Automated MRI-based radiotherapy planning for brain tumours

Emily Rose Johnstone

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds

School of Medicine

May 2019

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter Two is based on work from the jointly authored publication: *Systematic Review of Synthetic Computed Tomography Generation Methodologies for Use in Magnetic Resonance Imaging – Only Radiation Therapy*, Emily Johnstone, Jonathan Wyatt, Ann Henry, Susan Short, David Sebag-Montefiore, Louise Murray, Charles Kelly, Hazel McCallum and Richard Speight, International Journal of Radiation Oncology – Biology – Physics (2018), 100 (1), pp 199 – 217. Emily Johnstone conducted the systematic review search, collated and evaluated the literature, composed the submitted paper and responded and re-wrote the final accepted publication. Ann Henry and Richard Speight provided guidance on the review process. All co-authors reviewed the paper and provided constructive feedback on the written drafts and final paper, as well as Emily Johnstone's responses to reviewers.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Acknowledgements

First and foremost, I would like thank Prof Ann Henry and Dr Richard Speight for all their help and guidance over the course of this PhD. They have celebrated with me during the successes and supported me when things did not go as planned. I truly appreciate all of their mentorship over the last three and a half years.

I would also like to express my gratitude to Dr Louise Murray, Dr David Broadbent, Dr Daniel Wilson, Ms Sarah Bacon, Dr Animesh Saha and Prof Susan Short, who have given up a substantial amount of their time to help with the work which follows.

Thank you to the other members of my supervisory team: Prof David Sebag-Montefiore and Dr Bashar Al-Qaisieh. Their support throughout this PhD has been greatly appreciated.

I would like to acknowledge the help of our research collaborators, namely those at the Northern Centre for Cancer Research, Newcastle, UK; the Commonwealth Scientific and Industrial Research Organisation, Brisbane, Australia; Spectronic Medical AB, Helsingborg, Sweden and Leeds Tests Objects Ltd., Boroughbridge, UK. In particular, thanks to Mr Jonathan Wyatt and Dr Hazel McCallum, who I have enjoyed working closely with over the course of this PhD. I would additionally like to thank Dr Jason Dowling, who has not only provided an invaluable contribution to this work, but also made me feel welcome whilst I was working on the other side of the world.

I am extremely grateful to those in the radiotherapy physics department at the Leeds Cancer Centre, which is led by Dr Vivian Cosgrove. I especially appreciated the conversations I had with David Paynter and David Bird who were also going through the highs and lows of PhD life. I was very lucky to based with some wonderful physicists in my office, without whom I would not have enjoyed

the last few years as much as I did. I would especially like to acknowledge Samir Dawoud, who shared my fascination with aviation and Phil Rixham, who did not.

I would like to thank my family and friends for their encouragement over the course of this PhD, with particular thanks to Mum, Dad and Elinor. I always appreciate their support and guidance. Thomas and Alex also deserve recognition for providing so much enjoyment over the last few years. Finally, thank you to Drake who has spent the first years of his married life with a wife who was finishing her PhD. He has provided me with a bottomless pit of support and has never complained about it once.

This PhD is dedicated to Chloe. Although she was only here a short while, her memory lives on through those who loved her.

This research has been carried out by a team which has included Ann Henry (AH), Animesh Saha (AS), Bashar Al-Qaisieh (BAQ), Carl Siversson (CS), Charles Kelly (CK), David Broadbent (DB), David Sebag-Montefiore (DSM), David Smith (DS), Daniel Wilson (DW), Hazel McCallum (HM), Jason Dowling (JD), Jonathan Wyatt (JW), Louise Murray (LM), Richard Speight (RS), Sarah Bacon (SB), Sarah Wright (SW), Susan Short (SS) and Tim Clarkson (TC). My own contributions, fully and explicitly indicated in the thesis, have been outlined in the table below. The other members of the group and their contributions have also been specified in the table below.

	Work attributable to the candidate	**Work attributable to other team members**
Chapter 1	Research, write-up.	AH, RS and DB provided supervision.
Chapter 2	Study design, systematic review search, collation and evaluation of the literature, systematic review write-up. Updated literature review search and write-up.	AH and RS provided supervision. JW, AH, SS, DSM, LM, CK, HM and RS provided constructive feedback on the written review.
Chapter 3	Study design, image and contour preparation (including fusion and anonymisation), synthetic CT analysis, write-up.	AH and RS provided supervision. JD and CS created the models and produced the synthetic CT images.
Chapter 4	Study design, delineation guidelines, image preparation, automatic contour analysis, write-up.	AH, RS and LM provided supervision. JD created the models and produced the automatic contours. LM, SS and AS produced the manual contours. LM and SS provided expert advice on the delineation guidelines. LM classified

		the clinical acceptability of the automatic contours.
Chapter 5	Study design, phantom specification, phantom test-retest and monthly measurements at Leeds, analysis including statistical analysis, phantom susceptibility measurements and assessment, software validation assessment, write-up.	AH, RS and DB provided supervision. JW performed test-retest and monthly measurements at Newcastle. DB, DW and SB assisted with phantom measurements at Leeds. JW, DB, HM, DSM, SS, CK, LM, BAQ, AH and RS reviewed the temporal stability write-up.
Chapter 6	Study design, literature search, initial phantom design, MRI and CT measurements and analysis, scan segmentation and structure preparation, analysis and write-up.	AH and RS provided supervision. DB assisted with MRI measurements. BAQ, DS and SW assisted with CT measurements. RS, DB, SW, TC and BAQ provided guidance during the initial phantom design. The phantom was constructed by Leeds Test Objects Ltd., who were also represented at some meetings regarding the phantom specification and design.
Chapter 7	Write-up.	AH and RS provided supervision.

Abstract

Aims: To investigate the clinical feasibility of performing magnetic resonance imaging (MRI)-only radiotherapy planning for patients undergoing treatment for brain tumours.

Method: A systematic review was performed of methods which convert MRI scans into synthetic computed tomography (sCT) scans for the purposes of MRI-only radiotherapy planning. The feasibility of using two methods which had been developed at different centres for different anatomical sites, for the production of brain sCTs was assessed. The quality of automatically segmented brain structures generated using one of these techniques was determined. The temporal stability of geometric distortions of four MRI scanners over a year was investigated. A prototype of an MRI and CT compatible anthropomorphic head and neck phantom was developed for the purpose of quality assurance (QA).

Results: The review demonstrated that atlas-based and voxel-based techniques were clinically useful methods for sCT generation, with machine learning techniques beginning to develop. An atlas-based technique (previously developed for prostates) and a neural network technique (not previously tested for brains) were found to produce sCTs of acceptable quality for the majority of metrics, after modifications to the clinical MRI scanning protocol were introduced. Dosimetric deviations meant that further optimisation would be required. Larger brain structures were generally well segmented automatically, whilst smaller structures were segmented erratically. The stability of MRI geometric distortions varied between scanners and sequences. A prototype anthropomorphic head and neck phantom was designed and constructed, with future modifications needed in certain areas.

Conclusions: The feasibility of producing clinically acceptable brain sCTs along with automatically segmented structures, using models developed at different centres for different anatomical sites, has been shown. The variation in temporal stability of MRI geometric distortions means that careful consideration should be

paid to the frequency of distortion QA. An MRI and CT compatible head and neck phantom prototype has been developed.

Table of Contents

List of Tables

List of Figures

Abbreviations

AP	Anteroposterior
CBCT	Cone-beam computed tomography
CGD	Centre of gravity distance
CI	Conformity index
CNN	Convolutional neural network
CRT	Conformal radiotherapy
CSF	Cerebrospinal fluid
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CT	Computed tomography
DI	Distance-to-isocentre
DRR	Digitally reconstructed radiograph
DSC	Dice similarity coefficient
DTI	Diffusion tensor imaging
dUTE	Dual-echo ultra-short echo time
DVH	Dose-volume histogram
EPID	Electronic portal imaging devices
FFE	Fast field echo
FLAIR	Fluid attenuated inversion recovery
FLASH	Fast low angle shot
fMRI	Functional magnetic resonance imaging
FOV	Field of view
GAN	Generative adversarial network
GTV	Gross tumour volume
HU	Hounsfield unit
ICRU	International Commission on Radiation Units and Measurements
IGRT	Image-guided radiotherapy
IMRT	Intensity modulated radiotherapy
LCC	Leeds Cancer Centre
LTO	Leeds Test Objects
MA	Multi-atlas
MAE	Mean absolute error
MAPD	Mean absolute prediction deviation
MDC	Mean distance to conformity
ME	Mean error
MOLLI	Modified look-locker imaging
MPRAGE	Magnetisation prepared rapid gradient echo
MRI	Magnetic resonance imaging
MU	Monitor units

NCCC	Northern Centre for Cancer Care
NN	Neural network
NTCP	Normal tissue complication probability
OAR	Organ at risk
PDF	Probability distribution function
PETRA	Pointwise encoding time reduction with radial acquisition
ppm	Parts per million
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PRV	Planning organ at risk volume
PTV	Planning treatment volume
PVAc	Polyvinyl alcohol cryogel
QA	Quality assurance
RARE	Rapid acquisition with relaxation enhancement
ROI	Region of interest
RTP	Radiotherapy treatment planning
sCT	Synthetic computed tomography
SD	Standard deviation
SEM	Standard error of the mean
SNR	Signal to noise ratio
SPACE	Sampling perfection with application optimised contrasts using different flip angle evolution
SRS	Stereotactic radiosurgery
STL	Stereolithography
TCP	Tumour control probability
TE	Echo time
TOF	Time of flight
TPS	Treatment planning system
TR	Repetition time
TSE	Turbo spin echo
UTE	Ultra-short echo time
VMAT	Volumetric modulated arc therapy
WE	Water equivalent
ZTE	Zero-echo time

Chapter 1 Introduction

1.1 Introduction to the research question

This research project is concerned with the implementation of radiotherapy treatment planning using magnetic resonance imaging (MRI) as the sole imaging modality. The focus of the project is on the implementation of the technique for patients diagnosed with brain tumours, more specifically gliomas and meningiomas. However in order to achieve this, much of the work in this thesis investigates the clinical feasibility of the technique in general and would be relevant for MRI-only radiotherapy planning of other anatomical sites.

MRI offers certain advantages over computed tomography (CT) which has conventionally been used for planning radiotherapy treatments (1). This has led to the integration of MRI into radiotherapy planning. However it has been proposed that moving to an MRI-only radiotherapy pathway, i.e. no longer imaging with CT, would offer additional advantages for the patient, as well as for the treatment clinic (2).

Before this can be implemented however a number of challenges need to be overcome. An MR image does not directly provide the information needed to calculate the dose delivered by a radiotherapy treatment plan (3). Therefore the MRI dataset would need to be modified, allowing this information to be obtained from the image, before it could be used a sole reference for radiotherapy planning.

Additionally, it is well understood that MRI is less geometrically accurate than CT imaging (4). The severity and variability of these geometric distortions must be assessed before the images can be used clinically and recommendations regarding quality assurance (QA) procedures should be produced.

The introduction of new radiotherapy techniques normally involves the production of new QA protocols, enabling the technique to be delivered safely and effectively (5-7). QA requirements would need to be considered before MRI-only radiotherapy could be used clinically. This may involve the development of new QA phantoms.

In this introduction chapter, MRI-only brain radiotherapy is first discussed within the wider context of radiotherapy and radiotherapy imaging (Section 1.2). The MRI sequences and imaging parameters which are relevant to this project are then outlined (Section 1.3). With this background information in mind, further information regarding the specific challenges related to MRI-only radiotherapy is provided (Section 1.4). Finally, a project overview is given and the overall structure of the thesis is outlined (Section 1.5).

1.2 MRI-only radiotherapy planning

1.2.1 Radiotherapy for cancer treatment

Radiation therapy, or radiotherapy as it is commonly termed, is used to prevent uncontrolled proliferation of cancer cells. External beam radiotherapy is delivered using either photon beams or charged particle beams such as electrons using a linear accelerator, known as a linac. The radiation produces free radicals (created when an electron is removed from a water molecule) which damage the DNA within cells, stopping their proliferation.

The aim of radiotherapy is to deliver a tumourcidal dose to the area of disease, whilst minimising the dose absorbed by normal tissue. Increasing the delivered dose increases both the tumour control probability (TCP), but also the normal tissue complication probability (NTCP) (8). A therapeutic window is often used to describe the range of doses which produce a high TCP, with an acceptable NTCP. The range of doses in the therapeutic window is extended by fractionating the radiotherapy treatment (9); normally treatments are delivered every day over a number of weeks. This allows normal cells to repair non-lethal damage, whilst

tumour cells are only able to repair at a slower rate. Fractionation also allows for re-oxygenation of the tumour cells making them more radiosensitive (10).

1.2.2 Imaging in radiotherapy

Radiotherapy treatments have been revolutionised in recent decades by the integration of imaging into the planning pathway, as well as during radiotherapy treatments. Computed tomography (CT) was invented in the 1970s and, as well as allowing for visualisation of the anatomical location of the tumour and normal tissue structures, its use in radiotherapy treatment planning meant that 3D dose calculations could be performed for the first time (11). CT signal intensity is directly related to the electron density in the imaged volume. Knowledge of this is necessary to calculate the resulting dose distribution caused by a photon beam passing through the tissue.

Photons interact in tissue through three processes, namely the photoelectric effect (12), pair production (13) and through the Compton interaction (Compton scattering) (14), all of which result in the production of potentially energetic electrons. For beam energies used in external beam radiotherapy, photons primarily interact in the tissue through the Compton interaction, the result of which is a scattered electron. It is these energetic electrons that deposit dose in the tissue. Therefore the advent of CT imaging, with its resulting electron density information, allowed photon beam attenuation in tissue to be calculated and the resulting 3D dose distribution found. If the geometrical position of the tumour and normal tissue is accurately known then the dose distribution can be optimised to target the tumour, whilst avoiding normal tissue, as is the ultimate aim of radiotherapy (11).

The development of electronic portal imaging devices (EPIDs), which image the megavoltage photon beam, and more recently cone-beam CT (CBCT), through which kilovoltage imaging is performed, has enabled the introduction of image-guided radiotherapy (IGRT) into clinical practice. These devices are mounted onto the linac head, meaning that imaging during radiotherapy treatments can be performed. The purpose of IGRT is to take account of the inter-fraction variation

in tumour and normal tissue position, ensuring that the dose is delivered to the intended area. Inter-fraction variation may be caused, for example, by anatomical changes due to weight loss or by internal organ motion such as peristalsis or rectal and bladder filling (15).

1.2.3 MRI in radiotherapy

MRI is an imaging modality whereby variations in signal intensity are generated by variations in the presence and properties of hydrogen in the imaged object (16). This is useful when imaging human subjects and tissues which are largely made up of hydrogen, in the form of water, as well as lipids. Signal intensity in MRI depends on the density of hydrogen nuclei, which is related to the number of hydrogen atoms in the volume, and characteristic relaxation times. Spin-lattice interactions lead to the recovery of longitudinal magnetisation, characterised by the T_1 relaxation time, while spin-spin interactions cause dephasing of the transverse magnetisation, characterised by the T_2 relaxation time (16). Image contrast is a result of the combination of these factors, as well as other factors such as sequence parameters. It can be chosen which property to weigh the most for an image acquisition. Different tissues have different relaxation times, for example fluids have longer T_1 and T_2 values than fat tissues. This enables differentiation of soft tissues in MR images, which would not be possible with CT, where soft tissues have similar signal intensities due to the similarity of their electron densities.

Within the field of radiotherapy, there is therefore increasing interest about the integration of MRI into the patient pathway (2). MRI is favoured for the delineation of the target, i.e. the structure to be treated, and the organs at risk (OARs), i.e. the structures to be avoided, over CT, due to its superior ability to differentiate soft tissue (1) as discussed above. This is of particular importance due to the increasing use of intensity modulated radiotherapy (IMRT) techniques, whereby areas of high dose can be sculpted conformally around the target, with a steep fall off in dose outside of this region (1). Unlike CT, patients receive no ionising radiation through an MRI scan (1), and therefore this is no increased radiation

risk to the patient when performing multiple MRI scans before and during patient treatment.

The use of MRI scans for target and OAR delineation is typically achieved through transforming the MRI scan into the co-ordinate system of the CT scan, through a process known as co-registration (17). Thus despite the MRI and CT being acquired in separate scanning sessions with the patient imaged in different positions, the MRI anatomy can be matched with that of the CT image. The MRI can be transformed to match the CT anatomy through different techniques. Common techniques for performing this co-registration for radiotherapy purposes are through rigid registration (composed of rotations and translations along any three axes; here the distances between all points in the image are preserved (18)), affine registration (the same transforms as for rigid, but also composed of scaling, shearing and plane reflection along any three axes; here parallel lines remain parallel after the registration (18)) and deformable registration (where a unique displacement vector can be generated for every voxel in the image (18)). The latter technique is more commonly used for anatomical sites outside the head, where greater changes are observed in organ deformation (17).

Motivated by the superior soft tissue contrast associated with MRI, there has recently been significant developments in the field of MRI-guided radiotherapy, such as by the group at University Medical Centre Utrecht (19). Here, an MRI scanner is integrated with a linac in order that MR images can be acquired directly before and during patient treatment on each individual fraction. This would allow online adaptive radiotherapy using MRI to be performed, taking advantage of the anatomical and functional information provided by the modality (20). With this technique, the motion of the targets and organs at risk (OARs) could be clearly assessed on a daily basis, meaning that treatments could be modified or gated where necessary in order to deliver the dose to the correct location. In addition, there is the potential for margins around the tumour volume, needed to compensate for daily anatomical and set-up variation, to be reduced, sparing normal tissue (21).

1.2.4 MRI in brain radiotherapy

The use of MRI in the brain radiotherapy pathway is well-established. A survey of UK clinical neuro-oncologists in the last decade showed that 67% routinely used co-registered MRI and CT scans for treatment planning of high-grade gliomas (22). MRI has been shown to be a superior imaging modality compared to CT (1) for assessing brain lesions, as well as normal brain tissues and OARs, due to its superior soft tissue contrast within the cranium. It has also been shown that inter-observer variation when delineating the gross tumour volume (GTV), the demonstrable extent of the malignant growth (23), is significantly reduced by the use of MRI and CT co-registered information, compared to using CT alone (24).

The ability to use sophisticated MRI sequences also makes the modality appealing for brain radiotherapy planning. For example, MRI signal intensity can be made sensitive to the amount of diffusion within the brain tissue (25), resulting in a diffusion-weighted image. If a series of these images is acquired, the apparent diffusion coefficient of each image pixel can be determined (21). Due to the microstructure of the tissue, diffusion does not always occur isotropically. Diffusion tensor imaging (DTI) can be used to map the measured diffusion as a function of direction. By relating this to the internal brain tissue structure, the connectivity of fibres can be found (a process known as tractography) and this information could potentially be used as part of the radiotherapy planning process (21). Functional MRI (fMRI), which takes advantage of the different magnetic properties between diamagnetic oxyhaemoglobin and paramagnetic deoxyhaemoglobin can be used to detect increased blood flow to critical brain structures which can therefore be avoided during the planning process (21). The efficacy of using DTI and fMRI for brain radiotherapy has been demonstrated in the literature (26). This would allow the sparing of functional structures and fibre tracts within the brain during radiotherapy treatments.

1.2.5 The argument for MRI-only radiotherapy

The aim of MRI-only radiotherapy is to remove the planning CT scan from the workflow, and in its place use MR image(s) alone. As has been discussed in Section 1.2.3, conventionally for radiotherapy treatment planning, MR images are co-registered with CT scans which provide the electron density information required for dose calculations (2). This, however, is known to introduce a registration uncertainty, estimated to be in the range of 0.5-3.5 mm (1 standard deviation typically reported) for prostate and head patients (27-30), which is propagated as a systematic error throughout the treatment. The ability to use MRI alone would eliminate this error. In addition, the radiotherapy workflow would be simplified, meaning that patients would not have to attend as many imaging sessions and the concomitant dose received by the patient would be reduced. The latter is of particular benefit to paediatric patients and those requiring multiple imaging scans during their radiotherapy treatment (21). MRI-only planning is also increasingly appealing due to the development of MRI-guided treatment techniques (31). The ability to plan on the MRI-alone in this situation would increase the efficiency of the workflow, as well as removing registration uncertainty.

1.2.5.1 MRI-only radiotherapy for brain tumours

The use of MRI in neuro-oncology has been discussed previously in Section 1.2.4. Due to the clear benefits of using this modality in the planning of intra-cranial radiotherapy treatments, the brain is an excellent candidate for an investigation of MRI-only radiotherapy planning. The lack of organ motion within the brain, due to the rigid nature of the structures within the skull, also makes it an ideal site for preliminary testing of the technique. The position of the intra-cranial structures should not vary widely between the MRI and CT imaging sessions making the development and testing of the technique less likely to be confounded by anatomical variation.

The clinical implementation of MRI-only planning has some general challenges which will be discussed in Section 1.4, however there are also some specific

challenges related to its application to brain radiotherapy, which mean that a thorough investigation of its application to this site is warranted. These include the nature of the skull around areas where the tumour has been resected prior to radiotherapy treatment. As cranial bone cannot be well visualised using traditional MRI sequences, this may prove problematic. Unlike for some other anatomical sites, such as the prostate, brain tumours vary widely from patient to patient meaning that the target location is unpredictable. This means that sufficient dosimetric testing will be required to ensure that a range of tumour sites are not detrimentally affected by a move to MRI-only radiotherapy. Additionally, brain patients are typically treated in immobilisation masks and on head rests which may not be compatible with MRI scanners and traditional radiofrequency coils. These issues will be further discussed in Section 1.4.4.

1.3 MRI sequences and parameters of interest in this study

1.3.1 MRI sequences

Human subjects are primarily made up of hydrogen atoms, where the nucleus is composed of a single proton, which has a positive charge. When an external magnetic field, B_0, is applied, a proton will precess around the magnetic field, similar to the motion of a spinning top acting under the influence of a gravitational field. The frequency of precession of the proton is directly proportional to the strength of the applied magnetic field and is termed the Larmor frequency (32).

Protons also possess an intrinsic quantum property termed spin, which interacts with electromagnetic fields. When the spin of a hydrogen atom in an external magnetic field is measured, it is found that it can exist in two spin states (a low energy state parallel to the magnetic field and a high energy state anti-parallel to the magnetic field) (32). However it should be noted that as spin is a quantum phenomenon, when no measurements are being acquired, the majority of spins exist in a superposition of the two states (32).

