the ADC map was 0sec/mm2 and it is possible the findings of ADC entropy may at least in part reflect a perfusion difference component. These findings may not have been appreciated in prior quantitative imaging studies, most of which concentrate on assessing the vascular properties of tumors themselves, and further confirmatory (histologically correlative) studies are required.

Several T2 weighted features aid visual identification of TZ tumor; including homogeneously reduced lesion signal intensity, ill-defined lesion margins and lenticular lesion shape [19]. However, on a whole TZ basis, the current study does not confirm that quantitatively reduced whole TZ entropy (increased homogeneity) on T2 weighted images is a particularly good discriminator. This may relate to the relatively small area of tumor compared with background non-tumorous TZ and/or relate to the study size. Additionally, ADC and early post-contrast T1 images reflect processes occurring on a ‘microscopic’ level and can reflect field changes (for example subtle inflammatory change) that may explain the significance of differences between benign TZ and that harboring significant tumor. Whilst several studies have described homogenous reduced tumoral T2 signal compared to benign TZ, the remaining (and predominant approximately 95% by area) ‘normal’ TZ glandular tissue is variably heterogenous across patients and images are more reflective of macroscopic differences and subtle changes in T2 signal induced by background inflammation may be masked.

Skewness is a measure of histogram asymmetry; a zero value indicates a symmetrical distribution around the mean. Skewness was a relatively poor classifier of TZ cancer when applied to any image sequence. It has been speculated [23] that in prostate tumors densely packed with malignant cells, the resulting histograms from tumor ROIs are likely to be less skewed compared with tumors with more heterogeneous cellular density. However, we evaluated skewness for all pixels within the whole TZ and the proportion of tumor pixels may have been insufficient to influence skewness significantly particularly given such variation in cellular density is not captured by traditional histological grading [26].

To our knowledge, this study is distinct from other work as it assessed in-vivo textural features derived from each mpMRI sequence for the entire TZ. The study was conducted as a proof of concept for the purpose of detecting significant TZ tumor. Future work will examine the translation of this concept over whole TZ volume to examine utility in augmenting visual radiologist assessment (e.g. added value of prompting readers to areas for examination which may contain significant cancer) and/or automated detection. Interpretation of images by radiologists is complex [27] although an assessment of texture is made, to date this visual textural assessment has not been characterised objectively. We believe that textural analysis of each multiparametric image is analogous to the manner by which radiologists visually localise TZ cancer on mpMRI [4]. Other workers [e.g. 28, 29] have examined first and second order (e.g. two-dimension grey-level co-occurrence matrix) textural features from one or more mpMRI sequences from ROIs drawn around individual TZ tumors. Such approaches, while informative, have less clinical application since they are more computationally intensive and require the radiologist to first identify areas of concern. Furthermore, where textural features of small lesions are evaluated at larger spatial scales, these features can be biased significantly by region boundary. In contrast, assessment of the entire TZ texture requires less development as a clinical diagnostic tool (i.e. only segmentation of the TZ from the PZ which may be an automated/semi-automated procedure [30]) and minimizes boundary related effects.

Our study has some limitations. Most foremost the use of TexRAD software was based in part upon institutional links and as such is not free or open source (nor IBSI compliant) and as such we cannot propose this as the best option for textural analysis compared with other platforms. Another issue is the lack of independent validation (rather than LOO), such that we cannot be sure the inferences are not based upon overfitting/analysis. All mpMRI images were acquired on a 1.5T scanner and generalizability to 3T platforms was not assessed. However, our 1.5T mpMRI protocol was in keeping with recommendations from a European Consensus meeting [31]. We did not use endorectal coils, and note that the European Consensus Meeting failed to agree regarding their necessity even at 1.5T. We routinely perform DCE MRI using a higher spatial resolution and lower temporal resolution than some other centres as advocated by recent guidelines [3, 32]. Therefore, we expect the generalizability of post-contrast T1 weighted textural feature findings to be limited to centres that similarly maintain higher spatial resolution. We did not have access to radical prostatectomy specimens as a reference ground-truth. However, employment of a TPM based reference standard [33], which is a significant improvement over conventional TRUS biopsy [14], can help avoid the spectrum bias towards more severe cases that occurs with studies when prostatectomy specimens are mandated. We acknowledge that there could be some error when registering TPM histopathology to mpMRI. However, it should be noted that prostatectomy specimens themselves are not free from registration errors induced by processing (shrinkage, distortion, and tissue-loss) [34]. Lastly we manually contoured the single-slice TZ ROIs though we do not expect this to unduly affect the results and envisage automated segmentation could be used as available.