The spin distribution under the external magnetic field can be described by the Boltzmann distribution, where thermal interactions counteract the tendency for the spins to favour the low energy state when measured (33). It is statistically likely, however, that more spins will be aligned parallel to the applied magnetic field than in any other orientation and, as a result, a net magnetization vector parallel to the field is produced (32). The ratio of protons in the two states when measured at body temperature and with an applied B_0 field of 1.5 T is approximately 1.0000004. The net magnetisation is difficult to measure when it is aligned with the relatively large applied B_0 field. Instead the net magnetisation vector must be tipped into the transverse x-y plane to be observed more easily (16).

This is achieved through the application of a radiofrequency field, B_1, which is perpendicular to the external B_0 field. If the frequency of this field is close to the Larmor frequency, it will produce resonance effects and cause the net magnetization vector to rotate. The time for which this pulse is applied determines the angle through which the net magnetization vector is moved, with a 90º pulse being required to move the full component into the transverse plane (33). It is possible to measure this transverse magnetisation by detecting the voltage in a receive coil sensitive to magnetisation in the transverse plane (16). The detected transverse magnetisation is time-varying as it now precesses around the B_0 field.

The measured signal decays over time however. The loss of the coherent transverse signal can be described by two time constants, termed T_1 and T_2 (34). T_1 relaxation is caused by interactions between the spin system and the external environment which result in statistically more spins favouring the lower energy, spin-up state. This will cause the longitudinal component of the net magnetization to grow back to its original magnitude, while the transverse component will reduce to 0. T_1 relaxation can only occur through interactions close to the Larmor frequency (33). T_2 relaxation takes place over shorter timescales and is caused by the dephasing of the transverse magnetisation as a result of interactions at the atomic and molecular levels, which can occur at any frequency. The

relaxation time constant T_2* additionally takes into account the decay of the coherent transverse signal due to inhomogeneities in the B_0 field (33).

All MRI images are produced using a pulse sequence, which consists of radiofrequency pulses and gradient pulses which are set for predetermined and carefully chosen durations and timings (16). The main pulse sequence of interest in this study is a spin echo sequence with a T_1 weighting, as this is used clinically for radiotherapy brain imaging. An illustration of the radiofrequency pulse sequence and resulting signals for a spin echo sequence is shown in Figure 1-1.

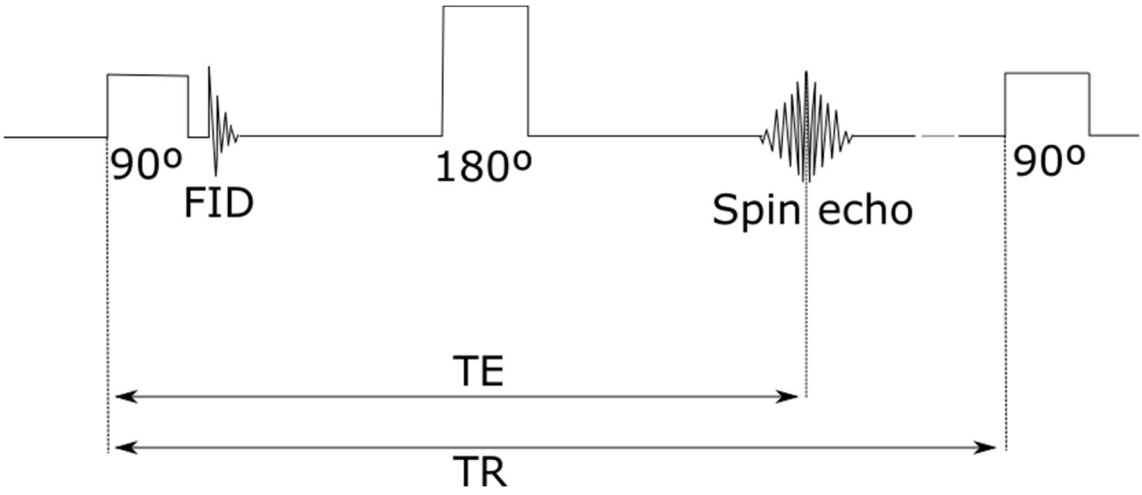

Figure 1-1 An illustration of the radiofrequency pulses and resulting signals from a spin echo sequence. The angles represent the flip angle of the pulses. FID is the free induction decay. This is the detected transverse signal which is dampened by T_2* dephasing. TE and TR represent the echo time and repetition time respectively.

In this sequence the signal lost through T_2^* dephasing is compensated for by applying a 180º pulse which refocuses the spins, resulting in another signal peak (echo) at echo time (TE). This sequence can have T_1 weighting by setting a short TE and a short repetition time (TR), the time between 90º pulses. The short TE reduces the amount of T_2 dephasing which can occur before the 180 º pulse is applied and the short TR means that the longitudinal magnetisations of different tissues have not fully recovered before the next radiofrequency pulse. This means that differences in the T_1 times of tissues are largely responsible for the contrast in the image (35).

The location of data in MRI is encoded using magnetic field gradients in the x and y (in-plane) and z (slice selection) direction (33). Here the gradients alter the frequency and phase of the precessing protons across the imaged volume. Therefore by analyzing the frequency and phase of a signal its physical location can be determined.

Commonly the turbo spin echo (TSE) MRI sequence, known more generically as Rapid Acquisition with Relaxation Enhancement (RARE), is used clinically. This sequence is composed of a series of evenly spaced 180º pulses after each 90° radiofrequency pulse. In this sequence, the phase encoding gradient is changed between successive echoes allowing different data lines to be acquired (16). The advantage of this sequence is that multiple lines of data can be acquired from a single 90º pulse, which may reduce imaging time.

Gadolinium-based contrast agents can also be administered to patients before or during MRI imaging. Gadolinium is paramagnetic and is useful in neuro-oncology as it rapidly builds up in a lesion where the blood-brain barrier is disrupted (16). The paramagnetic molecule decreases the T_1 and T_2 of nearby protons, thus altering the signal from the region, and in the case of oncology, highlighting the location of the tumour. For a T_1-weighted image, the reduction in T_1 times causes a higher signal intensity for tissues where the Gadolinium-based contrast agent has accumulated (16).

1.3.2 MRI parameters

The MR image signal, contrast and the acquisition time can be affected by a number of variable sequence parameters. These parameters include the field of view (FOV) and matrix size of the images. The FOV describes the dimensions of the image plane. The FOV must be large enough to image the anatomy of interest and to avoid wrap-around artefacts, which are caused by aliasing if the FOV does not cover the full extent of the imaging object in the phase encoding direction (16). The matrix size defines the number of pixels in the image plane. The matrix size and the FOV together determine the pixel size, which should be small

enough to allow sufficient resolution for the purpose the image is acquired for. In the case of MRI-only radiotherapy planning this in-plane resolution should be of the order of 1 mm to be consistent with current CT imaging (18) and meet delineation requirements.

The FOV and matrix size describe the x-y plane of the image. MRI image slices in the z direction can either by acquired through 2D or 3D imaging. In 2D multi-slice image acquisition MRI image slices are acquired using separate excitations whilst varying the slice selection gradient. In 2D imaging it is possible that slice gaps may be required due to imperfection in the radiofrequency pulses (16). These should generally be kept to a minimum as information in these gaps is not acquired. This practice would not be suitable in MRI-only radiotherapy planning. In 3D image acquisition the data is acquired by exciting a volumetric slab of tissue, rather than separate slices. Slice encoding, performed in the same way as phase encoding, is then used to separate the data into slices. 3D imaging generally results in a high signal to noise ratio (SNR), due to excitation of the whole imaging volume, and a low slice thickness with contiguous slices, however 3D imaging is generally associated with a longer scan time (35).

Similarly to pixel size, the slice thickness affects the image resolution and should be small enough for purpose. However with a smaller slice thickness, more slices are required to fill the volume, leading to an increase in scan time. Additionally thinner slices lead to a decrease in SNR (35) and therefore a balance must be reached. Signal averaging can be used to reduce noise, and hence improve the SNR, based on the principle that true signal is coherent whereas noise is random. Signal averaging however records the same signal more than once and therefore using this technique increases the scan time (36).

As has already been discussed TE, TR and the flip angle can be used to affect image contrast. These parameters can also affect the SNR. For example a lower SNR can be caused by a lower (less than 90º) flip angle, a short TR (due to incomplete signal recovery in the longitudinal direction) and a long TE (allowing for more decay of the transverse magnetisation) (35). A long TR however can

lead to a longer scan time and it should be ensured that the acquisition time is short enough for the patient to remain comfortable and avoid patient movement. The total number of scans to be acquired should be taken into account.

The receive bandwidth describes the range of frequencies sampled during the readout gradient application. Using a larger bandwidth reduces the chemical shift artefact which arises from protons in different chemical environments experiencing different magnetic field strengths (see Section 1.4.2.2 for further information on this artefact). However this also reduces the SNR as more noise is sampled relative to the signal (35).

Coils are used to transmit and receive the radiofrequency signals. The configuration and placement of these coils affects the image quality (35). The closer the receive coils are to the imaging area the better the SNR. For the purposes of radiotherapy planning, it must be ensured that the coils do not affect the patient contour. This has led to the development of coil bridges which allow positioning of the coil close to the patient, however not in contact with their skin.

It is necessary in MR imaging to define the cardinal axis. In this study the z direction is parallel to the direction of the B_0 magnetic field i.e. directed into the bore of the scanner. The x and y directions are defined perpendicular to this z direction. In this thesis the x and y directions are defined as shown in Figure 1-2. The z direction is also shown here for reference.

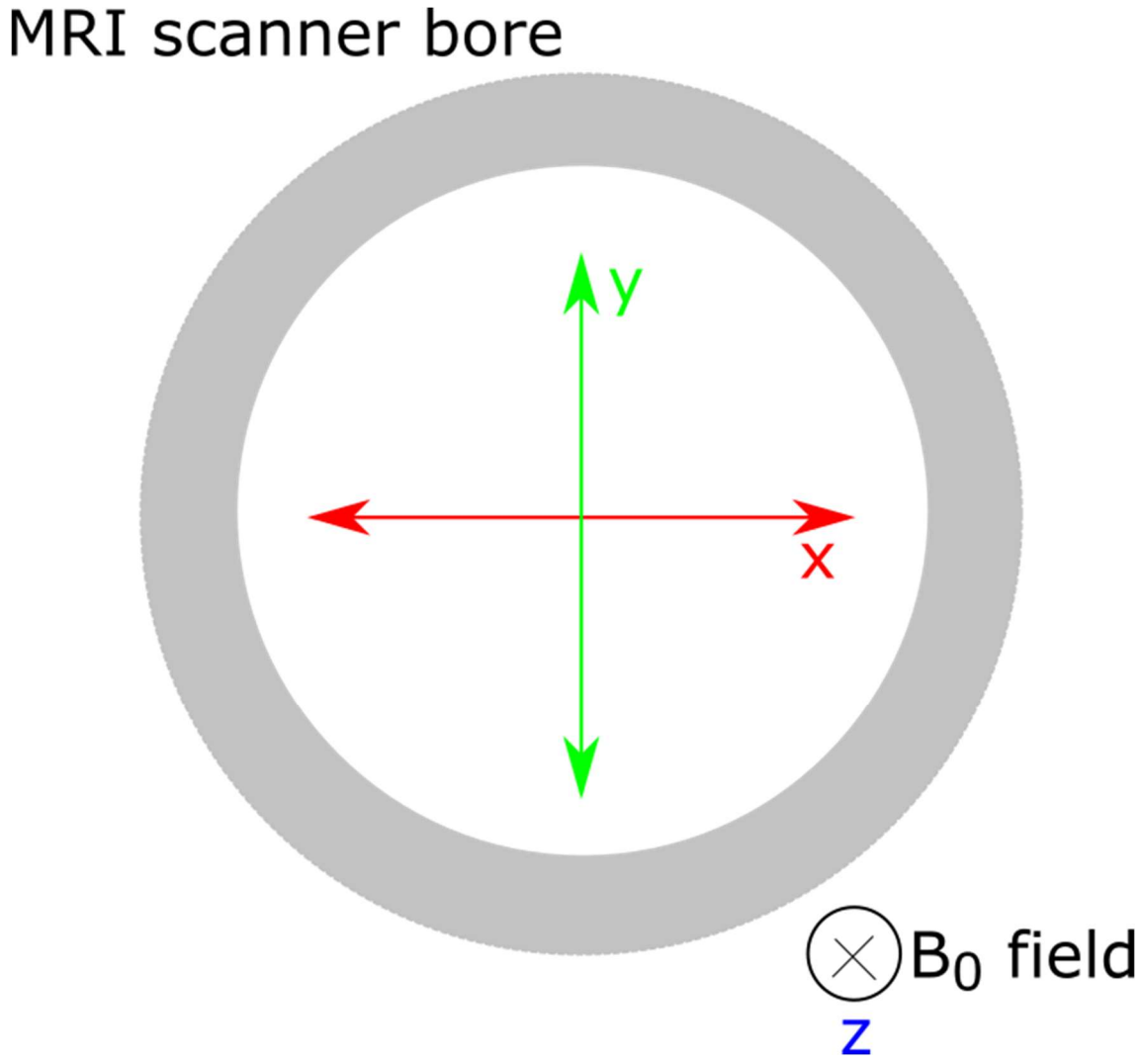

Figure 1-2 The MRI co-ordinate system used in this work. The MRI scanner bore is shown as if looking straight down the bore. The B0 field is aligned directly along this axis, shown here as going into the page. This aligns with the z direction. The x and y directions are orthogonal to this and their directions are demonstrated in the image.

1.4 Challenges to implementing an MRI-only radiotherapy workflow

1.4.1 Electron density information

A challenge when using MRI alone for radiotherapy planning is that MRI signal intensity does not uniquely relate to the electron density, which is needed for dose calculations, as is the case with CT (37). Instead, the MRI signal depends largely on the density of hydrogen nuclei, as well as tissue relaxation properties (38).

The result of this is that MR images cannot be used directly for radiotherapy treatment planning. Instead, a synthetic CT (sCT) scan (also commonly known as a pseudo or substitute CT), containing the relevant electron density information, must be produced for this purpose from the original MR image(s).

Methods for generating sCTs for the purpose of radiotherapy planning for various treatment sites is a topic of current interest. A systematic literature review (Chapter 2) was undertaken as part of this project in order to review the current techniques and to make an informed decision as to which would be the most useful to investigate with the aim of clinically implementing brain radiotherapy.

It is essential that the quality of sCTs is analysed using metrics which confirm the clinical acceptability of the scans for radiotherapy planning, compared to CT-based planning. To achieve this, the dosimetric differences of using both the CT and sCT as the basis of dose calculation for typical radiotherapy plans can be assessed.

A common method of comparing two dosimetric distributions in radiotherapy is to perform gamma analysis, which was proposed by Low et al., (39). Here it is determined whether an evaluated dose point is sufficiently similar to the reference dose, based on both the dose difference and the distance to dose agreement. The acceptable percentage dose difference can either be set relative to the local dose or to the global dose, for example as a percentage of the maximum or prescribed dose, generally resulting in a larger tolerance. The chosen parameters and tolerances require careful consideration based on clinical requirements.

It is also important to ensure that the appearance of the two scans is clinically acceptable through a comparison of the Hounsfield Units (HUs) of the CT images. This similarity in appearance is particularly important if the images are to be used for IGRT.

When imaging with conventional MRI sequences there is an absence of signal from cortical bone, including the skull. Therefore using images as references for

positional verification, which is essential for IGRT using both EPIDs and CBCT imaging, is an additional complication for MRI-only radiotherapy.

1.4.2 MRI geometric distortions

For radiotherapy treatment planning, it is essential that anatomy within the external body contour is represented accurately geometrically in the acquired images. This is also important if the images are to be used as references for IGRT. However, it is well documented that MRI is subject to geometric distortions, which become larger the greater the radial distance from the magnet isocentre (4, 40-42).

MRI geometric distortions can be divided into two categories, system and patient-induced distortions.

1.4.2.1 MRI system distortions

System distortions arise from non-linearities in the gradient fields, as well as inhomogeneities in the static magnetic B_0 field (43). In contemporary MRI scanners with superconducting magnets and sophisticated shimming coils, distortions arising from the latter are relatively small (44). The former however has become a more pertinent issue with the current generation of gradient systems (44) due to the restricted length of gradient coils, as well as the use of fewer coil turns, in order to achieve short gradient rise times (45). Additionally, the choice of pulse sequence affects the distortion resulting from field inhomogeneities, with gradient echo sequences being particularly susceptible (46).

Methods exist for the correction of system distortions, which are specific to the particular MRI scanner, for example by using spherical harmonic deconvolution methods (47). After these correction techniques have been applied, some residual distortion remains however, particularly as the distance-to-isocentre increases.

In order to create a homogeneous B_0 field, shimming of the magnetic field is performed. This can be carried out through fixed shimming. This involves the positioning of steel around the magnet (passive shimming) and through the addition of coils where the currents are predetermined to compensate for the known inhomogeneities (active shimming) (16). Dynamic shimming can also be performed using gradient coils to compensate for inhomogeneities on a per-patient basis. However again some residual magnetic field inhomogeneities can still exist after shimming has been performed.

1.4.2.2 MRI patient-induced distortions

Patient-induced distortions include chemical shift and magnetic susceptibility artefacts. Chemical shift artefacts arise from protons in different molecules within the body experiencing different magnetic field strengths due to electron shielding, for example the chemical shift of fat compared to water is approximately 3.5 parts per million (16), which corresponds to a resonant frequency difference of 224 Hz at 1.5 T. Spatial location in MRI is encoded using gradients and this means that differences due to chemical shift will be interpreted as spatial displacements.

Perturbation of the magnetic field itself can be induced by the susceptibility of tissues within the patient (48). Variations in this susceptibility can cause spatial differences in field strength and consequently lead to associated image artefacts including distortion and signal loss due to dephasing. These occur at tissue-air interfaces, for example superior to the shoulders, as well as around metal implants (16).

Patient-induced artefacts can be reduced by the use of an appropriately high bandwidth (4, 35), however caution should be employed as higher bandwidths lead to a lower SNR (16).

1.4.2.3 The impact of geometric distortions on MRI-only radiotherapy

Guidelines (6), which are now over two decades old, are commonly cited for the acceptability of MRI geometric distortions for radiotherapy purposes, advising

that distortions greater than 2 mm would require consideration. However, more recent studies state that distortions of less than 1 mm near targeted regions should be sought (46). Stereotactic radiosurgery (SRS) is a focal, high dose per fraction and highly precise treatment option used to treat very small lesions, for example, brain metastases (49, 50). It has been suggested that for SRS treatments, the required geometric accuracy of the planning images depends on the target size, with significant detriment to plan quality metrics being caused by distortions of less than 1 mm for targets smaller than 2 cm diameter (51). Therefore the required geometric distortion tolerance likely depends on the purpose for which the images are to be used.

In order to integrate MRI into the radiotherapy workflow, and more specifically to introduce MRI-only radiotherapy, a quality assurance (QA) programme, which includes measurement of MRI system distortions, would need to be implemented (52). The geometric distortions over the full field of view (FOV) should be assessed in order to ensure that the distortions are within a clinically acceptable range (53), both prior to the implementation of the technique and then at regular intervals after commissioning, to ensure consistency in the clinical process. The reason for this is that accurate representation of the anatomy within the body contour is needed for dosimetric calculations in radiotherapy planning (54), as well as for on-line verification during treatment. An accurate representation of the external body contour is also needed for patient set-up during treatment. Assessment of the temporal stability of MRI geometric distortions is necessary in order to guide discussions regarding the frequency of distortion QA for MRI-only radiotherapy. This has not been assessed over the full FOV in the literature and therefore this must be carried out in order to determine the appropriate QA frequency.

1.4.3 End-to-end quality assurance of an MRI-only radiotherapy workflow

1.4.3.1 End-to-end testing of a new workflow

When implementing a new procedure it is necessary to ensure that all aspects are operating correctly, with the correct outputs (5-7). A useful way of testing the whole system is with end-to-end QA, which ensures that the entire workflow is operating correctly (55). End-to-end testing is a sensitive method for identifying a problem in any part of the workflow, however it is not specific and identification of a fault does not mean that the cause of the fault is immediately known and a follow up investigation is likely to be necessary.

1.4.3.2 Anthropomorphic phantoms

Radiotherapy QA is often performed using phantoms which are sufficiently similar to human anatomy due to being unable to perform QA using real human tissue. There has therefore been a development of tissue surrogates which appear similar to human tissues when imaged using various imaging modalities and which have similar dose attenuation properties where appropriate.

In this work, a phantom which could be used for end-to-end testing of an MRI-only radiotherapy workflow was developed, due to the lack of a suitable commercial product. This phantom was designed to appear sufficiently similar to tissues found in the head and neck when imaged using both MRI and CT. The reason that CT imaging was also necessary is due to the need to image the phantom using CBCT during on-line image verification during patient treatment. Additionally, this would allow another way of comparing the dose calculated on a sCT scan to the dose calculated on a CT scan.

Phantoms and tissue surrogate materials have been developed which appear similar to anatomical tissues on both MRI and CT imaging (56-59). However a phantom has not been developed which has sufficiently similar properties to a head and neck when imaged with both modalities to be suitable for MRI-only end-

to-end QA. This therefore is an important area of development which was further investigated in this work.

1.4.4 Patient set-up requirements for MRI-only radiotherapy

A number of techniques are used in radiotherapy to ensure that patients are set-up in a reproducible position, which may pose a challenge when using MRI scanners for MRI-only radiotherapy planning. Patients are typically treated whilst lying on a flat-top couch which enables a reproducible treatment position to be achieved for each fraction (17). However MRI scans are normally carried out using a concave bed, often with padding, to increase patient comfort during scanning (21) which can last minutes, as opposed to CT scans which only last seconds. Additionally, using a flat top couch introduces a gap between the patient and the in-built receive array in the MR couch, which will reduce the SNR.

Brain and head and neck radiotherapy patients are routinely treated using an immobilisation mask and a rigid head rest. This means that a reproducible set-up can be achieved, with a consistent neck position between imaging for treatment planning and each treatment fraction. This can also reduce intra-fraction motion. In MRI, receive coils should be placed as close as possible to the patient surface in order to achieve an adequate SNR. Therefore for MRI-only radiotherapy it would need to be ensured that patients positioned using these immobilisation devices could be scanned in the MRI scanner using surface receive coils.

1.4.5 Automatic organ at risk delineation using MR images

Due to the superior soft tissue contrast of MRI compared to CT, MRI has been recommended for aiding in the contouring of various OARs in the brain during radiotherapy planning (1). It is hypothesised that MRI can be further used for the automatic contouring of brain structures. This follows on from work conducted by Wardman et al., (60) which suggested that automatic contouring of head and neck structures was equally or more accurate when using the MRI compared to the CT.

Automatic contouring is of interest in radiotherapy due to the objective nature of the delineation. It is well understood that inter-observer variation is a source of uncertainty in radiotherapy planning (61) for both target volumes and OAR delineations. The development of automatic contouring methodologies for these structures, where the generated contours are consistently acceptable or only require minor manual adjustments, would remove this uncertainty and improve delineation consistency. In addition the use of automatic contouring would make the workflow more time efficient, providing benefits for the clinic. This would therefore be a useful tool to develop in line with MRI only-radiotherapy planning.

1.5 Project overview

As has been discussed in this chapter, the implementation of an MRI-only radiotherapy planning workflow for brain patients is subject to a number of challenges, which are investigated in this thesis. These include the determination of electron density (Chapter 3), which is needed for radiotherapy dose calculations, and the evaluation of MRI geometric distortions (Chapter 5). As part of the latter assessment, the necessary frequency of geometric distortion QA measurements in order to ensure that geometric integrity is acceptable after implementation, is investigated. Additionally, attention is paid to other QA tests which should be included as part of an MRI-only workflow, leading to the design and production of an anthropomorphic phantom which can be imaged using both MRI and CT (Chapter 6).

As well as presenting solutions to these challenges, the difference in inter-observer variability when contouring normal brain structures on both MR and CT images independently is assessed. As has been discussed in this chapter, there is the potential for automatic contouring of these structures as part of an MRI-only workflow and the clinical acceptability of these automatic contours is also investigated (Chapter 4).

A review of the literature can be found in Chapter 2. Due to the current interest in MRI-only radiotherapy a systematic review of papers regarding the generation of

sCTs for dose calculation and patient positioning verification has been performed. Smaller literature reviews within other chapters of this thesis determine the current knowledge regarding the impact of MRI geometrical distortions on MRI-only radiotherapy planning, necessary QA associated with this, as well as the production of phantoms related to the implementation of this technique.

Chapter 2 Literature Review of synthetic CT generation methodologies

2.1 Introduction

As has been discussed in Chapter 1, a challenge when using magnetic resonance imaging (MRI) alone for radiotherapy treatment planning (RTP) is that MRI signal intensity does not directly relate to electron density, as is the case with computed tomography (CT) (37). Instead, the MRI signal depends largely on the density of hydrogen nuclei, as well as tissue relaxation properties (38). This means that MRI scans cannot be used directly for dose calculation during RTP, without some form of electron density correction. Additionally, in conventional MRI sequences there is an absence of signal from cortical bone. Therefore using images as references for positional verification, which is essential for image guided radiotherapy (IGRT), is an additional complication for MRI-only radiotherapy.

A number of techniques have been developed which attempt to introduce an MRI-only radiotherapy workflow. These methods produce a synthetic CT (sCT) (also commonly known as pseudo or substitute CT) from MR image(s) which can be used for radiotherapy treatment planning, and potentially positioning verification for IGRT.

This chapter analyses the results of a systematic review of methods used in the literature for the production of sCTs for the purposes of MRI-only radiotherapy treatment planning and use in an MRI-only radiotherapy workflow. A recently published review by Edmund and Nyholm (3) searched the Scopus database November 2015 for methods of sCT generation for MRI-only radiotherapy treatment planning and PET-MRI attenuation correction. The authors summarised performance metric values of sCTs and discussed issues related to reporting. The review in this chapter brings the search up-to-date and aims to provide a summary of different methodologies and their potential clinical

implementation, through a systematic search using the Medline and Embase databases.

In this chapter, Sections 2.1 - 2.5 are drawn largely from a systematic review by Johnstone et al., (62), of which I was the first author. An update to this review can be found in Section 2.6.

2.2 Method

A systematic review of techniques was carried out using the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines (63). The Embase and Medline databases were searched from 1996 to March 2017. Papers were included which related to both MRI and radiotherapy. Additionally, the papers included either referred to MRI-only, sCTs, bulk density or synonyms for these terms in their title or abstract. The search criteria used for identifying papers for the review from the Medline and Embase databases are as below.

1. exp Magnetic Resonance Imaging/
2. exp Radiotherapy/
3. (magnetic adj3 (only OR alone)).tw
4. (MR* adj1 (only OR alone)).tw
5. ((pseudo* OR synth* OR sub*) adj1 (CT OR comp*)).tw
6. (bulk adj3 densit*).tw
7. 1 AND 2
8. 3 OR 4 OR 5 OR 6
9. 7 AND 8

The use of ‘/’ following a word or phrase indicates that a key word, which papers are tagged with, is being searched for. The term ‘exp’ means a key word or phrase is being exploded i.e. all subheadings are included in the search as well as the main heading. The use of ‘adj’ followed by a number identifies papers referring to two terms within a certain number of words of each other. ‘*’ symbolises that any letters can follow the word or phrase. ‘.tw’ indicates that the term is being

searched for in the title or abstract of the paper. ‘AND’ and ‘OR’ are Boolean operators.

In summary, in searches 1 and 2, the key words of Magnetic Resonance Imaging and Radiotherapy were exploded. Searches 3 and 4 are concerned with identifying papers which refer to MRI (or magnetic resonance imaging) only in the title or abstract. Search 5 identifies papers referring to synthetic CT or synonyms of this term, again in the title or abstract. Search 6 searches here for references to bulk density or bulk densities. For a paper to be included in the search, it must fulfil the requirements of one of searches 3, 4, 5 or 6, as well as both search 1 and search 2.

Following the database search, duplicated papers were removed and records screened for eligibility. Papers were included which related to the generation of sCTs for use in an MRI-only radiotherapy workflow. Papers focussing on PET-MRI attenuation correction methods were not included. These methods use similar techniques to those in MRI-only RTP and have reported novel sCT generation methodologies producing results of high quality, however reviewing these papers systematically was outside the scope of this review. All papers identified during the search which related to PET-MRI were checked to ensure that no information relating to the use of sCTs in an MRI-only radiotherapy workflow was excluded.

This study considered external beam radiotherapy only and therefore brachytherapy studies were excluded. Brachytherapy (where the dose distribution is governed by the inverse square law due to the close proximity of the source to the tumour), as well as stereotactic radiosurgery (SRS) treatments can use an MRI-only radiotherapy workflow as standard practice. These assume the whole volume is water equivalent. Papers were excluded which related to SRS technicalities and procedures, however papers reporting on novel sCT production for SRS patients were included. Papers discussing the use of MRI in radiotherapy, the integration of MRI into a radiotherapy workflow, cancer screening using MRI and staging and delineation of tumours using MRI were not

included. MRI geometric distortion assessment, quality assurance (QA) of MRI-only radiotherapy workflows, fiducial marker assessment on MRI scans, and registration technique details are important aspects of implementing an MRI-only radiotherapy workflow. However performing a systematic review of these techniques was outside the scope of this review and therefore papers relating to these were excluded. Conference proceedings were not considered. These can contain valid methodologies, however the large number of relevant abstracts was not manageable in this review.

A citation search of the identified papers was performed. Each included study was assigned an sCT generation methodology category. For each category a table of data was constructed. These tables provide a summary of the published techniques, including the key findings of each study and other pertinent factors such as study size, anatomical site and, where appropriate, treatment technique. Studies are ordered based on publication year in each table. A discussion of the clinical feasibility of each methodology follows.

2.3 Results

A flowchart of the systematic search process can be seen in Figure 2-1. The database search yielded 517 records. After duplicate removal, 393 records remained. Out of these, 44 papers matched the inclusion criteria and, from the citation search, an additional 17 papers were identified. Therefore 61 papers in total were included in this review. Reasons for exclusion of papers from the review can be seen in Table 2-1. The number of papers excluded for each reason is given.

The generation of sCTs for radiotherapy treatment planning could be grouped into three main methodology categories. These were bulk density assignments, atlas-based and voxel-based techniques, with the latter being subdivided into techniques using standard MRI sequences alone and those utilising ultra-short echo time (UTE) sequences (which allow visualisation of cortical bone). Studies

have reported results using a range of metrics; issues related to the comparability of these will be discussed.

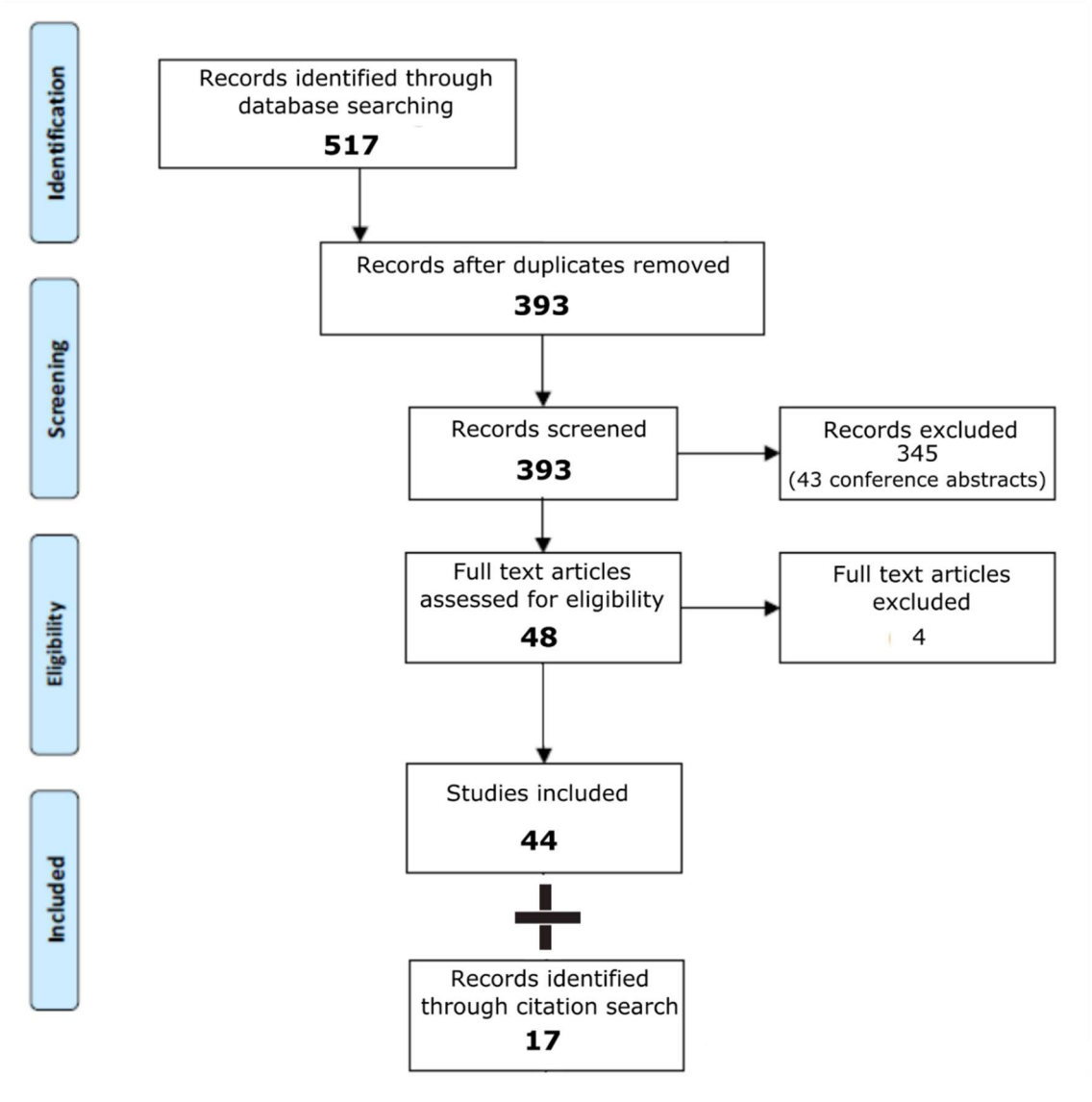

Figure 2-1 Flowchart of the systematic review process, which was in line with the PRISMA guidelines (63).

It should be noted that only the results of 56 studies are reported. This is due to 5 papers from the search, whilst discussing the generation of sCTs for an MRI-only radiotherapy workflow and hence being used in the citation search, not directly testing any novel sCT methodologies.

Reason for exclusion	Number of papers
The general use of MRI in cancer treatment, not focussed on MRI-only planning methods	141
Brachytherapy	37
The role of PET/ SPECT in cancer treatment	22
Stereotactic radiosurgery (SRS)/ radiosurgery technicalities	19
The integration of MRI into the radiotherapy pathway	16
PET-MRI	13
Target delineation with MRI	10
Staging with MRI	9
Not related to cancer	8
Screening with MRI	7
Chemotherapy	5
Not using MRI	4
Geometric distortions	4
Fiducial marker assessment	3
Quality assurance	2
Magnetic nanoparticles	2
The use of ultrasound in cancer treatment	1
Registration assessment	1
MRI sequence details	1
MRI-linac technicalities	1

Table 2-1 Reasons for the exclusion of articles from the systematic review. The number of articles excluded for each reason is given.

2.3.1 Bulk Density Override Techniques

The simplest method to generate a dataset for dose calculation from an MR image is to apply a bulk density override to the entire patient volume, assigning it as water equivalent electron density. This has been tested for brain sites (64-68), as well as for prostate and head and neck studies (69-72).

It can be seen from Table 2-2 that assuming a homogenous density across the volume can lead to dose discrepancies greater than 2% compared to planning heterogeneously on the CT. Korsholm et al., (73) has suggested that a 2% error in MRI dose calculation is clinically acceptable (assuming a 1% dose calculation error when using CT). In addition, with this technique, it is difficult to create reference images that could be used for patient positioning verification due to the lack of bone segmentation.

An alternative methodology is to separate the tissues in the MR image into different classes and assign every class an electron density or Hounsfield unit (HU) value. In most cases this involves two or three classifications; soft tissue and bone (and in some cases air). Improved dosimetric results have been reported using these techniques compared to using a homogeneous density override, namely for prostate, brain and head and neck sites (71-81). Whilst most studies undertook segmentation of structures on the MRI, others (73, 74) contoured the bone on the CT and then transferred the structures to the MRI before overriding densities. In an MRI-only radiotherapy workflow, this would not be possible. These studies were included in this review however, as the results are useful for assessing bulk density techniques for MRI-only planning. The overrides applied in the literature are summarised in Figure 2-2.

	Number of patients	Anatomical site	Densities used	Treatment technique	Key findings	Positional verification
(Schad et al., 1994)(64)	10	Brain	WE	SRS	Dose difference <2%	N/A
(Ramsey and Oliver, 1998)(65)	Single phantom	Brain	WE	CRT	Dose difference <2% (beams passing through cranium), 2-4% (beams passing though air cavities)	DRRs (bones assigned using MRI intensities based on TE and TR) were structurally equivalent.
(Ramsey et al., 1999)(82)	16	Brain	Bones (low MRI values) assigned a pseudo density	DRR study only	DRR study only	DRR misalignments >3 mm could be identified.
(Lee et al., 2003)(79)	5	Prostate	1.WE 2.Bone assigned 320 HU	CRT	Dose difference 1. <5% high dose regions, 2. <2.7% PTV. Most high dose region <2%	N/A
(Chen et al., 2004a)(69)	15	Prostate	WE	IMRT	Dose difference PTV <2%	DRRs created by manually contouring bone (applied density 2 gcm^{-3}). Accuracy <4mm.

(Chen et al., 2004b)(70)	15	Prostate	WE	CRT	Dose difference PTV <2.5%	DRRs created by manually contouring bone (applied density 2 gcm^{-3}). Accuracy <3mm.
(Stanescu et al., 2006)(81)	4	Brain	Bone assigned 1.47 gcm^{-3}	Not specified	Isodose distributions, DVHs, TCP <1%	N/A
(Chen et al., 2007)(83)	20	Prostate	Bone assigned 2.0 gcm^{-3}	DRR study only	DRR study only	DRR max difference 3 mm
(Prabhakar et al., 2007)(66)	25	Brain	WE	CRT	Difference in DVH parameters statistically insignificant. Dose difference <2%	N/A
(Wang et al., 2008)(67)	6	Brain/ Intracranial	WE	IMRT	3-5% dose error	N/A
(Weber et al., 2008)(68)	10	Brain	WE	CRT	Planning brain tumours homogeneously is clinically acceptable	Set-up differences compared to CT-DRRs ranged from 1-4 mm
(Kristensen et al., 2008)(77)	11	Brain	1. WE. 2. Bone assigned 1.6 gcm^{-3}	CRT	Clinical acceptable dose deviations if bone is segmented	N/A
(Stanescu et al., 2008)(80)	4	Brain	Bone assigned 1000 HU	IMRT	Difference in D_{max} and D_{mean} <1%. Difference in TCP <4%	Quality of DRRs not assessed

(Eilertsen et al., 2008)(71)	10	Prostate	1. WE 2. Bone assigned 1.30 gcm^{-3} 3. Bone assigned 2.10 gcm^{-3}	CRT	Dose difference in target volume 1. <2.8% 2. <1.6%, 3. <9.7%	N/A
(Jonsson et al., 2010)(76)	40	Prostate, lung, head and neck, brain	Densities from ICRU 46 (cranium assigned 1.61 gcm^{-3}, femoral bone assigned 1.33 gcm^{-3})	CRT	Largest difference in MU 1.6%	N/A
(Karotki et al., 2011)(72)	10	Head and neck	1. WE 2. Bone assigned 1.50 gcm^{-3}, air assigned 0gcm^{-3}	IMRT	1. 4-5% dose deviations reported 2.Target parameter dose difference <2%	N/A
(Lambert et al., 2011)(78)	39	Prostate	1. WE 2. Bone assigned 1.19 gcm^{-3}	CRT	Dose difference 1. 2.6 ± 0.9% 2. 1.3±0.8%	N/A
(Korsholm et al., 2014)(73)	57	Head and neck, prostate, vesica, pelvic	1. WE. 2. Bone assigned using ICRU 46 (age dependent) 3. In the head and neck. Bone assigned as in 2., air assigned 0 gcm^{-3}	VMAT	2% dose difference in PTV coverage for 95% of patients fulfilled by all bulk density groups for DVH points, D_{median} and $D_{2\%}$	N/A

(Chin et al., 2014)(74)	7	Head and neck	1. WE 2. Bone assigned 1.53 gcm^{-3} 3. Bone assigned as in 2., air assigned 0 gcm^{-3}	IMRT	1. Overestimated target coverage by 15-30% 2. Reduced difference to <2% 3. Improved conformity at air-tissue interfaces	N/A
(Doemer et al., 2015)(75)	10	Prostate	Bone assigned 480 HU (based on average CT measurements)	IMRT	Dose difference <1%	CBCT-MRI cf CBCT-CT largest difference in AP direction (0.15 ± 0.25 cm)

Table 2-2 Summary of the findings of bulk density override techniques. The remaining tissue is assigned of water equivalent unless otherwise stated. (AP = anteroposterior; CBCT = cone beam CT; CRT = conformal radiation therapy; CT = computed tomography; DRR = digitally reconstructed radiograph; DVHs = dose-volume histograms; HU = Hounsfield units; ICRU = International Commission on Radiation Units and Measurements; IMRT = intensity modulated radiation therapy; MRI = magnetic resonance imaging; MU = monitor units; N/A = not applicable; PTV = planning treatment volume; SRS = stereotactic radiosurgery; TCP = tumour control probability; TE = echo time; TR = repetition time; VMAT = volumetric modulated arc therapy; WE = water equivalent.)

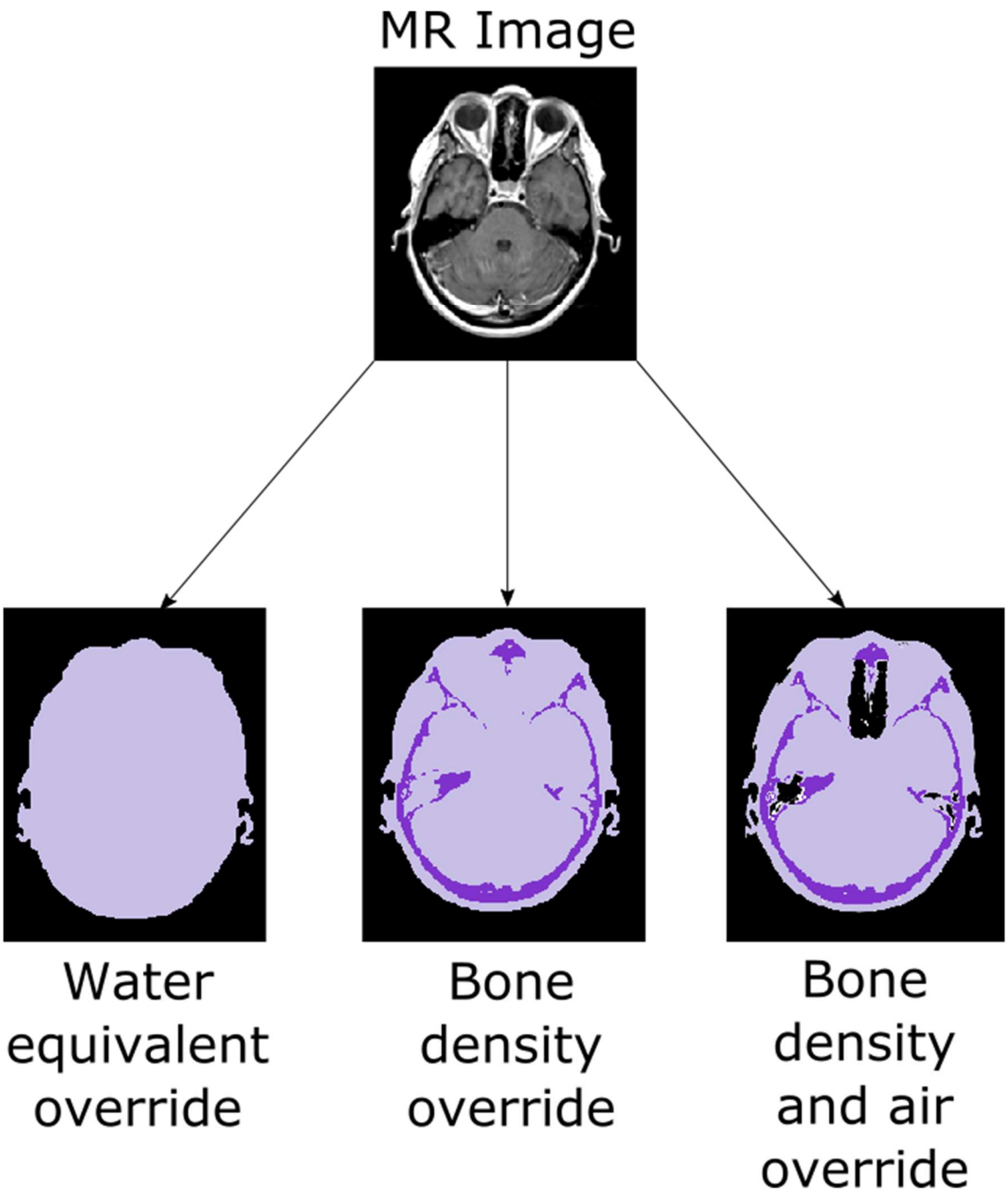

Figure 2-2 An illustration of bulk density override techniques using a brain site as an example. These can be (left to right) a water equivalent override, bone density override, or bone and air override. In the latter two cases, the remaining tissue is assigned as water equivalent.