Our results have clinical relevance, and confirm textural features may inform CAD software to highlight images on which significant tumor could be present within the TZ. In developing this whole TZ approach using MRTA as an imaging biomarker, it is recognised that a series of studies would be required to further validate the findings as per various ‘roadmap’ approaches in contemporaneous guidelines [35]. This approach could be adapted to several scenarios, subject to demonstration of efficacy, for example by drawing attention to slice(s) which may require further examination for radiologists, probability of significant tumor being present across entire TZ and predictive information if there are relevant ‘field’ changes across the TZ. The United Kingdom NICE guidelines have adopted mpMRI for detection of prostate cancer in patients with a negative non-targeted TRUS biopsy [36]. Our textural evaluation technique may have particular relevance to such patients who are more likely to have TZ tumors that are systematically undersampled by TRUS [34].

7.6 REFERENCES

1. Delongchamps NB, Rouanne M, Flam T, Beuvon F, Liberatore M, Zerbib M, Cornud F. Multiparametric magnetic resonance imaging for the detection and localization of prostate cancer: combination of T2-weighted, dynamic contrast-enhanced and diffusion-weighted imaging. BJU Int. 2011; 107(9):1411-811.

2. Langer DL, van der Kwast TH, Evans AJ, Trachtenberg J, Wilson BC, Haider MA. Prostate cancer detection with multiparametric MRI: logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI. J Magn Reson Imaging 2009; 30:327–334.

3. Barentsz JO, Richenberg J, Clements R, Choyke P, Verma S, Villeirs G, Rouviere O, Logager V, Fütterer JJ. ESUR prostate MR guidelines 2012. Eur Radiol. 2012; 22(4):746-57.

4. Davnall F, Yip CS, Ljungqvist G, Selmi M, Ng F, Sanghera B, Ganeshan B, Miles KA, Cook GJ, Goh V. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? Insights Imaging 2012; 3:573-89.

5. Ganeshan B, Miles KA. Quantifying tumour heterogeneity with CT. Cancer Imaging 2013; 13:140-9.

6. Eliat PA, Olivie D, Saikali S, Carsin B, Saint-Jalmes H, de Certaines JD. Can dynamic contrast-enhanced magnetic resonance imaging combined with texture analysis differentiate malignant glioneuronal tumors from other glioblastoma? Neurol Res Int. 2012; 195176.

7. Parikh J, Selmi M, Charles-Edwards G, Glendenning J, Ganeshan B, Verma H, Mansi J, Harries M, Tutt A, Goh V. Changes in primary breast cancer heterogeneity may augment midtreatment MR imaging assessment of response to neoadjuvant chemotherapy. Radiology 2014; 272(1):100-12.

8. De Cecco CN, Ganeshan B, Ciolina M, Rengo M, Meinel FG, Musio D, De Felice F, Raffetto N, Tombolini V, Laghi A. Texture analysis as imaging biomarker of tumoral response to

neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3-T magnetic resonance. Invest Radiol. 2015 Apr;50(4):239-45.

9. Miles KA, Ganeshan B, Hayball MP. CT texture analysis using the filtration-histogram method: what do the measurements mean? Cancer Imaging 2013 Sep 23; 13(3):400-6.

10. Onik G, Barzell W. Transperineal 3D mapping biopsy of the prostate: an essential tool in selecting patients for focal prostate cancer therapy. Urol Oncol 2008; 26:506–510.

11. Taira AV, Merrick GS, Galbreath RW et al. Performance of transperineal template-guided mapping biopsy in detecting prostate cancer in the initial and repeat biopsy setting. Prostate Cancer Prostatic Dis 2010; 13:71–77.

12. Barzell WE, Melamed MR. Appropriate patient selection in the focal treatment of prostate cancer: the role of transperineal 3-dimensional pathologic mapping of the prostate—a 4-year experience. Urology 2007; 70:27–35.

13. Lecornet E, Ahmed HU, Hu Y, Moore CM, Nevoux P, Barratt D, Hawkes D, Villers A, Emberton M. The accuracy of different biopsy strategies for the detection of clinically important prostate cancer: a computer simulation. J Urol. 2012; 188(3):974-80.

14. Ahmed HU, Hu Y, Carter T et al. Characterizing clinically significant prostate cancer using template prostate mapping biopsy. J Urol. 2011; 186:458–464.

15. Stamey TA, Freiha FS, McNeal JE et al: Localized prostate cancer. Relationship of tumor volume to clinical significance for treatment of prostate cancer. Cancer 1993; 71:933-938.