The dosimetric results from these studies suggest that this technique has the potential to be used clinically, with dose differences of less than 2% typically reported when bone is segmented (Table 2-2). The appearance of cortical bone in conventional MR images however limits its advantages. Cortical bone has a very short T_2^* relaxation time (2) and therefore in conventional MRI it is

represented as a signal void. This makes it difficult to distinguish bone from air and has led many studies to resort to manual bone contouring. This is time-consuming and not practical for routine clinical use. In addition, artefacts such as those associated with dental implants can make segmentation in the head more difficult.

Stanescu et al., (81) attempted a semi-automatic method of bone segmentation in the head. Here, a point was placed close to the structure which required segmentation. Thresholding was then used to segment the structure. The authors noted that manual adjustment afterwards was required in some cases, particularly towards the lower section of the skull. Stanescu et al., (80) used an atlas-based segmentation method to separate the bone, prior to bulk density override. Again manual adjustment was used if necessary. Methods such as these could mean that bulk density techniques are more useful in a clinical workflow in the future, although manual adjustment of contours would not be desirable.

There is debate in the literature over the most appropriate bone density assignment to use (Table 2-2). Densities assigned range from 1.19-2.10 gcm^{-3}. Hoogcarspel et al., (84) stated that dose errors have arisen due to assigning a single bone density rather than separating the bone into individual components.

Varying degrees of dose accuracy for bulk density methods have been reported. This can in part be explained by the use of a different number of tissue classifications as well through assigning different bone density values.

Although most studies do not state the planning algorithms used for dose calculations, earlier studies are likely to use simpler models. Therefore it is likely that there is more uncertainty in these studies in terms of accurately modelling areas of inhomogeneity, particularly low density changes, as well as photon and electron scattering. This should be taken into account when assessing dosimetric differences.

By segmenting bone and assigning bulk densities, reference images for patient positioning can be created. Digitally reconstructed radiographs (DRRs) created using MRI with bulk density overrides have been compared to CT-derived DRRs for prostate and brain patients (65, 68-70, 82, 83). Doemer et al., (75) compared cone-beam computed tomography (CBCT)-to-MRI with CBCT-to-CT agreement for prostate patients. Differences in shift position maximal in the anterior-posterior (AP) direction of 0.15 ± 0.25 cm were reported. The authors postulated the reason disagreement was greatest in this direction was due to bowel preparation issues during MRI scanning.

In the following sections other methods of sCT generation are discussed. Some studies referenced compare the results for generated sCTs to bulk density techniques. Where this is the case, the bulk density results are listed in the relevant table.

2.3.2 Atlas-Based Techniques

Atlas-based techniques (see Table 2-3) typically use a single, standard MRI sequence in order to produce an sCT (85). This ensures that scanning time is kept to a minimum, reducing the chances of patient movement (85). It also means that the scanning protocol is straightforward to implement in a clinical environment. The process for sCT production can be fully automated and reference images for positional verification can be produced as well as automatic contouring of OARs. Sjolund et al., (86) remarked how atlas-based techniques are relatively robust to image artefacts due to their reliance on prior training information.

	Number of patients	Anatomical site	Atlas type (number used)	Treatment technique	Key findings	Positional verification
(Greer et al., 2011)(87)	39	Prostate	Single	Not a planning study	No significant difference in HU values for organs of interest	N/A
(Dowling et al., 2012)(88)	36	Prostate	Single	CRT	Point dose difference <2%. No significant difference in Chi values or OAR HU values. Mean DSC 0.79, 0.70, 0.64, 0.63 for bone, prostate, bladder, rectum respectively.	DRRs could be generated
(Uh et al., 2014)(89)	14	Paediatric brain	1. Single random 2. Arithmetic mean (6) 3. Pattern recognition (6) 4. Pattern recognition (12)	CRT/ IMRT	Multiple atlases outperformed single atlases. For these, dose difference $D_{95\%}$ and $V_{95\%}$ <2%. Pattern recognition had an equal performance cf the mean. The root mean square difference was greater for WE override, and showed greater DVH differences.	N/A

(Dowling et al., 2015)(85)	39	Prostate	Multi-atlas (38) with local patch-based pattern recognition	IMRT	ME = 0.6 ± 14.7 HU. MAE = 40.7 ± 8.2 HU. Mean DSC>0.80 for all organs. Change in MU = 0.3 ± 0.8%. 1.00 gamma pass rate (2%, 2mm)	DRRs generated
(Siversson et al., 2015)(90)	10	Prostate	Multi-atlas (15) with local pattern recognition	VMAT	MAE = 36.5 ± 4.1 HU. Average target dose difference = 0.0 ± 0.2%. Average gamma pass rate 99.9% (2%, 1mm)	N/A
(Sjolund et al., 2015)(86)	10	Head region	Multi-atlas (9) using iterative voxel-wise average	Not a planning study	Voxel-wise median performs better than voxel-wise mean	N/A
(Edmund et al., 2015)(29)	6	Brain	Multi-atlas (5) with local pattern recognition	Not a planning study	Average median absolute error 184 ± 34 HU	Largest deviations cf CT-CBCT were <1 mm and 1°
(Andreasen et al., 2015)(29)	5	Brain	Multi-atlas (3) patch based method	CRT	<0.5% dose difference in the target	Quality of DRRs not investigated

(Andreasen et al., 2016)(91)	10	Prostate	Multi-atlas `leave-one-out' patch based method, with highly dissimilar atlases ignored	VMAT	Average MAE 54 HU. Average MAE WE path length 1.2 mm. Median deviation <0.4% relevant DVH points. Average gamma pass rate 97.0% (1%, 1 mm). Performed significantly better than WE override for the majority of metrics.	N/A

Table 2-3 Summary of the findings of atlas-based techniques. (DSC = dice similarity coefficient; MAE = mean absolute error; ME = mean error. Other abbreviations as in Table 2-2.)

The simplest atlas techniques use a single or average atlas, for example as developed by Dowling et al., (88) for prostate planning. With an average atlas technique, pairs of MRI and CT scans from a database of patients are co-registered. An average MRI atlas is then created, potentially with a matching set of organ contours. By determining the deformations which need to be applied to each MRI scan in the database to reach the average atlas, an average CT atlas can be created by applying the same deformations to the corresponding registered CTs and finding the average of these. In order to create an sCT for an incoming MR image, the average MRI atlas is registered to the incoming MRI scan. These deformations are then applied to the average CT atlas resulting in a corresponding sCT. The organ contours can be propagated similarly.

Dowling et al., (88) validated their method through a ‘leave-one-out’ approach, which is commonly used in sCT evaluation (results in Table 2-3). Here the training atlas is determined using all patients except one. The scan of this excluded patient is used as an input in order to test the model. Differences in dose were found to be largely attributable to changes in external patient contour between MRI and CT scanning. Additional work by the group confirmed no significant difference in HU values for the main OARs between sCTs and CTs (87).

Demol et al., (27) used a single atlas as a baseline for brain MRI-only radiotherapy. Here, the co-registered MRI and CT of one patient is used rather than an average atlas. The authors reported statistically significant dosimetric errors using this method. Additionally, it was found that for a test patient where a large section of skull had been removed, the sCT was assigned as bone in this area.

Several groups have reported improved quality sCTs when multiple atlases, combined with local patch-based pattern recognition methods, are used (Figure 2-3). By combining an atlas technique with these methods, the effect of uncertainty in image registration is reduced (89).

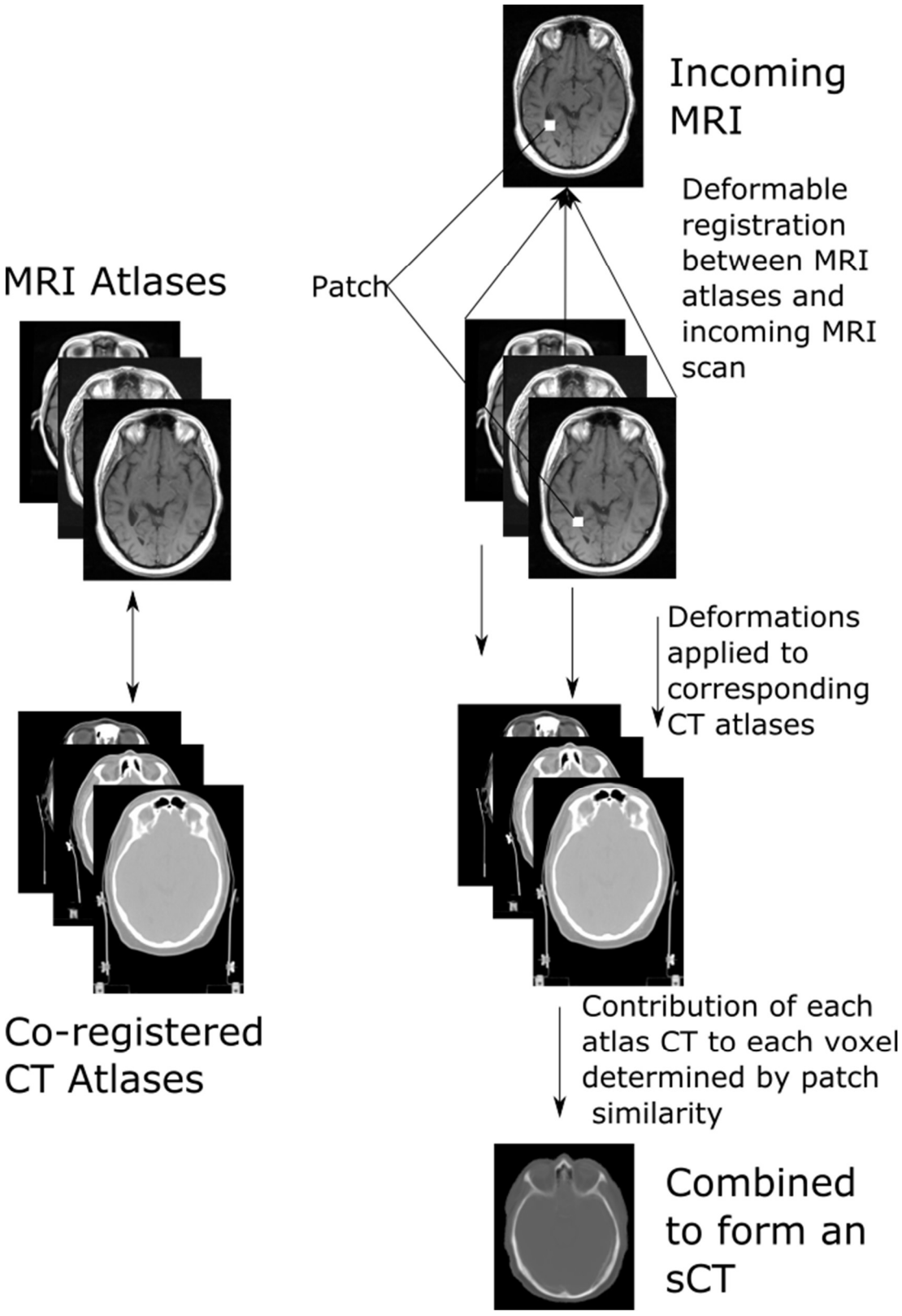

Figure 2-3 Illustrating a multi-atlas technique for creating synthetic CT images. A single synthetic CT image is created using weighting techniques based on pattern recognition in most cases. A patch consists of a pre-defined number of voxels.

Uh et al., (89) used a multi-atlas method for brain sCT production. Here, pairs of CT and MRI scans were rigidly co-registered. When a new patient MRI was input, all MRI atlases were deformed to match the incoming image and these deformations were subsequently applied to the corresponding atlas CTs. The final sCT was calculated by combining the deformed CT atlases using a pattern recognition approach. Here, the intensity of each voxel in the sCT was a weighted average of voxel intensities from the deformed atlas CTs. The voxel in the same location in the atlas, as well as a defined number of neighbouring voxels (a patch), contributed to the prediction. For each voxel, the weighting of each atlas's contribution to the final sCT was determined by assessing the similarity of the patch between the MRI atlas and the incoming MR image. The smoothing of patient-specific anatomy was less pronounced with this weighted atlas technique compared to taking the average of the atlases. The dosimetric results of this study can be found in Table 2-3.

Sjolund et al., (86) used multiple atlases to produce sCTs in the head. The collection of deformed CTs was iteratively registered to their joint voxel-wise mean. It was found however that using the voxel-wise median of the deformed CT dataset gave superior results.

Andreasen et al., (29) used a patch-based regression model for generation of brain sCTs using a multi-atlas approach based on affine registrations. Corresponding MRI patches and CT target values were extracted. A database of these was created for every patch location for every patient. For a patch in the test patient MRI, the CT number was assigned by performing an intensity-based nearest neighbour search of the patch database. A similarity measure was used to weight the contribution of the patches. Additionally, the structural similarity measure (92), based on the mean and variance of the patches, was used in order to discard highly dissimilar patches and hence reduce the number of similarity comparisons required. The method was found to produce comparable quality sCTs to a multi-atlas method using non-linear registration.

This method was also tested on prostate patients (91). Before the patch search, atlases which were highly dissimilar to the patient scan were discarded. A significant reduction in the time needed to produce an sCT was achieved by implementing an approximate nearest-neighbour search of the patch database.

Dowling et al., (85) used a multi-atlas local weighting patch-based method to produce sCTs for prostate MRI-only planning. In this study, a training database was created by co-registering individual patient CT scans to their respective T_2-weighted MRI scan. This was performed using a rigid registration, followed by a structure-guided deformable registration (in order to improve the registration of the bladder and rectum volumes which had previously been contoured). To create a new sCT the following method was used (again a leave-one-out approach was employed):

1. All MRI scans in the atlas database were registered to the incoming MRI scan. This was performed through a rigid registration followed by a deformable registration (this was not structure-guided due to the lack of structures on the incoming MRI). For each atlas, the same deformations were applied to binary images representing pelvic structures which had been contoured on each MRI atlas.

2. A patch (radius of 2 voxels) around each voxel in the incoming MRI was compared to patches in the same location in each of the atlas MRIs. The sum of squared differences was used to assess the similarity between the two. For each voxel, the weighting of the individual MRI scans was normalised to 1.

3. This normalised weighting was then used to determine the contribution of each co-registered atlas CT to the HU value of same voxel in the final sCT. The same weights were used to determine the contribution of each of the deformed binary images to the final pelvic structure segmentation.

4. A 1 mm expansion to the body contour was added in the sCT. The authors proposed this was necessary to compensate for missing signal from collagen in this area in the T_2-weighted MR images (85). This was assigned a bulk value of 47 HU based on patient CT scans.

Siversson et al., (90) used a multi-atlas method for the creation of sCTs for prostate MRI-only planning. The incoming MRI was auto-segmented into 5 structures; prostate, bladder, colon, bone and fat using a multi-template approach with machine learning. This was an automated segmentation algorithm which had been trained using MRI scans in the template database, along with their associated delineations. This was followed by a non-linear warping procedure whereby the template MRI atlases were deformed such that their segmented structures matched those of the incoming MR image. Linear deformations were applied to the tissues, both within and between structures. A further constrained non-rigid registration was carried out in order to align fine-grained structures. A voxel-wise weighted median HU value of all deformed CT atlases was determined with the weighting based on the resemblance of the candidate sCT to the incoming MRI.

Andreasen et al., (29) used a patch-based multi-atlas method in order to assess the feasibility of using the resulting sCTs for set-up verification of brain patients. Using the positional corrections for a CT-CBCT image match as a reference, the corrections for using the MRI and sCTs as reference images were assessed (results in Table 2-3).

As patch-based pattern recognition methods are typically used to weight the contribution of each atlas to the final sCT on a voxel-wise basis, the techniques are in part voxel-based. Although these methods fall naturally into the category of an atlas-based method, it is note-worthy that this overlap exists.

2.3.3 Voxel-Based Techniques

An alternative method for the generation of sCTs is through voxel-based techniques. These can involve the use of standard or specialised sequences,

such as UTE imaging. Some techniques use a mixture of the two. These methods create sCTs using MRI intensities from a number of sequences. With voxel-based techniques the need for accurate registration of an incoming MR image to an atlas is not necessary (85, 93) and no segmentation of images is needed if statistical methods are used (94). These techniques are well-equipped to handle patients with atypical anatomy (85, 93) and have shown ability to separate bone from air (75, 94).

2.3.3.1 Voxel-Based Techniques: Standard MRI sequences

A number of authors have developed voxel-based techniques using routine, clinical MRI sequences (results in Table 2-4).

A group (95) in Helsinki, Finland devised a method for sCT generation using T_1/T_2*-weighted MR images for pelvic sites. MRI and CT images were registered using bony anatomy. The MR images were normalised based on the average intensity of a muscle which was located closest to the bone, in order to compensate for intensity level variation between patients. For each patient 40 voxels within the cortical bone, trabecular bone and bone marrow were chosen at random. The corresponding HU values and MRI intensities of the identified points were used to generate a model. The authors also converted the MRI intensities of tissue outside bone (96). One thousand points were used to analyse the relationship between HU values and MRI intensities for soft tissue in the pelvis. The model divided MRI intensities into threshold-based sections for muscle, fat and urine, assigning bulk HU values. Between these tissue classes, the MRI intensities were converted into HU values using linear interpolation.

This study also developed a conversion model for brain patients using 700 points in bone and soft tissue. Separate models were applied for bone and soft tissue after auto-contouring the bone. Post soft tissue assignment, bulk overrides were applied for fluid, white matter, grey matter and the scalp. Between these tissue classes, linear interpolation was used to convert MRI intensities to HU values.

	Number of patients	Anatomical site	Sequences used	Treatment technique	Key findings	Positional verification
(Kapanen and Tenhunen, 2013)(95)	10	Prostate	T_1/T_2* 3D gradient echo	CRT	Mean prediction error 135 HU. Maximal dose difference improved compared to WE and bone bulk density override.	Bone edge error <1 mm
(Korhonen et al., 2013)(97)	Phantom	Phantom	T_1/T_2* 3D gradient echo	6MV and 15MV CRT	Dose difference <1.3% (6MV) and <1.0% (15MV) behind bones (corresponding bone bulk density override values 2.7% and 2.0% respectively).	N/A
(Korhonen et al., 2014a)(96)	10	Prostate	T_1/T_2* 3D gradient echo	IMRT/ VMAT	Average local absolute difference 11 HU for soft tissue, 99 HU for bone. PTV dose difference <0.8%. 94% (IMRT) and 92% (VMAT) passed (1%, 1mm) gamma analysis. Significantly superior (dose, HU agreement) to WE override.	N/A

(Korhonen et al., 2014b)(98)	15	Prostate	T_1/T_2* 3D gradient echo	IMRT/ VMAT	1 mm bone segment error equivalent to 0.4% change in prostate dose level. Need <2mm bone segmentation error to achieve 2% dose consistency.	N/A
(Yu et al., 2014)(99)	20	Brain (SRS)	T_1	DRR study only	DRR study only	Maximum distance difference <1.88 mm. Mean geometric difference 0.05 ± 0.85 mm
(Kim et al., 2015b)(100)	9	Prostate	T_1	IMRT	MAE 74.3 ± 10.9 HU. Difference in mean target dose 0.63 ± 0.34%. Gamma analysis pass rate 99.9 ± 0.1% (2%, 2mm)	DSC for AP and lateral DRRs 0.90 ± 0.04 and 0.92 ± 0.05 respectively
(Kim et al. 2015a)(101)	15	Prostate	T_1/T_2/Turbo field echo	IMRT/ VMAT	Better dose agreement for $D_{95\%}$, $D_{99\%}$, D_{mean} (not statistically significant), significant improvements in the bladder cf bone bulk density and WE overrides. Gamma pass rate (1%,1mm) 97.2%, exceeded bulk overrides.	N/A

(Korhonen et al., 2015)(102)	5	Prostate	T_1/T_2*	CBCT/ DRR study only	CBCT/ DRR study only	Standard deviation of difference <1.7 mm. Similarity metrics improved cf bulk sCT-DRRs.
(Koivula et al., 2016)(103)	20	10 Prostates and 10 brains	T_1/T_2*	Intensity modulated proton therapy	MAE 34 HU (brain) and 42 HU (prostate). Maximum absolute dose difference in clinical target volume 1.4% (brain), 0.6% (prostate) (cf 1.8%, 8.9% in the brain and 1.2%, 3.6% in the prostate for bone bulk and WE override respectively). > 91% passed gamma criteria (1%,1mm) (improved cf bulk density and WE).	N/A

Table 2-4 Summary of the findings of voxel-based techniques using standard MRI sequences. (3D = 3-dimensional. Other abbreviations as in Table 2-2 and Table 2-3.)

The same group (97) carried out a phantom study determining dose deviations behind bones in RTP when using this technique. This work was extended (98) to investigate bone outline errors on sCTs of prostate patients and the effect of these on dose calculation. Prostate sCT-derived DRRs were also assessed (102). Dose calculation accuracy of prostate sCTs created using this method for proton therapy have been investigated (103).

Kim et al., (100) used T_1 and T_2-weighted MR images, with co-registered CT images, to create sCTs for prostate patients. The bone was manually contoured. All remaining low intensity voxels on the MRI were assigned as air and a bulk HU override was used for these regions. A truth table was created in order to assign the remaining voxels a tissue class based on their MRI intensities. The signal intensities for these voxels were calculated using a weighted sum of all the MR images. sCTs generated using this method were compared to those created using bulk density override methods (101).

Yu et al., (99) used T_1-weighted images to contour the airways (manually and through interpolation) for head and neck patients. Compact bone, spongy bone and soft tissue masks were generated using statistical characteristics of MRI intensities. The MRI intensities of the voxels were mapped to their respective CT number ranges for each tissue.

2.3.3.2 Voxel-Based Techniques: Ultra-short echo time imaging

A challenge of MRI-only planning is that cortical bone is difficult to differentiate from air using standard MRI sequences. This has led some authors to undertake time-consuming manual contouring of either bones or airways, which in certain cases relied on CT information. Dual-echo UTE (dUTE) allows imaging of tissues with short T_2* relaxation times such as bone (104) allowing air and bone to be more easily segmented.

In UTE sequences data acquisition starts as soon as possible after the radiofrequency pulse is complete (105) (a dead time of approximately tens of microseconds exists due to switching between transmit and receive modes

(106)). The readout gradients are switched on at the start of data acquisition, resulting in a centre-out (“koosh-ball”) trajectory being used to sample k-space (105). These techniques allow signals from short T_2* tissues to be sampled, along with those from longer T_2* constituents. A dual-echo sequence involves acquiring two images at two different echo times. Post-processing and subtraction of these images can be used to emphasise signal from the short T_2* components (106).

Results of techniques using these sequences can be seen in Table 2-5. Some methods use dUTE sequences alone, whilst others combine them with standard sequences. These techniques have so far only been clinically tested on brain patients.

Rank et al., (107) and Rank et al., (108) used a 2D turbo spin echo (TSE) sequence with proton density weighting, as well as a 3D dUTE sequence. The model parameters of a tissue classifier were determined by finding the voxel-wise correlation between the corresponding MRI and CT image sets for 2 brain patients. This classifier had as input MRI intensities from the image sets, as well as neighbourhood and co-ordinate information. For a test patient, the probability of a voxel belonging to a specific tissue class was determined using this model.

A group (109), (94) from Umea, Sweden used regression models in order to generate sCTs in the head region. dUTE was used, along with a T_2-weighted 3D spin echo based, sampling perfection with application optimised contrasts using different flip angle evolution (SPACE) sequence. The additional use of the SPACE sequence enabled tissues with a long T_1 value to be distinguished from air. Each of the MR images and the CT image were considered a variable in the model, with the signal intensity of each voxel a sample of the variable. Two additional images for each MRI scan were derived using the mean and standard deviation of voxels in a 27-voxel neighbourhood. These were also input as model variables. Using Gaussian mean regression the expected CT number of each voxel was determined using the variables in the model. The method takes spatial location into account in order to help discriminate between tissues located at different interfaces (110). Johansson et al., (111) attempted to use parallel

	Number of patients	Anatomical site	Sequences used	Treatment technique	Key findings	Positional verification
(Johansson et al., 2011)(109)	5	Brain	T_2/ dUTE	N/A	MAE 137 HU. Accurately discriminated between bone and air.	N/A
(Johansson et al., 2012)(94)	14	Brain	T_2/ dUTE	N/A	MAPD 140 HU. Largest error at air-soft tissue and bone-soft tissue interfaces.	N/A
(Johansson et al., 2013)(110)	9	Brain	4 UTE sequences	N/A	MAPD 130 ± 18 HU. Accuracy improved by the inclusion of spatial information.	N/A
(Jonsson et al., 2013)(93)	5	Brain	T_2/ dUTE	CRT	Dosimetrically identical to air and bone bulk density assigned plans. Improved cf WE.	sCT-DRRs sufficient for treatment set-up, but inferior quality around air cavities.
(Hsu et al., 2013)(112)	10	Brain	T_1/ T_2/ Dixon/ dUTE/ TOF	N/A	UTE showed significant improvement in discriminating bone and air.	sCT-DRR comparable to CT-DRR. Bony edges of the skull were well visualised.
(Rank et al., 2013a)(107)	3	Brain	Proton density/ dUTE	Ions/ Photons	MAE <165 HU. Deviations at air cavities, bones and interfaces. PTV mean dose difference <2% (ions) and <0.2% (photons).	N/A

(Rank et al., 2013b)(108)	Phantom	Phantom	Proton density/ UTE/ T_2/ MPRAGE	Ions	MAE <95 HU. Deviations at interfaces. PTV mean dose difference <3.1%	N/A
(Johansson et al., 2014)(111)	23	Brain	dUTE (parallel imaging investigation)	N/A	SPIRiT parallel imaging marginally improved sCT quality compared to gridding (but increased computational time)	N/A
(Edmund et al., 2014)(113)	5	Brain	2 dUTE	CRT	Statistical regression better geometrically cf Bayesian and threshold methods. All methods superior to WE. Statistical regression and threshold better dosimetrically than Bayesian.	N/A
(Jonsson et al., 2015)(114)	5	Brain	2 dUTE	Not specified (optimised plans)	Mean difference D_{min} and D_{max} <0.3%. 100% of target voxels passed gamma analysis (1%, 1mm). 99.4-100% all voxels passed gamma analysis (3%, 3mm). DSC 0.8-0.9 for bones.	No appreciable differences for DRRs
(Hsu et al., 2015)(115)	12	Brain	UTE/ PETRA	N/A	UTE has significantly greater discriminating power in separating air and bone than PETRA.	N/A

(Paradis et al., 2015)(116)	12	Brain	TSE Dixon/ MPRAGE/ TOF/ UTE	VMAT	Mean dose parameter differences in target <1%. No significant change in MU. OAR D_{max} difference -2.2 - (+1.9) Gy.	N/A
(Zheng et al., 2015)(117)	10	Brain	UTE/ Dixon/ T_1 FFE/ T_2 TSE/ FLAIR	Not specified	MAE 147.5 ± 8.3 HU. Gamma analysis pass rate 99.4 ± 0.04% (1%, 1mm).	N/A
(Gutierrez et al., 2015)(118)	6	Pre-clinical	T_1/ T_2/ ZTE/ UTE	CRT/ arcs	For arced beams, the ZTE/ UTE (2 ms echo time) sequence combination was the most accurate: 0.7% deviation	N/A
(Price et al., 2016)(119)	Phantom/ 12 patients	Brain	UTE/ Dixon/ T_1 FFE/ T_2 TSE/ FLAIR	Not a planning study	MAE 149.2 ± 8.7 HU	sCT-DRRs within 1 mm of CT-DRRs
(Yang et al., 2016)(120)	7	Brain	UTE (2 echo times)	Not a planning study	Bone match <1 mm different cf CT	Registration error between DRRs <1 mm

Table 2-5 Summary of the findings of voxel-based techniques using UTE sequences. (dUTE = dual UTE; FFE = fast field echo; FLAIR = fluid attenuated inversion recovery; MAPD = mean absolute prediction deviation; MPRAGE = magnetisation prepared rapid gradient echo; OAR = organ at risk; PETRA = pointwise encoding time reduction with radial acquisition; TOF = time of flight; TSE = turbo spin echo; UTE = ultra-short echo time; ZTE = zero-echo time. Other abbreviations as in Table 2-2 and Table 2-3.)

imaging in order to reduce the imaging time needed for sCT generation. The authors evaluated different methods of parallel imaging.

Jonsson et al., (93) and Jonsson et al., (114) reported on the use of this method for intracranial targets, assessing DRRs. These authors reported that the greatest discrepancies were around the posterior nasal cavities. Yang et al., (120) compared UTE-MRI-derived DRRs to conventional DRRs for brain patients.

The Ann Arbor, Michigan group (112, 116) used statistical regression combined with spatial information in order to create brain sCTs. The authors used dUTE along with time-of-flight (TOF) angiography to image blood vessels. TSE Dixon (used to separate fat and water) and T_1-weighted magnetisation prepared rapid gradient echo (MPRAGE) images were acquired (116). Air masks and vessel masks were created. Fuzzy c-means clustering with a spatial constraint was used to assign the remaining voxels a probability of belonging to each of 5 classes; fat, fluid, grey matter, white matter and bone. This allowed for a mixture of tissue types within one voxel. DRRs derived from this process were compared to CT-derived DRRs. The authors (115, 116) compared using standard UTE to using pointwise encoding time reduction with radial acquisition (PETRA) sequences.

PETRA is a type of UTE imaging. In standard UTE sequences, data is acquired during gradient ramp-up, which can lead to image artefacts, and a radial `koosh-ball' trajectory is used to sample k-space (121). In PETRA, data acquisition begins after gradient ramp-up. In order to avoid a resulting gap in the centre of k-space, PETRA uses both radial and Cartesian sampling, the latter being used to fill the middle of k-space (121).

Zheng et al., (117) modified the method developed by Kim et al., (100) and applied it to brain studies. Here bone-enhanced images (created using inverted UTE and Dixon sequences) and air masks were input into the previous workflow, along with bone-enhanced fluid attenuated inversion recovery (FLAIR) and UTE imaging. MR images were segmented into 5 tissue classes: air, bone, fat, brain matter and cerebrospinal fluid using a Gaussian mixture model, and sCTs were

generated using the voxel-based technique described previously (100). The same group tested the sCTs for use in IGRT (119).

Edmund et al., (113) undertook a review of the use of UTE in the creation of sCTs for brain patients. Three approaches were investigated; a threshold-based approach often performed for PET-MRI studies (113), a statistical regression approach, and a Bayesian method whereby for each voxel a probability of belonging to each Gaussian distribution (i.e. tissue class) is estimated. The voxels are assigned to the tissue class with the highest probability. The authors compared the methods to a bulk density override, setting the entire volume as water equivalent.

There has been preclinical work investigating zero-echo time (ZTE) imaging (118) (where the readout gradients are switched on prior to the radiofrequency pulse (105)) combined with other sequences, including UTE, which reported more accurate dose calculations compared to using UTE alone. This review has not identified any clinical investigations using this sequence.

2.3.4 Hybrid methods: Atlas and Voxel-based techniques

Hybrid methods using elements of voxel-based and atlas-based techniques have been tested; examples are described below (results in Table 2-6).

Gudur et al., (122) used a voxel-wise technique with a Bayesian framework to create sCTs for brain patients. T_1-weighted MR images were acquired and deformable image registration between an MRI atlas and the patient MRI was performed. The intensity of each voxel in the MRI scan and the knowledge of the geometry of the voxel compared to the reference anatomy were used to create two conditional probability distribution functions (PDFs). The mean value of the PDFs for each voxel was used to determine its electron density. Bone and air could be differentiated on the T_1 images due to the use of an atlas, and the impact of registration issues associated with an atlas were reduced by the additional use of intensity information. The main difficulty lay in the compromise between

	Number of patients	Anatomical site	Sequences used	Treatment technique	Key findings	Positional verification
(Gudur et al., 2014)(122)	9	Brain	T_1 Gradient echo	N/A	MAE 126 HU (cf 282 HU WE override). For 90% sensitivity in bone detection, 86% specificity.	N/A
(Demol et al., 2016)(27)	21	Intracranial	T_1	SRS	Single atlas led to significant dose differences. Hybrid method performed better. Here, for 85% of the patients, mean dose difference to PTV <2%.	N/A

Table 2-6 Summary of the findings of hybrid methods for synthetic CT generation. (Abbreviations as in Table 2-2 and Table 2-3.)

accurately representing detailed structures in the anatomy, whilst avoiding becoming over reliant on a single registration.

Demol et al., (27) compared a method using a single atlas, to a method combining atlas and intensity methods for brain patients. For each voxel, a search of the nearest 81 voxels was performed on the deformed MRI-atlas to identify voxels within 10% of the input intensity. The sCT value of the voxel was determined by averaging the CT atlas voxel values corresponding to those selected on the MRI atlas. This was found to give superior dosimetric results to using a single atlas.

2.4 Discussion

A number of methodologies for generating sCT scans from MR images have been identified. Using water equivalent homogenous overrides for the entire patient volume, whilst simple, gives unacceptable dosimetric results in some cases, for example when the beam passes through an air cavity in the head (65). It is not possible to use these images as references for verification of patient positioning. Bulk density overrides can be used by separating out different tissue classes. Whilst this can give better dosimetric results, the need to segment bone, which is carried out manually in the majority of cases, makes this technique unappealing. It is possible that the use of automatic segmentation techniques seen particularly in atlas-based (85) and some voxel-based techniques, such as Koivula et al., (103), would aid clinical implementation of bulk density overrides for some anatomical sites.

Atlas-based techniques are promising methods for MRI-only planning. They can be fully automated and use routine MRI sequences. The techniques can be carried out using a single MRI sequence, ensuring scan time is kept to a minimum and reducing chances of patient movement. The techniques have been shown to produce results with good geometric and dosimetric accuracy for prostate and brain patients, particularly when multiple atlases are used, with dose deviations typically reported below 1%. It is feasible to produce accurate reference images for treatment verification, provided accurate registration between the atlas and

incoming MR image is achieved. Additionally, the use of an atlas means that structures can be contoured automatically; a process which may result in a reduction of contour variability and improvement in clinical efficiency.

The drawbacks of atlas-based techniques largely lie in their ability to handle patients with atypical anatomy. Uh et al., (89) noticed larger errors in atlas deformation in cases where patients had a large tumour volume or surgical void. Use of a single atlas alone has been found to give unacceptable dose deviations (27). This is to be expected as a single atlas would be unable to handle atypical anatomy. The ability of the technique to generate an accurate sCT depends on the accuracy of the registration techniques used (89). This uncertainty in image registration, particularly for patients with atypical anatomy is a concern (93, 109). The quality of the MRI scans, which need to have a FOV large enough to encompass the entire body contour is also important. The need for multiple pairwise registration of images is computationally intensive (85). Johansson et al., (94) commented that atlas-based techniques, although considered robust in terms of average pixel intensity, are associated with geometrical uncertainties particularly outside of the head region.

It can be seen (Table 2-3) that a number of different atlas sizes have been employed. Siversson et al., (90) suggested that there is limited benefit in increasing the atlas size beyond 15 patients, however there does not seem to be a consensus in the literature. Andreasen et al., (91) used an atlas pre-selection process, excluding highly dissimilar atlases before the patch search. These authors commented that the optimal number of atlases to select would vary depending on the similarity of the atlases to the incoming MRI (91). The optimal number of atlases may therefore be site specific. Further investigation into appropriate atlas numbers should be a focus of future work.

Voxel-based techniques have been shown to produce clinically acceptable geometric and dosimetric results. As with atlas methods, dose differences typically below 1% have been reported and the production of accurate reference images for IGRT has been shown to be feasible. These techniques have

developed in recent years with the integration of UTE sequences, which have made automatic classification of cortical bone possible. These methods have the ability to better handle patients with atypical anatomy, due to not being reliant on an atlas. There is also no requirement for accurate registration of a new incoming MRI scan, although accurate registration is normally essential during the learning steps.

One drawback of voxel-based methods is the use of multiple sequences. These improve tissue classification but result in a longer scan time, increasing the potential for patient movement. This may also lead to the scans needing to be accurately co-registered. Additionally, methods which rely solely on standard MRI sequences often require some manual contouring of bone or airways which would limit their use in the clinic.

A large proportion of voxel-based methods use non-standard sequences, such as UTE, however these are often not in routine clinical use, particularly in radiotherapy departments. UTE is associated with poor image quality (122, 123) and streak artefacts which become more severe outside of the head region meaning application to other sites may prove difficult (109). Areas such as the nasal septa continue to be problematic with voxel-based techniques. Improvements of the UTE sequence may however occur in the future.

It is noteworthy that across the techniques there is inconsistency in the criteria used to evaluate sCT quality and accuracy. This issue needs to be addressed in order to aid method comparisons. This could be achieved, for example, with the consistent reporting of average absolute deviations in HU values and by calculating dose deviations and DVH parameters for volumes of interest within the patient for dosimetric studies, when comparing sCTs to CTs.

The problem is apparent when comparing dosimetric agreement between sCTs and CTs between different studies. Many studies used gamma analysis (39) to evaluate similarity in dose distributions. However gamma analysis pass rate is dependent on a number of factors, including dose and distance-to-dose

agreement criteria, the percentage dose below which points are excluded from the analysis, whether global or local gamma analysis has been carried out and whether it has been performed in 2 or 3 dimensions. This number of variables makes a direct comparison of different studies difficult. Andreasen et al., (91) suggested for example that for their gamma criteria (dose difference = 1%, distance-to-dose agreement = 1 mm, 10% dose threshold, 2D global gamma analysis) an average pass rate of 97% is acceptable clinically, however a value of 94% should be questioned. This however would only apply to this specific criteria.

Edmund and Nyholm (3) have discussed the difficulty in comparing methodologies even in the case where the same metrics and parameters are being reported. Patient selection and exclusion criteria, as well as the amount of data pre-processing will affect the reported results (3). Aspects such as treatment technique, beam quality and target and OAR variability will affect dosimetric results.

The mean absolute error (MAE) of sCT HUs compared to those of the real CT is a common reporting metric for sCT quality analysis. However, it should be used with caution as it is influenced by which voxels are included in the comparison. For example, if voxels outside the body are included this will likely result in MAE which suggests better results than in the body alone, the latter being the only area of clinical interest. Additionally, including bowel gas can skew results, leading to poorer results than in reality. Bowel gas is not consistent between MRI and CT scanning, however it is unlikely to be present in the same anatomical region during treatment. Reporting techniques as in the study by Demol et al., (27) whereby the error across the whole HU range is shown, would also be useful.

Andreasen et al., (91) impressed the need to assess MAE values for each site separately. For example, brains sCTs generally have higher MAE values than prostate sCTs. This is due to the differing amount of soft tissue relative to air and bone (91). Edmund and Nyholm (3) reported that for prostates, the typical MAE was around 40 HU, however for brains it was in the range of 80 - 200 HU. The

current review supports this, although it should be noted that some lower MAEs have recently been reported for brains; Koivula et al., (103) reported MAEs for the head which were similar to those reported for prostates.

Some centres have implemented MRI-only planning clinically for specific sites. The Helsinki, Finland group have treated 400 prostate patients with a dual regression approach and the Michigan group have treated brain patients using a probabilistic approach (3, 96, 112). Additionally, centres in New York, USA and Turku, Finland have recently started using a commercial solution developed by Philips for clinical sCT generation for prostate patients (124). An Australian group are running a trial involving 25 prostate patients, where planning is performed on the sCT (125).

The growing enthusiasm for MRI-only planning solutions is linked to the development of MRI-guided radiotherapy treatments. MRI-linacs are being developed worldwide, meaning that MRI-only planning would be beneficial for full on-line plan adaption. The majority of clinics will not have access to an MRI-linac in the immediate future, however increasing access to MRI-simulators and MRI diagnostic scanners would mean that many centres would be able to benefit from the advantages of MRI-only planning. The possibility of performing IGRT using either 2D or 3D image verification in an MRI-only radiotherapy workflow, without an MRI-linac, has been demonstrated in the reviewed literature. In order to use online adaptive planning, it would be desirable for sCT production time to be as short as possible. In the literature, where studies have reported sCT generation times, they have been in the range of 1-6 minutes for voxel-based techniques, a few minutes for average atlas studies and up to 80 mins for multiple atlas studies. Many studies do not report on the generation time; this should be included to assess the clinical suitability of the method.

2.5 Conclusion

A systematic review has been performed to identify methods of sCT generation for MRI-only radiotherapy. Three main methods have been identified, with atlas-

based and voxel-based techniques being the most clinically useful. Some of these studies have reported dosimetric differences of <1% between planning on the sCT and CT and <1 mm deviations when using sCTs for positional verification. It is recommended that a consensus regarding the preferred metrics for reporting on the quality of sCTs should be reached.

2.6 Update to the literature review

Since the initial literature search was performed further studies regarding sCT generation methodologies have been published. The main significant development in the field has been the emergence of deep learning based techniques which have the advantage of not requiring interpatient registrations. The application of the technique to the generation of sCTs was first proposed by Han (126) where deep convolutional neural networks (CNNs) were used to perform mapping between the MRI and CT of brain patients, allowing generation of sCTs. High quality sCTs (average MAE of 84.8 ± 17.3 HU) were produced using this method, however a dosimetric assessment of using the sCTs compared to CTs was not carried out in this study. Additionally the function converted a 2D MRI slice to a 2D sCT slice, rather than creating a 3D neural network model due to the large computational memory the latter would require. This has the disadvantage that 3D spatial differences across organs are not considered in the model (127) and discontinuities between slices could exist.

Other studies, such as those conducted by Emami et al., (128), have employed generative adversarial networks (GANs) to generate sCTs. With this method a discriminator network is used alongside the generator network, to attempt to distinguish between synthesised and real data. The method was compared to using a CNN method with superior quality sCTs reported, although again no dosimetric assessment was carried out.

Recently, studies have begun to emerge which assess the dosimetric differences of using pelvic and brain sCTs generated using CNNs and GANs compared to using CT scans (129, 130). These have reported generally acceptable dosimetric

deviations with results in-line with those reported in other studies i.e. dose differences of less than 1% for pelvic studies. The sCT generation for an individual patient was fast and of the order of seconds, which would be an advantage for online adaptive MRI-only radiotherapy. It should be noted however that neither study used a full 3D model, due to the high computation requirements this would require. In general, deep learning methods have a high computational cost and require large databases for training (127). Additionally, it is unclear how sensitive the techniques are to differences in sequence acquisition (127).

The use of deep learning in the generation of sCTs has recently emerged as a promising technique. The main drawback appears to be the high computation cost associated with the technique, however it is likely that more studies will be carried out in this area in the future and the technique may be a candidate for future clinical use. This is especially true when considering the use of sCTs for MRI-guided radiotherapy, where the production of high quality sCTs in a short timescale is particularly pertinent.

2.7 Concluding remarks

This review has shown that the generation of sCTs for use in MRI-only planning is currently a subject of great interest, with a number of different techniques showing promise. These include atlas-based and voxel-based techniques, with deep learning techniques also having recently been used with acceptable results.

Additionally, this review has outlined the discrepancy in reporting metrics related to the quality of sCTs between different studies. Recommendations on metrics which should be reported consistently by the community should be the focus of future work.

Chapter 3 Assessment of brain synthetic CT images

3.1 Introduction

As can be seen from Chapter 2, atlas-based techniques are a promising method for MRI-only radiotherapy planning. These techniques can be fully automated and commonly use a single, routine MRI sequence. The method has been shown to produce results with good geometric and dosimetric accuracy, particularly when multiple atlases weighted using patch-based pattern recognition techniques are used, with dose deviations typically reported below 1%. The main disadvantages of atlas-based techniques lie in their ability to handle patients with atypical anatomy and the need for a high quality training database built using highly accurate registration techniques.