16. Hanley JA, McNeil BJ. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology 1982; 143:29-36.

17. Fukunaga K, Hummels DM. Leave-one-out procedures for nonparametric error estimates. IEEE Trans Pattern Anal Mach Intell 1989; 11:421–423.

18. Dikaios N, Alkalbani J, Sidhu HS, Fujiwara T, Abd-Alazeez M, Kirkham A, Allen C, Ahmed H, Emberton M, Freeman A, Halligan S, Taylor S, Atkinson D, Punwani S. Logistic regression model for diagnosis of TZ prostate cancer on multi-parametric MRI. Eur Radiol. 2015; 25(2):523-32.

19. Akin O, Sala E, Moskowitz CS, et al. Transition zone prostate cancers: features, detection, localization, and staging at endorectal MR imaging. Radiology 2006; 239(3): 784–792.

20. Oto A, Kayhan A, Jiang Y, et al. Prostate cancer: differentiation of central gland cancer from benign prostatic hyperplasia by using diffusion-weighted and dynamic contrast-enhanced MR imaging. Radiology 2010; 257(3): 715–723.

21. Wibmer A, Hricak H, Gondo T, Matsumoto K, Veeraraghavan H, Fehr D, Zheng J, Goldman D, Moskowitz C, Fine SW, Reuter VE, Eastham J, Sala E, Vargas HA. Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. Eur Radiol. 2015 Oct;25(10):2840-50.
22. Coffey N, Schieda N, Cron G, Gulavita P, Mai KT, Flood TA (2014). Multi-parametric (mp) MRI of prostatic ductal adenocarcinoma. J Magn Reson Imaging. 2015; 41: 1639–1645.
23. Donati OF, Mazaheri Y, Afaq A et al (2014) Prostate cancer aggressiveness: assessment with whole-lesion histogram analysis of the apparent diffusion coefficient. Radiology 2014; 271:143–152.
24. Stark T, Livas L, Kyprianou N. Inflammation in prostate cancer progression and therapeutic targeting. Transl Androl Urol. 2015; 4(4):455-63.
25. Gurel B, Lucia MS, Thompson IM, Jr, et al. Chronic inflammation in benign prostate tissue is associated with high-grade prostate cancer in the placebo arm of the prostate cancer prevention trial. Cancer Epidemiol Biomarkers Prev 2014;23:847-56.
26. Langer DL, van der Kwast TH, Evans AJ, et al. Intermixed normal tissue within prostate cancer: effect on MR imaging measurements of apparent diffusion coefficient and T2—sparse versus dense cancers. Radiology 2008; 249(3):900–908.
27. Koontz NA1, Gunderman RB. Gestalt theory: implications for radiology education. AJR Am J Roentgenol. 2008; 190(5):1156-60.
28. Vignati A, Mazzetti S, Giannini V, Russo F, Bollito E, Porpiglia F, Stasi M, Regge D. Texture features on T2-weighted magnetic resonance imaging: new potential biomarkers for prostate cancer aggressiveness. Phys Med Biol. 2015; 60(7):2685-701.
29. Niaf E, Rouvière O, Mège-Lechevallier F, Bratan F, Lartizien C. Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. Phys Med Biol. 2012; 57(12):3833-51.
30. Garvey B, Türkbey B, Truong H, Bernardo M, Periaswamy S, Choyke PL. Clinical value of prostate segmentation and volume determination on MRI in benign prostatic hyperplasia. Diagn Interv Radiol. 2014; 20(3):229-33.
31. Dickinson L, Ahmed HU, Allen C et al. Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: recommendations from a European consensus meeting. Eur Urol. 2011; 59:477–494.

32. Kirkham AP, Haslam P, Keanie JY, McCafferty I, Padhani AR, Punwani S, Richenberg J, Rottenberg G, Sohaib A, Thompson P, Turnbull LW, Kurban L, Sahdev A, Clements R, Carey BM, Allen C. Prostate MRI: who, when, and how? Report from a UK consensus meeting. Clin Radiol. 2013;68(10):1016-23.

33. Ahmed HU, Emberton M, Kepner G, Kepner J. A biomedical engineering approach to mitigate the errors of prostate biopsy. Nature Reviews. Urology 2012; 9:227–23131.

34. Xu S, Kruecker J, Turkbey B et al. Real-time MRI-TRUS fusion for guidance of targeted prostate biopsies. Comput Aided Surg. 2008; 13:255–264.