Deep learning techniques for sCT generation have also recently started to be developed and results appear promising. The main disadvantage of these techniques is the high computation cost involved in the modelling. It is also uncertain how sensitive these models are to differences in the MRI acquisition protocol and equipment.

The next step in this investigation is to further explore previously developed sCT generation methods, to determine the generalisability of the models. In other words, whether it is possible to use a method developed at a different institution with different MRI equipment and protocols, and potentially for a different anatomical site, to produce sCTs which can be used clinically.

A lack of consistency in sCT quality reporting metrics was also identified in the literature review. Therefore in this study, the HU and dosimetric sCT analysis metrics have been chosen based on findings in the review. This includes analysing mean absolute deviations in HUs within the patient contour, evaluation of this error as a function of the HU value, dosimetric comparisons involving

relevant DVH criteria and gamma analysis calculations performed within the body contour with the evaluation parameters carefully selected.

3.2 Synthetic CT generation using diagnostic MR images

3.2.1 Introduction

This project involved collaboration with the Australian research group based at the Commonwealth Scientific and Industrial Research Organisation (CSIRO) e-research centre. This group had previously developed an MRI-only radiotherapy atlas models for prostate patients (85) using their own in-house data. The aim of this project was to test the generalisability of the model i.e. the feasibility of developing a similar model for brain patients and to assess this using data from other centres.

In order to do this, CT and MRI scans of multiple brain patients needed to be acquired. Currently at Leeds Cancer Centre (LCC), brain patients undergo a routine post-surgery MRI scan, which aids in target and OAR delineation during radiotherapy planning. This scan is however not acquired with the patient in the treatment position i.e. an immobilisation mask is not used (as this is produced later in the radiotherapy pathway) and no head rest is present. The FOV of the MRI scan may also be smaller than the FOV of the planning CT, as the main purpose of the MRI is to display the tumour and OARs, rather than the whole body contour. It was investigated whether these scans, along with the corresponding planning CTs, could be used to develop atlas models for brain MRI-only planning. If not, adjustments to the scanning protocol would be required.

3.2.2 Method

Ten meningioma patients from LCC who had planning CT and diagnostic MRI scans acquired for the purpose of radiotherapy treatment planning were used in the initial study. Ten stereotactic radiosurgery (SRS) patients from the Northern Centre for Cancer Care (NCCC) were also included in order to increase the number of patients in the study. SRS patients were used from NCCC as these

are the only brain patients who routinely receive a post-surgery MRI scan at this centre.

At both centres, diagnostic CT scans had been acquired on a Siemens Somatom Sensation Open scanner (Siemens Healthineers, Erlangen, Germany) with a 2 mm slice thickness at LCC and 1 mm slice thickness at NCCC. MRI scans had been acquired at NCCC on a 1.5 T Siemens Magnetom Espree and at LCC on a 1.5 T Siemens Magnetom Avanto, Aera and Symphony. At NCCC a T1-weighted post-contrast 3D VIBE sequence was used with 1 mm slice thickness and at LCC a T1-weighted post-contrast 2D spin echo sequence was used with 2 mm slice thickness. Vendor specific 3D distortion correction was applied to the NCCC scans and 2D distortion correction to the LCC scans.

At LCC the diagnostic MRI scans and CT scans were exported from the Monaco (Elekta AB, Stockholm, Sweden) treatment planning system (TPS) and the scans were anonymised. For each patient, the MRI and CT scans, along with the associated structure sets, were rigidly registered using Mirada RTx (Mirada Medical, Oxford, UK). The ten NCCC MRI and CT scans were similarly anonymised and exported from the TPS. The MRI and CT scans were rigidly registered using in-house software at the Australian e-research centre, Brisbane, Australia.

The multi-atlas technique developed by Dowling et al., (85) to produce sCTs for prostate MRI-only planning was used in this study. This has been described previously in detail (Chapter 2). In brief, CT and MRI scans are co-registered to form a multi-atlas. Each of the atlas MRI scans is registered to the new incoming MRI scan and these deformations are then applied to the corresponding atlas CTs. A patch of voxels in the incoming MRI is compared to a patch in the corresponding location in each registered MRI and a similarity coefficient metric (the sum of square differences) is computed. This metric is then used to weight the contribution that each co-registered CT atlas has to the HU value of the voxel in the final sCT.

A multi-atlas database was created for each cohort of patients (two cohorts from the two centres with ten patients in each, i.e. one for the meningioma and one for the SRS patients). sCTs were created using a leave-one-out approach and therefore there were 9 patients in each multi-atlas. The similarity coefficient was calculated based on a patch size radius of 4 voxels.

The sCTs were compared to the planning CTs through a Hounsfield unit (HU) value comparison. For each patient the CT was subtracted from the sCT using the open source 3D Slicer software (version 4.5.0-1, http://www.slicer.org) (131). The mean error (ME) and mean absolute error (MAE) inside the patient contour were determined using Equation 3-1 and Equation 3-2 respectively, where $HU(sCT)_i$ and $HU(CT)_i$ are the HU values of the ith voxel in the sCT and CT respectively and N is the total number of voxels. The standard deviation (SD) of the errors and the absolute errors were also calculated for each patient. These values were determined using an in-house script written in the GNU Octave environment (version 4.0.0, http://www.octave.org) (132).

$$ME = \frac{\sum_{i=1,2,3....}^{N}(HU(sCT)_i - HU(CT)_i)}{N}$$

Equation 3-1

$$MAE = \frac{\sum_{i=1,2,3....}^{N}|HU(sCT)_i - HU(CT)_i|}{N}$$

Equation 3-2

The mean ME and MAE for each patient cohort were determined. The standard error of the mean (SEM) was calculated for each using Equation 3-3, where n is the total number of sCTs analysed in the model and sd is the standard deviation of ME/ MAE values.

$$SEM = \frac{sd}{\sqrt{n}}$$

Equation 3-3

The similarity between the sCT and CT images for each patient was also assessed using gamma analysis (39) (see Chapter 1). Typically with this method, a dose point in a certain location in a plan is compared to the dose in the associated location in the second plan. In order for the point to pass the dose comparison the dose point must either be within a certain percentage dose difference of the equivalent point in the second plan or within a certain distance of the same dose in the second plan. In this investigation however, gamma analysis was performed in terms of HU values, rather than dose.

In this study the criteria of 3% difference in HU and 3 mm distance to agreement were used, where the percentage is with reference to the maximum HU in the planning CT. A criterion of 3% of the maximum HU was chosen, as this corresponded to approximately 100 HU, which has been suggested as an acceptable HU deviation (86). A relatively large spatial criterion of 3 mm was selected in order to identify gross errors in this feasibility investigation. This analysis was carried out inside the patient contour using the radiotherapy dose comparison module in the 3D Slicer software (http://www.slicer.org) (133).

3.2.3 Results

An example of slices through a sCT of a representative meningioma case can be seen in Figure 3-1 along with the corresponding planning CT and MRI slices. A misclassification of soft tissue as bone towards the inferior section of the brain can be seen in the sCT, as well as blurring around air cavities.

Upon inspection of the subtraction images (the planning CT subtracted from the sCT) large deviations (greater than 1000 HU) were located at skull – soft tissue and soft tissue - air interfaces suggesting issues with registration and a blurring of anatomical information in the atlas generation process. An example of a subtraction image is shown in Figure 3-2.

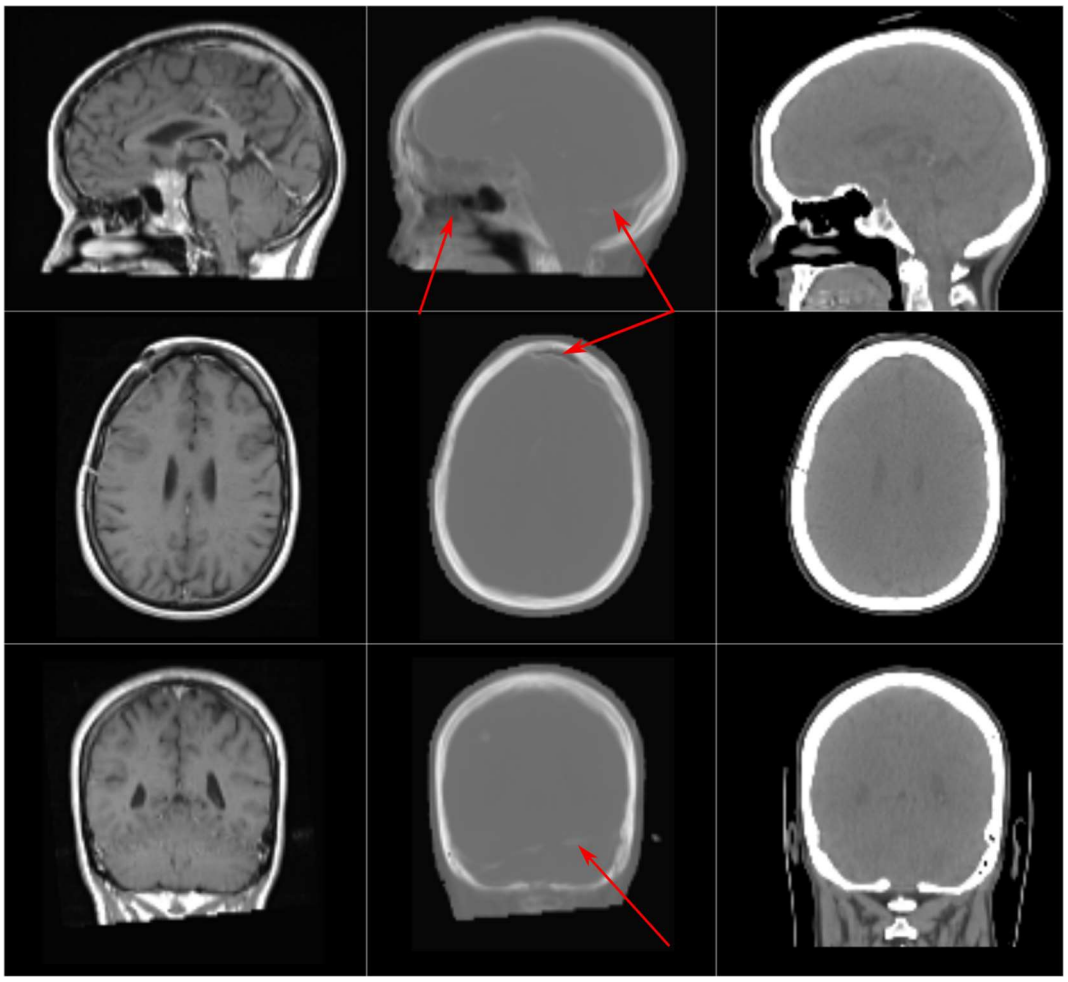

Figure 3-1 Showing representative slices in three planes through a meningioma sCT generated using the multi-atlas technique (middle column) with the corresponding slices through the MRI scan (left column) and the planning CT scan (right column). The red arrows show the areas where soft tissue has been misclassified as bone and where there is blurring around air cavities.

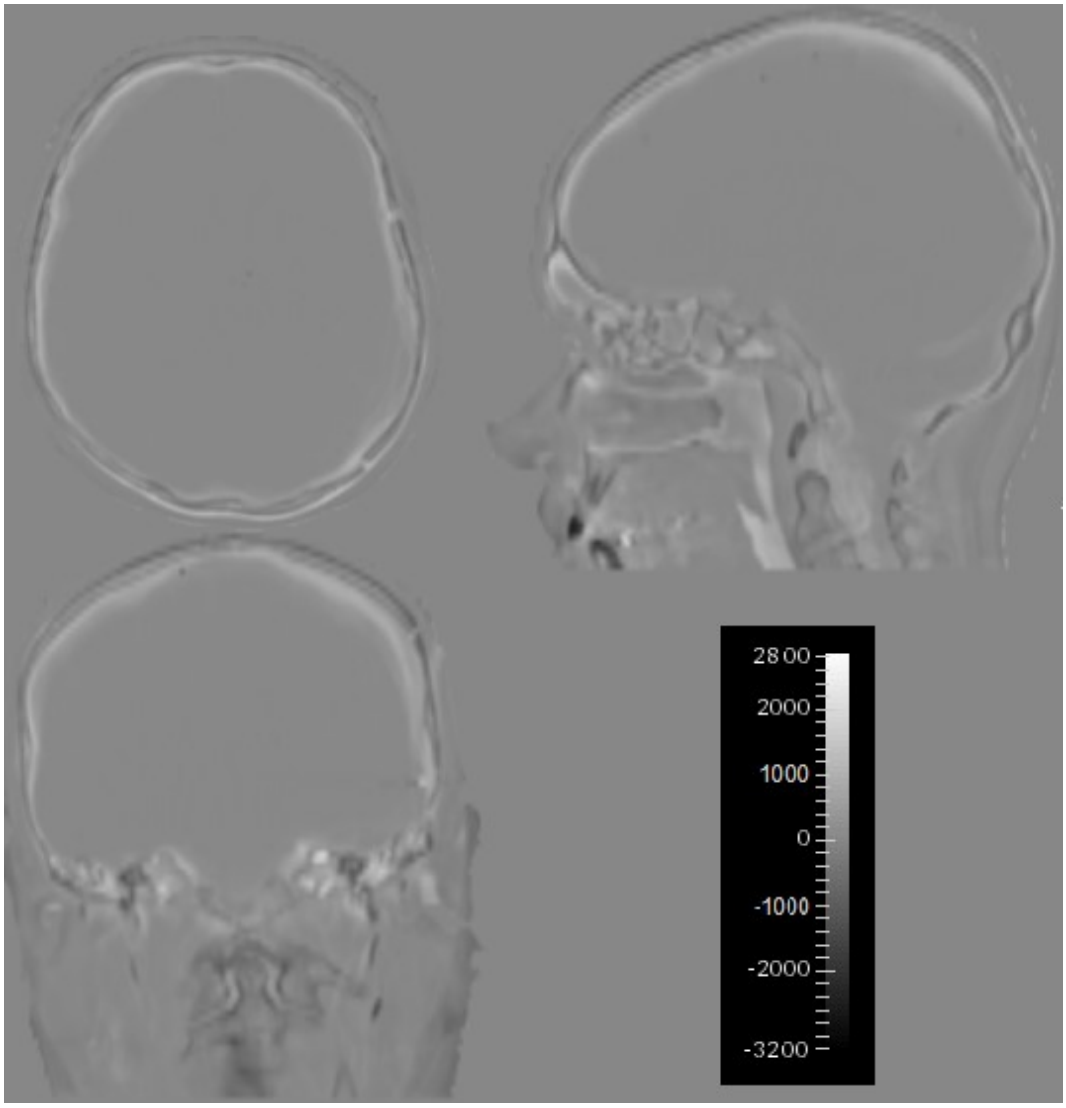

Figure 3-2 An example of a subtraction image (planning CT subtracted from sCT) for a meningioma patient showing slices in three planes. The scale in Hounsfield units is shown at the bottom right of the image.

A summary of the average ME and MAE inside the patient contour for both cohorts of patients can be seen in Table 3-1. The gamma analysis pass rate is also shown.

	Mean ME/ HU	Mean MAE/ HU	Mean gamma pass rate/ %
Meningiomas	-43.1 ± 9.8	170.4 ± 10.4	79.1
SRS	-1.6 ± 13.0	175.0 ± 3.8	87.0

Table 3-1 The average (mean) mean error (ME), mean absolute error (MAE) and gamma analysis pass rates inside the patient contour for both cohorts of patients. The errors represent the standard error of the mean.

3.2.4 Discussion

Agreement between the sCTs and planning CTs was generally good in the brain tissue, however discrepancies between the two appeared at air - soft tissue and bone - soft tissue interfaces as illustrated in Figure 3-1 and Figure 3-2. The blurring around these areas in the sCT, as well as some incorrect classification of bone, could potentially lead to both dosimetric issues during RTP as well as problems with image alignment during on-treatment positional verification.

The average MAEs are higher than have been reported for prostate sCTs produced using the same model (85) where a value of 40.5 ± 8.2 HU was found. The average ME for the SRS data was comparable to the average ME of the prostate data reported in Dowling et al., (85) of 0.6 ± 14.7 HU. However in the literature, reported MAEs are typically higher for brain images than prostate images due to the different ratios of soft tissue and bone between the two. For example, MAE values between 35 HU and 200 HU have been reported (3, 62). The results presented here for both datasets are consistent with the literature, however they lie towards the higher end of the reported values. It can be seen from Table 3-1 that there is an offset in the meningioma dataset where the sCT voxel intensity measured lower than the intensity in the planning CT. The low gamma pass rates in Table 3-1 are indicative of a poor HU agreement.

This study was limited by the number of patients in each multi-atlas, which was equal to 9 patients. This means that it is unlikely that the patient anatomy of a large enough number of patients has been modelled (134, 135). Additional issues were caused by the MRI scans having been acquired for purposes other than MRI-only radiotherapy planning. The meningioma dataset, for example, contained inconsistent fields of view (FOVs) between the MRI scans which, as well as causing problems during the atlas generation process, also meant that in some MRI scans anatomy, such as the nose, was missing. Whilst this is not a problem for imaging the tumour and OARs, information about this anatomy is needed for radiotherapy treatment planning to determine the beam attenuation and calculate the dose distribution. Additionally, this information is also needed for generating accurate atlas models. Interleaving artefacts were present on some scans and the images consisted of 2D non-isotropic voxels, which are less suited than 3D image acquisition for registration purposes. The patients had not been imaged in the treatment position during MRI scanning which most likely contributed to MRI - CT registration problems.

As a result of this work changes to the MRI scanning protocol for brain patients were implemented. Patients were scanned on a head rest so that the position of the back of the neck was more consistent between patients, as well as being closer to the planning CT scan position, which should lead to a more accurate registration. Whilst resulting in a more similar position, it should be noted that the patient was still not imaged in the exact treatment position, as they were not scanned in the immobilisation mask and the generic head rest differed from the radiotherapy treatment head rest. This was due to practical reasons in the clinic, as the MRI scan is currently performed before the immobilisation mask has been produced, and the base of the head rest was modified to fit in the MRI head coil. Due to the lack of immobilisation mask, it should be ensured that no motion artefacts are present in the MRI training images.

Additionally, the FOV of the MRI scans was increased in order to include all necessary anatomy such as the nose, as well as being more consistent between

patients and ensuring that no wrap around artefacts are present (see Chapter 1). A scan of a meningioma patient imaged using the new MRI protocol, alongside a patient imaged with old protocol for comparison, can be seen in Figure 3-3.

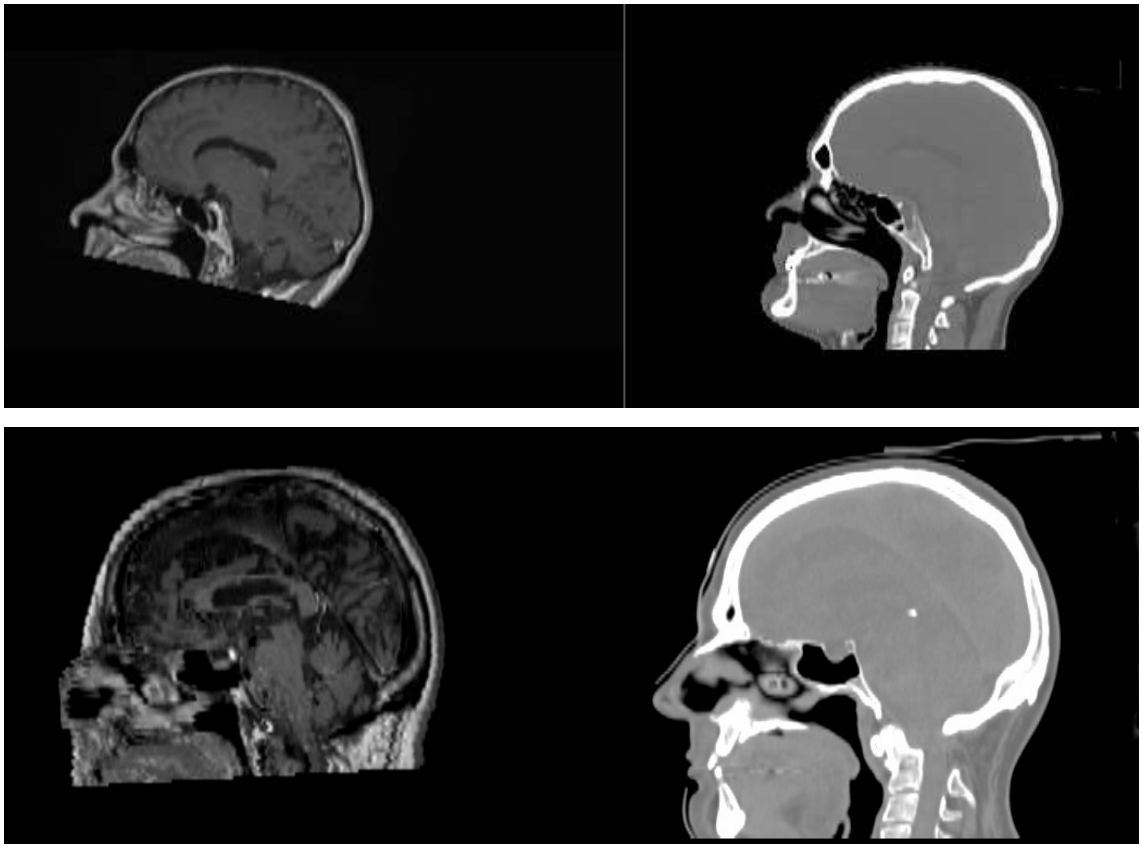

Figure 3-3 Example of an MRI slice acquired using the new protocol (top left) with the corresponding planning CT slice (top right). All relevant anatomy has been included in the MR image and the patient has been scanned on a head rest. This can be compared to an MRI slice acquired with the patient in a diagnostic position (bottom left) with the corresponding planning CT slice (bottom right).

3.2.5 Conclusions

The aim of this work was to test the feasibility of generating sCTs for brain patients using a multi-atlas technique previously developed for prostate MRI-only planning. Only routine MR images, which had not been optimised for MRI-only radiotherapy, were used in the generation of the atlases, however this led to poor quality sCT scans, including areas of misclassification of bone and blurring around air cavities. In order to improve the quality of the sCT scans, changes were made to the MRI scanning protocol as a result of this study, specifically

imaging patients on a head rest and increasing the FOV. Additionally, head motion should be kept to a minimum, in order to reduce the likelihood of artefacts.

The aim of future work is to evaluate sCTs produced using MR images which have been prospectively acquired using the new protocol. This will include a HU comparison, as well as a dosimetric comparison investigating differences when planning on the sCT scans with planning using the CT scans. This investigation is described in Section 3.3.