35.http://www.cancerresearchuk.org/sites/default/files/imaging_biomarker_roadmap_for_cancer_studies.pdf

36. http://www.nice.org.uk/guidance/cg175/chapter/recommendations

SECTION D: THESIS SUMMARY, DISCUSSION AND CONCLUSIONS

8.1 AIMS:

In this thesis, I aimed to identify and follow the development of MR derived imaging biomarkers at a range of stages, as driven by clinical need, to exemplify the biomarker development roadmap for cancer studies and how it can be applied to oncological research imaging. I have focused upon a small number of biomarkers ranging from early stage novel uses of T1 mapping in cancer evaluation and response assessment from technical/analytical validation studies to early clinical qualification, through to more conventional establish biomarker use in head and neck cancer clinical qualification as well as more advanced modelling on these data to generate further biomarker information and then finally looking at techniques for using existing images to generate and model biomarkers for cancer evaluation.

8.2 CHAPTER 2

Aim: To analytically validate a clinically feasible whole body T1 mapping methodology for use in cancer assessment.

Outcome: The VFA-3D method demonstrated good accuracy when correlated with gold standard and surrogate IR T1 measures and precision performed in test-retest and temporally separate scans in a single imaging platform in a single centre.

Further work: There is a need to expand this validation to different platforms within the same centre and ultimately across centres to truly facilitate clinical utility in real world setting.

8.3 CHAPTER 3

Aim: To explore clinical qualification of T1 mapping in sarcoma and breast cancer- can it predict or assess early response to systemic chemotherapy.

Outcome: Prospective single centre study demonstrated that T1 changes could be used to assess early response to chemotherapy and was comparable to more established measures such as diffusion weighted imaging.

Further work: Larger scale trials, ideally multicentre, will be required in more homogenous cancer cohorts (i.e. larger series in a single tumour type receiving similar treatment) to confirm the findings. We also need to understand the biophysical changes underpinning the observed differences.

8.4 CHAPTER 4

Aim: To explore clinical qualification of T1 mapping in metastatic prostate cancer- can it predict or assess early response to systemic therapy.

Outcome: Prospective single centre study demonstrated that T1 changes could be used to assess response to systemic therapy in cancerous lymph nodes and was comparable to more established measures such as diffusion weighted imaging. Additionally baseline lymph node T1 values (but not ADC) seemed to be able to predict response before treatment was initiated. In bone deposits, T1 was the only biomarker which was able to discriminate response and actually outperformed size change in significance.

Further work: Larger scale trials, ideally multicentre, will be required to confirm the findings. We also need to understand the biophysical changes underpinning the observed differences.

8.5 CHAPTER 5

Aim: To explore clinical qualification of existing BOLD imaging biomarkers in head and neck cancer- can it predict response to chemoradiation.

Outcome: Prospective single centre study demonstrated cancerous lymph nodes in patients with Head and Neck squamous cell carcinoma (HNSCC) which subsequently respond to chemoradiotherapy (CRT) demonstrate significant lengthening of R2* relaxation times when switching from breathing air to 100%-oxygen, implying a paradoxical increase in deoxyhaemoglobin on breathing 100%-oxygen in nodes which respond favourably to CRT.

Further work: This finding is novel and counter to most clinical trials (albeit these sampled part of the node with invasive sampling). Larger scale trials, ideally multicentre, will be required to confirm the findings. We also need to understand the biophysical changes underpinning the observed differences.

8.6 CHAPTER 6

Aim: To explore clinical qualification of existing DWI imaging biomarker generated modelling in head and neck cancer- can pre-treatment nodal microstructural heterogeneity, reflected through diffusion MRI characteristics, classify diseased from normal nodes and/or predict local disease recurrence of patients with head and neck squamous cell carcinoma.

Outcome: Prospective single centre study demonstrated pre-treatment inter-voxel diffusion MRI heterogeneity in lymph nodes, assessed by ADC histogram analysis, can be used as disease identifier. Lymph nodes intra-voxel and inter-voxel MR diffusion heterogeneities assessed by the stretched exponential model didn’t classify normal from diseased nodes or predict treatment response in this study.

Further work: Future studies are required to further investigate the utility of the intra-voxel diffusion heterogeneity, derived by αvalue, as an imaging biomarker for disease classification and for treatment response.

8.7 CHAPTER 7

Aim: To evaluate multiparametric-MRI (mpMRI) derived histogram textural-analysis parameters for diagnosis detection of transition zone (TZ) prostatic tumour.

Outcome: Retrospective single centre study showed textural features of the whole prostate TZ can discriminate significant prostatic cancer through reduced kurtosis of the ADC-histogram where significant tumor is included in TZ-ROI and reduced T1 entropy independent of tumor inclusion.

Further work: Future studies are required to confirm findings in larger centre prospective settings. Whilst textural analysis seems useful, contrasting with other biomarkers, the mechanisms for the observed changes are difficult to intuit and will need to be explored.