3.3 Synthetic CT generation using MR images acquired prospectively using a new protocol

3.3.1 Introduction

It was hypothesised that the main issues with using the routine brain MRI scans for sCT generation in Section 3.2 was firstly the difference in the neck position between the CT and MRI scans. This contributed to poor registration during atlas training, resulting in blurring and incorrect tissue assignment. Patients attending for their planning CT scan and radiotherapy treatments are positioned inside fitted immobilisation masks and on a head rest. This allows the neck to be kept in a reproducible position during each treatment fraction, as well as in the planning scan sessions. Within the current radiotherapy pathway, patients attend for their MRI scan prior to their immobilisation mask being fitted, and therefore it was not possible in this study to scan patients in their true treatment position within the mask. However an MRI-compatible head rest was produced, which was similar to those used in CT scanning. This enabled patients to be scanned with their neck in a similar position to when CT scanning was performed.

Another difficulty with the routine MRI scanning was that frequently the FOV was not large enough to include all the patient anatomy necessary for MRI-only planning i.e. the full head and neck contour. For example, commonly missed areas were noses, ears and the superior section of the skull. These areas do not need to be included in MRI scanning for routine radiotherapy planning, as they

are not essential for either a diagnosis or for target or OAR delineation. However it is essential that these are included in the FOV for MRI-only radiotherapy, in order to accurately calculate the dose deposited inside the patient. Additionally, missing anatomy in the MRI atlas, that is present in the CT atlas, will lead to a poor quality atlas, and hence contribute to poor quality sCTs. A consistently large FOV in the MR image should also lead to a higher quality atlas compared to a varying size FOV.

In order to ensure that these changes to protocol were introduced an electronic message was created in the MRI scanner systems under the brain radiotherapy protocol. This reminded the MRI radiographers to image all patients scanned using this protocol on the head rest provided, as well to include all relevant head anatomy in the scan and to use a large FOV which included the superior section of the skull.

Following some continued challenges with the implementation of the new protocol, it was found that there was a problem in that some of the brain patients were being referred for the incorrect MRI brain scan type after surgery, therefore meaning that the radiographers did not see the message prior to scanning. Once this was resolved, a rise in compliance with the new protocol was observed. This was an important learning point in the importance of taking the entire patient pathway into account when considering the implementation of an MRI-only workflow.

3.3.2 Method

3.3.2.1 Preparation of images

Ethical approval was granted for this study (IRAS number 183964). This approval allowed anonymised MRI and CT data from consented patients to be shared with the CSIRO e-research centre, as well as the company Spectronic Medical AB, Sweden in order to develop and assess sCT generation models.

Meningioma and glioma patients who had been referred for radiotherapy were identified. The MRI scans of all patients who met this criteria were viewed to see if the scans were consistent with the new protocol i.e. the patients had been scanned on a head rest and all relevant anatomy had been included in the FOV. It was also ensured that there were no large motion artefacts in the MRI scan. Only MRI scans acquired with a T_1-weighted spin echo sequence (post gadolinium contrast administration), with a 2 mm slice thickness were approached for inclusion in the study, in order to maintain consistency in the training database. Patients who met this criteria were approached for consent to the study. In total 16 patients were consented.

Once a patient had consented to be included in the study, they were assigned an anonymised name, patient ID and date of birth. The MRI and CT scans were then anonymised, with all patient-identifying tags replaced with the anonymised versions. A key was produced linking each patient with their anonymised ID. The anonymised MRI and CT scans were brought into the Oncentra Masterplan treatment planning system (Elekta AB, Stockholm, Sweden) as separate cases associated with each anonymised patient. For each CT and MRI scan the patient contour was outlined, ensuring that all patient anatomy was included, but that no couch or immobilisation mask was delineated.

These were then exported from Oncentra Masterplan to the Mirada RTx system. For each patient, the CT scan, along with the associated patient contour, was registered to the MRI scan. This was achieved through automatic rigid registration, followed by manual rigid registration if required. Rigid registration was used so as to maintain the geometric integrity of the CT image and, as the images were of the cranium, there was not expected to be any problem resulting from internal deformation.

The patient images and external contours were then shared with the CSIRO e-research centre and Spectronic Medical AB. The MRI and CT scans were used to build a sCT brain model by CSIRO as described in the work by Dowling et al., (85) using a multi-atlas patch-based technique (see Chapter 2 for further details).

This technique had been originally developed for prostate patients. The Spectronic model was a pre-release of the commercial MRI Planner 2.0 software (Spectronic Medical AB, Helsingborg, Sweden), which is a neural network-based method. This method identifies features in an MRI image and, through training with MRI and CT databases, converts the MRI into a sCT scan. The model has been created as non-site specific i.e. with the aim that it can be used for any anatomical site.

All 16 patients were used in the Australian multi-atlas patch-based model using a leave-one-out approach, so that 15 patients were used in the atlas for each generated sCT. This is an increase in the previous atlas size of 9 patients and the number appears acceptable as Siversson et al., (90) suggested that there is little benefit in using an atlas larger than 15 patients. In the Spectronic model, data from other patients had already been used to create a preliminary model. However 13 patients from LCC were also added to the model in order to develop the training dataset and aid the generation of high quality sCTs for LCC patients. This meant that 3 patient scans could be used to generate sCTs for unbiased testing of the model.

Due to the different number of patients tested using both the Australian multi-atlas and Spectronic neural network model, as well as the different number of patients in the training dataset, it is not the aim of the study to compare the two different sCT generation techniques. The main purpose is to report on the clinical feasibility of using each technique for the generation of brain sCTs, and whether for each, there are any additional adjustments needed before they could be implemented clinically. Additionally, by examining two different techniques, a consistent reporting protocol can be development for reporting the quality metrics of sCTs. In Chapter 2 this was identified as an area which is missing from the current literature.

3.3.2.2 Synthetic CT assessment

Firstly it was ensured that all sCTs were assigned the correct anonymised patient ID, name, date of birth and sex. It was found that the sCTs from the Australian

multi-atlas model were assigned a random value for each of these, and therefore the DICOM tags were manually changed. This was reported back to the centre. This step was not necessary for the Spectronic neural network sCTs.

It should be noted that the sCTs are returned in the same FOV as the original MR image. This is as expected in an MRI-only workflow where the MR image is used as the only input. Additionally, this means that the sCT and MRI can be viewed in the same frame of reference during the planning process.

3.3.2.2.1 Hounsfield unit (HU) comparison

A useful method for assessing the difference between CT and sCT scans is to perform a HU value comparison. The reporting of ME and MAE allows a determination of any offset in the assigned HU values and the overall level of error in tissue assignment respectively. Therefore for each of the generated sCTs the ME and MAE were determined using Equation 3-1 and Equation 3-2 respectively.

The CT, in the same frame of reference as the MRI, and sCT were first masked using the patient outline on the MR image. The MRI patient contour was used as this was the same position from which the sCT had been generated. In order for the analysis to be fair, the same outline was therefore used to mask the CT image. In the open source Slicer software, the sCT and CT were each multiplied by the binary patient mask from the MRI. The masked CT was then subtracted from the masked sCT, resulting in a HU subtraction image masked with the MRI patient outline. In order to ensure that voxels outside the mask were not included in further analysis, the subtraction image was divided by the MRI binary patient mask. This resulted in a true subtraction image within the body contour, with voxels outside the contour being assigned “nan” values or not a number.

An in-house script in GNU Octave was written which determined the ME and MAE of the voxels inside the body contour. To determine the ME, all subtraction voxels where the corresponding mask voxel value was not equal to zero, were summed together. The number of these voxels was counted, and was confirmed to be

equal to the number of voxels equal to 1 in the mask image. The ME was then calculated using Equation 3-1. The MAE was calculated in a similar manner, but with the subtraction image first converted to an absolute subtraction image. The final MAE value was determined using Equation 3-2.

The average ME and MAE for each generation methodology were also determined. The standard error of the mean (SEM) for each was calculated using Equation 3-3.

As stated in Chapter 2 reporting a single figure for ME and MAE may not be the most appropriate or insightful measure of sCT quality. This is because the single figure, although useful in gaining an insight into the approximate level of deviation between the CT and sCT image, can mask large discrepancies in the HU deviation, even if the SD of the errors is also reported. For example, a deviation in a part of the skull may not affect enough voxels over the entire cranium to result in a large MAE value. This may be even more pertinent when reporting the ME as, for example, an underestimation of HU in one area could be compensated for by an overestimation of HU values in another.

Therefore it may be useful to report the error in HU values between the sCT and CT across the range of HU values found in the original CT. This has been performed in a small number of sCT generation studies (27, 29), however it is not widely reported. Analysing the error as a function of HU value allows determination of any particular tissues in the sCT scan where there is a large deviation from the original CT image. It is hoped that this form of analysis, will be particularly useful if multiple techniques are being assessed for clinical use, to understand the typical HU error output of each over the HU range. In this study, this analysis was performed for both the ME and the MAE.

For each patient the maximum and minimum HU value in the planning CT was determined. The ME and MAE were then calculated using a similar method as described previously, but only for voxels where the CT HU value was within a certain range. The CT HU values were binned into ranges of 100 HU, starting at

the minimum HU value and continuing until the maximum HU value had been included. 100 HU was chosen as this was considered an appropriate range of HU values required to assess the error over different tissues within the body. Prior to the analysis the masked CT image was also divided by the binary patient MRI mask, in order to convert voxels outside the mask to nan values. This was performed using an in-house GNU Octave script.

The ME and MAE were then plotted as a function of HU, allowing the variation of the error to be assessed. The average ME and MAE in each HU bin were also calculated for each model, allowing plots of average HU error against HU value to be created.

As a validation, it was ensured that $\left(\sum_{i=1}^{i=t}(ME_i \times c_i)\right)/c = ME$, where ME_i and c_i are the mean error and voxel count in the *i*th bin, c is the total voxel count within the patient mask and *t* is the total number of HU bins. A similar validation was also performed for the MAE. It was also ensured that $\sum_{i=1}^{i=t}(c_i) = c$.

3.3.2.2.2 Dosimetric comparison

A dosimetric analysis of the sCTs was performed using the Monaco treatment planning system, for patients whose original plans had been created in this system. There were 12 patient sCTs within this category from the Australian multi-atlas model and 2 from the Spectronic neural network model.

The CT, planning structures and the original radiotherapy plan were exported from the clinical system and imported as a new case into a research version of the Monaco TPS. All of these patients had been treated with volumetric modulated arc therapy (6 MV energy, with either 1 or 2 arcs, with a dose-fractionation of either 60 Gy in 30 fractions (6 patients), 54 Gy in 30 fractions (4 patients), 66 Gy in 33 fractions (1 patient) or 50.4 Gy in 28 fractions (1 patient)). In this treatment the patient is treated using an x-ray beam which is moved through an arc(s) whilst varying the gantry speed, the position of multi-leaf

collimators and the dose-rate. Through this technique the dose can be sculpted around the target volume, whilst sparing normal tissue.

The sCT was first rigidly registered to the CT within Monaco TPS in order to bring them into the same frame of reference so that a comparison could be performed. The sCT was then transferred to the research version of the Monaco TPS. Two cases now existed in the planning system, one for the planning CT and one for the sCT. Structures, including target and OAR delineations and the original patient contour, were copied from the CT to the sCT, possible as they were now co-registered. As dose is only calculated inside this body contour, it was necessary to ensure that all the sCT anatomy was included, and adjustments to this contour were made if necessary. All internal contours were left unmodified. The couch top was simulated and density overrides were applied for both the CT and sCT image set, as is carried out in the clinical protocol.

The original plan was associated with the CT and sCT image. The isocentre was copied from the clinical plan to ensure that this matched the isocentre of the new plans. The dose resulting from the treatment plan was then calculated on the CT and the sCT. Monaco uses Monte Carlo techniques to calculate the dose distribution inside the patient. At the planning stage the desired uncertainty in dose calculations can be entered, with lower uncertainty coming at the expense of additional computing time. In this study an uncertainty of 1% was selected for both plan calculations, in line with local clinical protocols.

To assess an intensity modulated radiotherapy plan, it is recommended by the International Commission on Radiation Units and Measurements (ICRU) in report 83 (136) that dose-volume histograms (DVHs) are used for dose specification and reporting. Using these, the dose that a certain volume of a target or OAR receives can be reported. For clinical brain plans at LCC, dose-volume specifications have been determined which should be met in order for a plan to be used clinically. These criteria were used to assess the clinical acceptability of plans calculated on sCTs. Additionally these dose-volume parameters were recorded on both the sCT and the CT and the difference between the two

determined. In this way, both the clinical acceptability of the sCT plan and linked with this, the difference from the plan calculated on the original CT could be assessed. The DVH criteria (from the local clinical protocol) which were used for this assessment can be seen in Table 3-2.

Gamma analysis can be used to assess the dose difference between two plans, as described by Low et al., (39) (see Chapter 1). In this study, gamma analysis was carried out within the sCT patient body contour. Additionally, the gamma analysis was also calculated over the planning treatment volume (PTV) which includes the microscopic disease plus a margin to account for errors in set-up or due to anatomical movement. The gamma analysis was also assessed over the planning organ at risk volume (PRV) of dose-critical structures in the brain, which must receive a dose below 54 Gy. These are the brainstem, optic chiasm and optic nerves.

The two dose distributions, resulting from planning on the CT and sCT, as well as the structures, were exported from the Monaco treatment planning system. The gamma analysis between the two dose distributions was then calculated using the MICE toolkit (NONPI Medical AB, Umea, Sweden), with the relevant structure used as a mask. Gamma criteria of 1%/ 1mm and 2%/ 2mm were used to determine the difference in dose for plans calculated on the sCT and CT. Local gamma analysis, where the percentage dose difference is with reference to the local voxel dose, and global gamma analysis, where the percentage dose difference is with reference to the prescribed dose, were both calculated. A calculation step size of 0.1 mm, a tenth of the pixel size, was set for all gamma analysis.

Dose-volume criteria	Limit (no overlap of PTV and 54 Gy OARs)	Limit (overlap of PTV and 54 Gy OARs)
PTV D99%	>90%	>85.5%
PTV D95%	>95%	n/a
PTV D50%	± 1 Gy	± 1 Gy
PTV V105%	<5%	<5%
PTV V107%	<2%	<2%
(PTV – 54Gy OAR) D99%	n/a	>90%
(PTV – 54Gy OAR) D95%	n/a	>95%
Brainstem D5%	54	54
Brainstem D1cm^3	54	54
Brainstem mean dose	52	52
Optic chiasm D1%	54	54
R optic nerve D1%	54	54
L optic nerve D1%	54	54
L globe D1%	45	45
R globe D1%	45	45
(Brain – PTV) D10%	57	57
(Brain – PTV) mean dose	24	24
L lens D1%	6	6
R lens D1%	6	6
L cochlea D50%	45	45
R cochlea D50%	45	45
L lacrimal D1%	30	30
R lacrimal D1%	30	30
Pituitary maximum dose	45	45

Table 3-2 The dose-volume constraints taken from the clinical protocol at Leeds Cancer Centre for the planning of volumetric modulated arc therapy treatments for brain patients. PTV is the planning target volume and L and R represent the left and right organs respectively. Dx% is the dose received by the hottest x% of the volume, Vx% is the volume that receives x% of the prescribed dose, Dx cm3 is the dose that is received by the hottest x cm3 of the volume. 54 Gy OARs are the organs at risk that should receive a dose of less than 54 Gy, and are the brainstem, optic chiasm and optic nerves. When a dose-volume limit is given with reference to a percentage dose or a dose deviation, this is in relation to the prescribed dose. All doses are in Gy. Two different dose-volume constraints are given depending on whether or not the PTV overlaps with the 54 Gy OARs.

3.3.3 Results

An example of two axial slices through sCTs generated from the multi-atlas and neural network method respectively are shown in Figure 3-4. The same patient was used to create the sCT. The patient CT is also shown for comparison. The results of subtracting the CT from the sCTs for these slices can be seen in Figure 3-5.

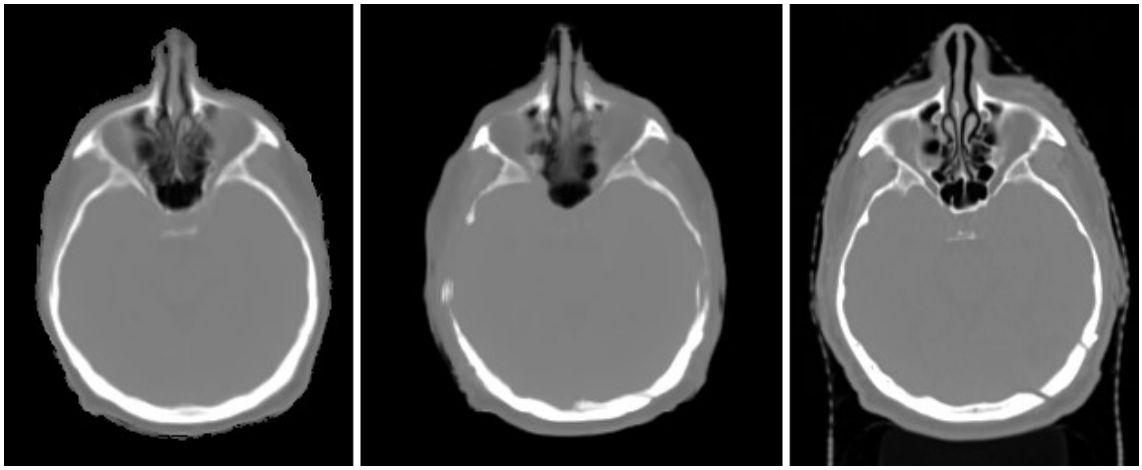

Figure 3-4 A slice through the sCTs generated by the multi-atlas method (left), neural network method (middle) and the original planning CT (right). The same slice for the same patient is shown and the same window level and width have been used for all images.

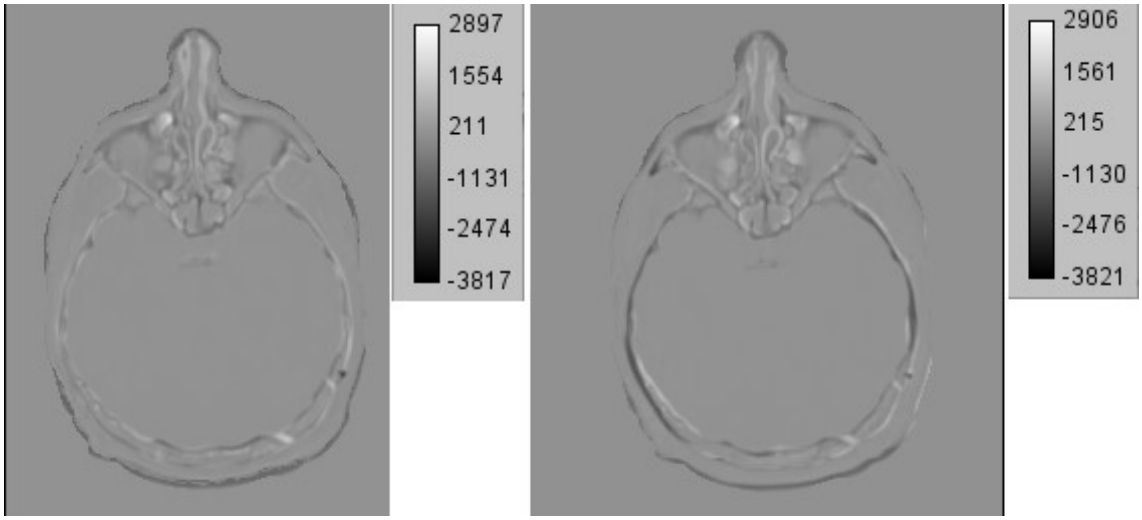

Figure 3-5 Results of the subtraction of the planning CT from the sCTs for the slices shown in Figure 3-4. The subtraction using the multi-atlas generated sCT is shown on the left and the subtraction using the neural network generated sCT on the right. The scale of each image (in Hounsfield units) is shown to the right.

3.3.3.1 Hounsfield unit assessment

The results of the average ME and MAE results for the multi-atlas and neural network generation methodology, as well as the SEM, over the entire patient volume can be seen in Table 3-3. For the ME, a positive value indicates that the HU value of the sCT voxel is greater than the HU value of the corresponding CT voxel and a negative value that the HU value of the sCT is lower than that of the CT.

sCT generation model	Average ME ± SEM/ HU	Average MAE ± SEM/ HU
Multi-atlas	-41.1 ± 4.3	129.7 ± 5.3
Neural network	-24.8 ± 7.8	114.4 ± 4.6

Table 3-3 The average mean error (ME) and mean absolute error (MAE) for sCTs generated through both methodologies. The error represents the standard error of the mean.

The averages for the two generation methods are not comparable due to the small sample size of the neural network method (3 patients) compared to the

sample size used in the multi-atlas method (16 patients). The negative offset in the ME values was consistent for all patients and for both methods.

A box-plot of the MAE values for the multi-atlas generated sCTs is shown in Figure 3-6. The minimum, maximum, first and third quartile and median MAE are indicated. This was only performed for the multi-atlas model due to the large sample size.

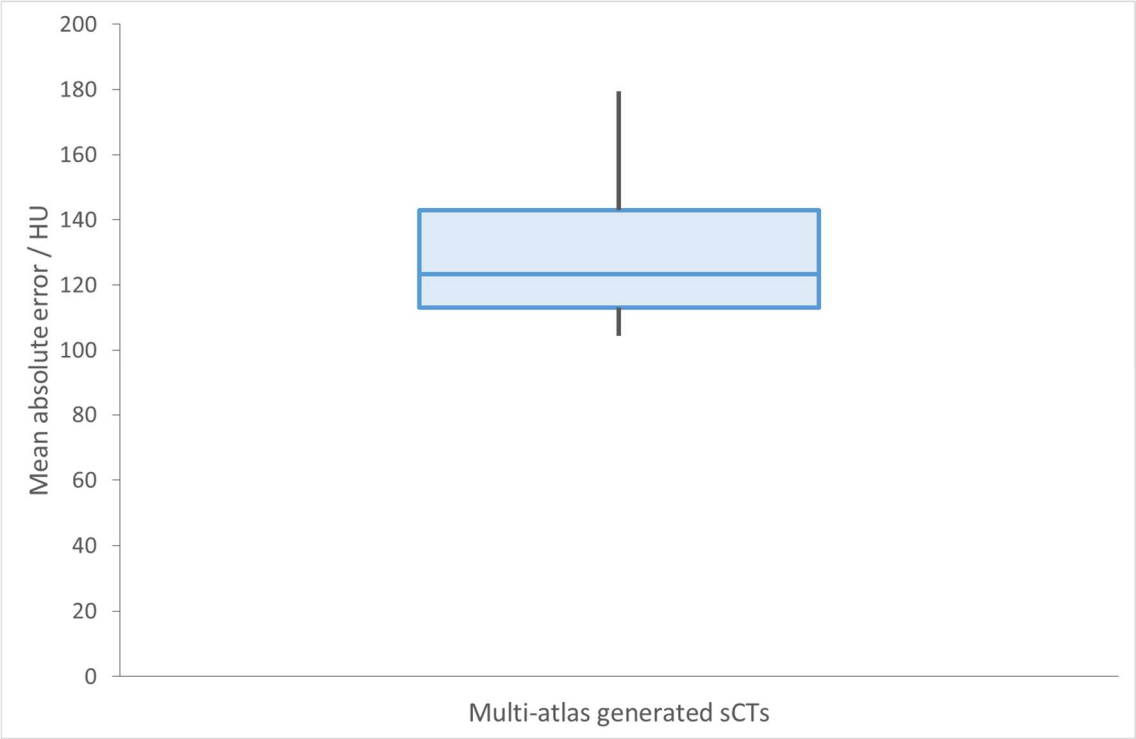

Figure 3-6 Box-plot of the mean absolute errors of the sCTs generated from the multi-atlas method.

The HU error was also assessed over the range of HU in the CT image. The ME and MAE variation can be seen as a function of HU value in Figure 3-7 and Figure 3-8 respectively. The average error variation of the sCTs generated through the multi-atlas method and of the sCTs generated through the neural network method are shown.

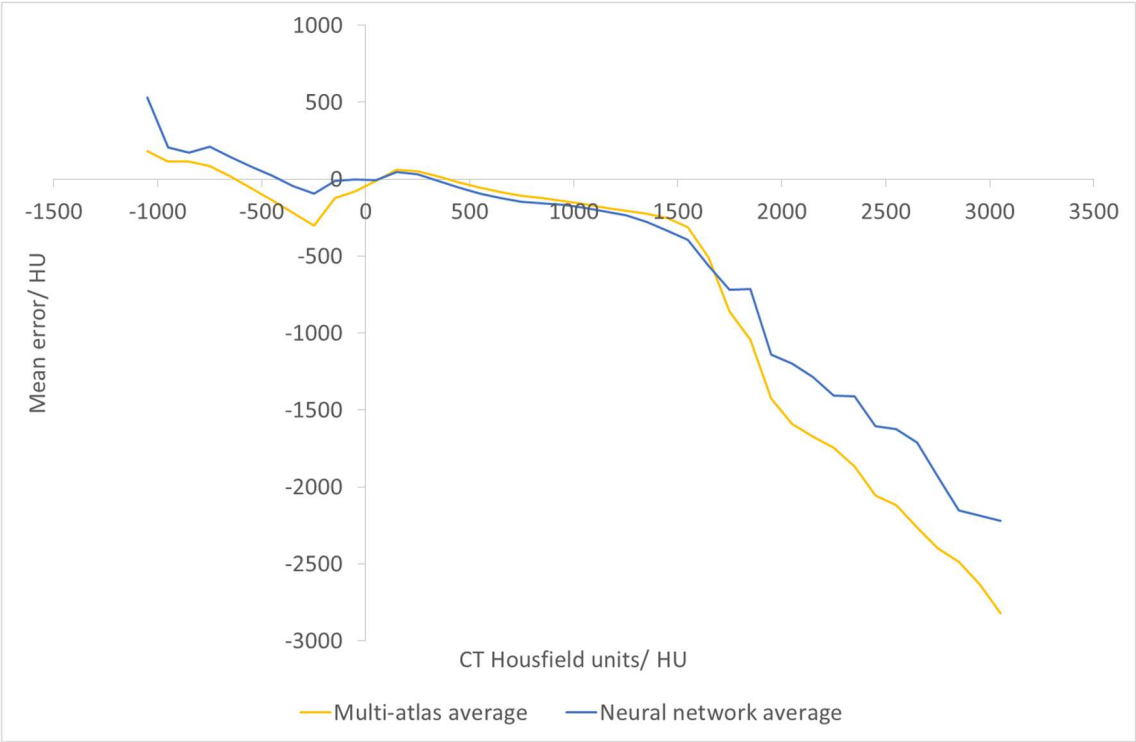

Figure 3-7 The variation in mean Hounsfield unit error (ME) against the Hounsfield unit in the planning CT scan. The average of the sCTs generated by the multi-atlas and the neural network method are each shown.

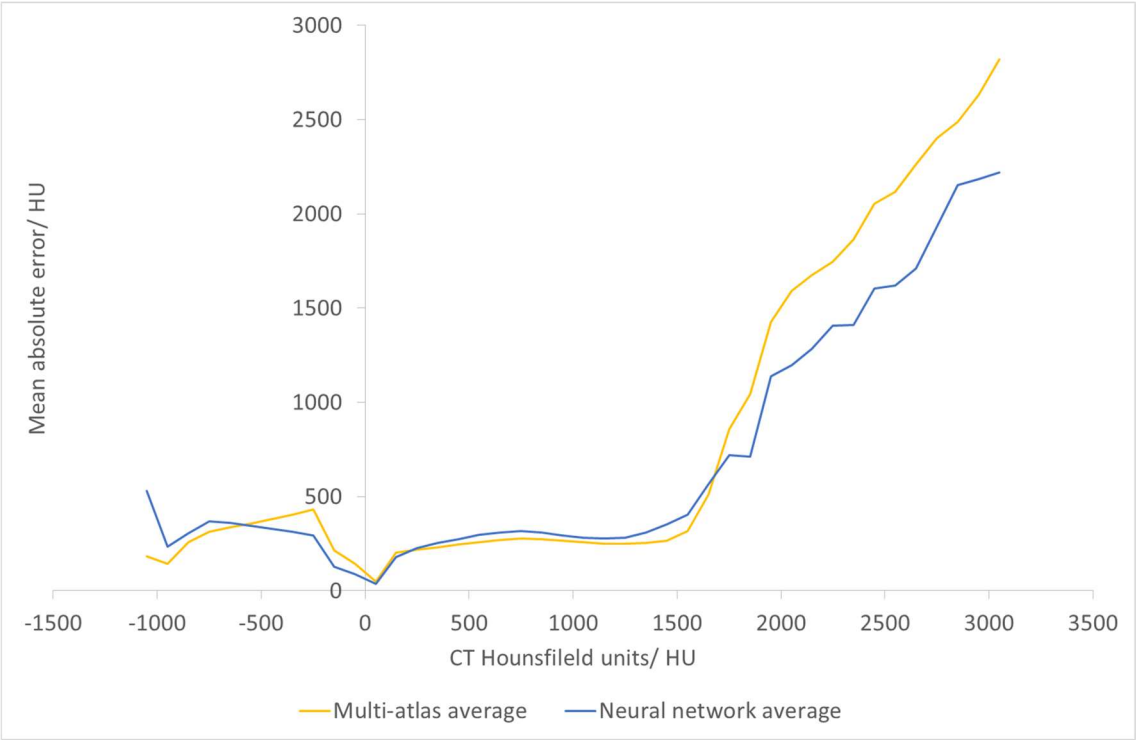

Figure 3-8 The variation in mean absolute Hounsfield unit error (MAE) against the Hounsfield unit in the planning CT scan. The average of the sCTs generated by the multi-atlas and the neural network method are each shown.

3.3.3.2 Dosimetric assessment

In order to assess the clinical acceptability of sCT scans, it is essential that the dosimetric differences between planning radiotherapy treatments on the sCT and original planning CT scan are assessed. In this investigation this was carried out by assessing the dose-volume criteria of the target and OARs as well as performing gamma analysis within certain volumes of the patient.

The results of the dose-volume constraints in Table 3-2 for radiotherapy plans created on the CT and sCT were found. The percentage difference of these constraints between the sCT and CT plans can be seen in Table 3-4. Any deviations greater than 2% have been formatted in bold. A 2% dosimetric deviation of the sCT from the CT has been suggested to be a clinically acceptable level of discrepancy in the literature (3, 73).

Dose-volume criteria	MA1	MA2	MA3	MA6	MA7	MA8	MA9	MA10	MA11	MA12	MA14	MA15	NN14	NN 15
PTV D99%	1.1%	0.5%	0.1%	0.4%	0.1%	0.8%	0.1%	0.0%	-1.6%	0.5%	0.7%	-0.6%	0.9%	-0.4%
PTV D95%	0.9%	0.4%	-0.1%	0.3%	-0.6%	0.7%	0.3%	-0.2%	-0.4%	0.3%	0.3%	-0.5%	0.6%	-0.4%
PTV D50%	1.0%	0.6%	0.8%	0.0%	0.4%	0.6%	1.3%	-0.1%	0.0%	0.2%	0.4%	-0.4%	0.5%	-0.4%
PTV V105%	**372.1%***	**245.5%**	**109.3%**	-0.9%	**87.8%***	**95.4%**	**82.3%**	**9.5%**	**80.5%**	**60.2%**	**191.3%**	**-35.1%**	**207.2%**	**-31.2%**
PTV V107%	**Inf.**	**100.0%**	**57.1%**	**11.8%**	**128.3%**	**200.0%**	**142.5%***	**22.2%**	**100.0%**	**150.0%**	**400.0%**	0.0%	**800.0%**	0.0%
(PTV – 54 Gy OAR) D99%	n/a	n/a	0.0%	0.6%	0.2%	0.7%	0.4%	0.0%	n/a	n/a	0.7%	n/a	0.9%	n/a
(PTV – 54 Gy OAR) D95%	n/a	n/a	-0.1%	0.3%	-0.6%	0.6%	0.4%	-0.2%	n/a	n/a	0.3%	n/a	0.5%	n/a
Brainstem D5%	-0.9%	0.3%	0.7%	0.0%	0.8%	0.7%	0.1%	0.0%	0.3%	0.5%	1.0%	0.6%	0.6%	-0.2%
Brainstem D1cm^3	-0.8%	0.2%	0.6%	0.0%	0.8%	0.7%	0.1%	0.0%	0.3%	0.6%	1.1%	0.5%	0.7%	-0.1%

Brainstem mean dose	**-3.0%**	0.4%	0.9%	0.3%	1.5%	0.9%	0.2%	-0.2%	0.2%	0.4%	1.0%	0.2%	0.6%	-0.4%
Optic chiasm D1%	**-4.6%**	0.6%	1.1%	0.7%	0.8%	0.5%	0.0%	0.4%	-0.2%	1.0%	1.4%	0.3%	0.7%	-1.6%
R optic nerve D1%	1.2%	-0.3%	1.3%	0.3%	0.4%	0.0%	1.0%	-0.8%	0.7%	**3.0%**	0.1%	0.2%	0.2%	0.7%
L optic nerve D1%	0.9%	-0.5%	0.3%	-0.1%	-1.4%	0.3%	1.2%	-1.5%	-1.6%	-0.6%	1.6%	1.0%	1.6%	-0.3%
L globe D1%	**-6.7%**	-0.7%	0.8%	-0.3%	-0.7%	0.0%	-0.4%	0.5%	-1.7%	0.7%	0.0%	0.9%	0.3%	1.1%
R globe D1%	-0.2%	1.5%	0.3%	0.4%	-0.7%	0.6%	1.0%	-0.2%	-0.6%	0.6%	0.0%	-0.7%	-0.3%	0.0%
(Brain – PTV) D10%	n/a	n/a	0.8%	-0.3%	0.6%	0.6%	n/a	-0.1%	n/a	n/a	0.6%	-0.2%	0.6%	-0.3%
(Brain – PTV) mean dose	n/a	n/a	0.6%	-0.3%	0.5%	0.6%	n/a	-0.5%	n/a	n/a	0.5%	-0.2%	0.5%	-0.4%
L lens D1%	**-3.5%**	-0.8%	-1.5%	**-5.8%**	**-2.4%**	-1.7%	-1.2%	**-6.1%**	-0.3%	-1.6%	**-4.6%**	**3.8%**	-0.9%	0.0%
R lens D1%	**31.6%**	1.9%	**-3.4%**	**-7.7%**	1.3%	**4.0%**	1.3%	-1.7%	**-3.1%**	**3.2%**	-1.6%	**8.0%**	-1.4%	**9.7%**

L cochlea D50%	**-4.9%**	**-2.1%**	-0.3%	1.2%	**3.2%**	0.0%	1.1%	0.0%	**-2.5%**	-1.6%	0.6%	**2.3%**	0.2%	0.0%
R cochlea D50%	**2.6%**	0.0%	-0.7%	0.4%	**2.1%**	1.0%	**2.4%**	-0.4%	-1.0%	-0.2%	-0.2%	0.0%	-0.9%	-1.4%
L lacrimal D1%	**-8.6%**	1.5%	0.6%	n/a	-0.4%	**-3.3%**	-0.2%	**-4.0%**	-0.5%	-0.6%	**-2.1%**	-0.4%	-0.5%	-1.9%
R lacrimal D1%	n/a	**2.8%**	-0.7%	n/a	-1.0%	**-2.0%**	**2.0%**	**-2.4%**	-1.3%	-1.9%	1.9%	**-2.0%**	1.4%	-1.4%
Pituitary maximum dose	-1.9%	**2.1%**	-0.1%	-0.1%	0.6%	1.1%	0.9%	-1.3%	-0.4%	0.0%	1.3%	1.9%	0.8%	1.7%

Table 3-4 The percentage difference of the dose-volume criteria specified in Table 3-2 between radiotherapy plans created using the CT (reference dataset) and the sCT (evaluated dataset). Deviations greater than 2% are highlighted in bold. Any of these deviations which were additionally more than 1 Gy or 1% different in absolute value from the CT value, have been highlighted in yellow. Any cases where the criteria failed in the sCT, but passed in the CT are highlighted in green (if already highlighted the criteria are marked with an *). For cases which passed in the sCT, when previously they had failed in the CT, the criteria are highlighted in blue. Any criteria labelled as "n/a" were not contoured in the original plan. The labelling "Inf" refers to infinity and is due to a division by 0 (i.e. the criteria was equal to 0 in the original CT plan). Patients labelled with "MA" represent sCT created from the multi-atlas method, and those labelled with "NN" are generated from the neural network method. The number labelling represents the patient number. Only patients who had radiotherapy plans created in the Monaco treatment planning system were included.

Any dose-volume criteria with a deviation greater than 2% were investigated further. The majority of these deviations related to low original values, hence a small absolute change in value when planning on the sCT related to a large percentage deviation. However any investigated deviations where this was not the case (identified by a difference in absolute value between the sCT and CT of greater than 1% or 1 Gy) were recorded. These can be seen in Table 3-4 as highlighted in yellow.

The majority of dose-volume criteria which passed in the original CT plan, also passed in the sCT plan. Similarly, the majority of dose-volume criteria which failed in the original CT plan, also failed in the sCT plan. However, there were some cases where a small difference caused a criterion to fail in the sCT plan, when it had passed in the CT plan and also, infrequently, the other way around. The former events are highlighted in green and the latter in blue in Table 3-4. The only occasions where an identified large deviation (i.e. highlighted in yellow in Table 3-4) caused a change in whether a criterion passed or failed the dose-volume limit was for the V105% for the multi-atlas technique for patients 1 and 7 and the V107% for the multi-atlas technique for patient 9. These are marked with an asterisk (*) in Table 3-4.

Gamma analysis was performed using the body contour, the PTV and the 54 Gy PRVs as masks. The mean pass rates for 1%/ 1mm and 2%/ 2mm gamma criteria, for both local and global gamma analysis can be seen in Table 3-5.

	sCT generation method	1%/ 1 mm global	1%/ 1 mm local	2%/ 2 mm global	2%/ 2mm local
Average pass rate in the body contour/ %	MA	96.1 ± 0.5	83.4 ± 1.0	99.3 ± 0.2	95.4 ± 0.6
	NN	96.1 ± 1.1	84.3 ± 1.1	99.2 ± 0.2	95.4 ± 0.3
Average pass rate in the PTV/ %	MA	86.2 ± 2.8	84.9 ± 2.5	95.5 ± 1.6	95.5 ± 1.5
	NN	90.0 ± 1.9	90.1 ± 1.8	98.0 ± 0.7	98.1 ± 0.7
Average pass rate in the 54 Gy PRV/ %	MA	88.8 ± 5.2	77.6 ± 5.3	95.4 ± 3.8	91.9 ± 4.5
	NN	92.9 ± 5.1	69.0 ± 0.5	99.8 ± 0.1	93.4 ± 1.9

Table 3-5 The mean number of voxels passing the gamma analysis between dose deviations from plans on the sCT and those planned using the CT. The errors represent the standard error of the mean. Results of 1 %/ 1 mm and 2 %/ 2 mm gamma criteria, for local and global dose assessment, are shown for each sCT generation method (MA = multi-atlas, NN = neural network). Results within the patient body contour, the PTV and the 54 Gy PRVs (PRVs of the brainstem, optic chiasm and optic nerves) are given.

3.3.4 Discussion

The aim of this study was to investigate differences in the HU intensity of sCT images, which had been created from imaging patients using the new MRI protocol, from the original CT images. It was also important to assess whether these differences caused deviations in the dosimetry of radiotherapy plans.

The MAE results found for both models are comparable to those in the literature, where values in the range of 35 to 200 HU have been reported (3, 62). MAE values are typically higher for the brain compared to the more common prostate models, due to the differing amounts of soft tissue, bone and air present in the anatomy (91). Therefore the mean MAE values of 129.7 ± 5.3 HU and 114.4 ± 4.6 HU for the multi-atlas and neural network generated sCTs

respectively are not in themselves a concern. The calculated MAEs for individual sCTs in this study are also in line with the literature as demonstrated for the multi-atlas results by the boxplot in Figure 3-6. It is important, however, to assess the effect of this MAE on the dosimetry of generated plans.

It was noted that all MEs across both models, appear to produce a systematic negative offset. This has not been previously reported in the literature, however it is not common to discuss the ME of individual plans, merely to report the average ME. Upon inspection of the subtraction images, rings of low value (~-1000 HU) intensity were observed around the edge of the patient contour, indicating the sCT was systematically smaller than the CT. One possible cause of this negative offset is a difference in the patient contour position between the CT and the sCT, as a result of the CT and the MRI, used to generate the sCT, being acquired in separate sessions. This would not cause a clinical problem in an MRI-only radiotherapy pathway, where the patient would be imaged in the treatment position. Additionally, Dowling et al., (85) reported that, for the prostate model, a 1 mm expansion to the sCT body contour was needed to account for missing signal from collagen in the MR images. Therefore it is probable that this is also needed for the brain model and would reduce the observed negative offset. Furthermore, it can be seen in Figure 3-4 that the outer surface of the patients in the sCTs do appear blurred and jagged in places, and therefore this negative offset may also be caused by real peripheral errors in the sCTs.

Negative subtraction values were also consistently seen at tissue interfaces within the head, such as surrounding the skull (see Figure 3-5). This suggests that neither model handles the soft tissue – bone boundaries consistently well across patients. The next step is to investigate if this causes a clinical problem when using the images to plan radiotherapy treatments.

Demol et al., (27) analysed ME and MAE variation over a HU range of -1000 HU to 1800 HU. Variation in HU error was found to follow a similar trend to that reported in this study. For example, Demol et al., (27) found ME to range between approximately 0 HU to 300 HU for negative CT values, and to then to generally

decrease to around -600 HU at a CT value of 1800 HU. The MAE varied between 0 HU and 600 HU across all CT values. Andreasen et al., (29) similarly reported MAE variation between 0 HU and 450 HU for negative CT values and 0 HU and approximately 300 HU for atlas and patch-based methods and between 0 HU and 800 HU for regression methods for positive CT values, up to a CT value of 1500 HU.

The results found in this study regarding the variation in ME and MAE with CT HU value were similar to these limited results in the literature. The ME and MAE ranged from 0 HU to approximately (±)500 HU for negative CT values, and between 0 HU and approximately (±)350 HU for positive HU values up to 1500 HU. However beyond this range, the ME and MAE increased greatly with values of up to ±1000 HU error found at 1800 HU and up to approximately ±2000 HU error found at 2500 HU. Upon investigation of areas of the subtraction images associated with CT values in this range, it appears these are large negative errors in the region of dental artefacts, which are caused by the high HU value of these areas on the CT, which were not replicated on the sCT. These errors will adversely affect the MAE results, however it is unlikely that these will have a clinical impact. Dental implants are typically located away from the treatment volume, however if they are close to any critical structures or to the volume, the CT value around the artefact will be overridden and the beam placement arranged to avoid traversing the artefacts where possible.

This HU error variation investigation has shown that ME and MAE are lowest between approximately -900 and 1400 HU. This HU range covers the approximate HU values of air cavities up to those of the skull (137), therefore suggesting that the sCTs are likely to be acceptable substitutes for radiotherapy planning. However this can only be definitively assessed through a dosimetric comparison.

Generally good dosimetric agreement was found between plans created on the sCTs and the CTs. This is reflected in the high mean gamma pass rates within the body contour for both techniques (>99% and >96% with a 2%/ 2 mm and

1%/ 1mm global pass criteria respectively) which are comparable with those reported for multi-atlas prostate sCT generation methods in the literature (85, 91). The dose-volume criteria were also largely within 2% of each other (see Table 3-4). Where this was not the case, it was normally due to small original dose values and the difference was either within 1 Gy or 1% in absolute terms. However, it was common for the volume of the target receiving a high dose (105% or 107% of the prescription dose) in the sCT plan to greatly exceed that recorded in the CT plan. In three cases this caused the result to be out of tolerance, when it had previously been within tolerance in the original CT plan. This would be unacceptable for clinical implementation, as the effect that this would have in a clinical setting is that out of tolerance results may be falsely reported for the volume of the target receiving high doses. This may cause plan alterations to be carried out, when this is not necessary. The observed difference was investigated further by performing gamma analysis within the PTV.

The difference between local and global gamma analysis is small within the PTV. This is to be expected as the local dose should be close to the prescribed dose. It was found that on average more than 95% of voxels within the target passed the 2%/ 2mm gamma analysis for both techniques, suggesting a good agreement. However the lower average pass rate between 84.9% – 90.1% for the 1%/ 1mm gamma analysis, suggests some minor dose variation. It is possible that this minor variation ties in with the variation in the volume of the PTV receiving high doses. It should be noted that this gamma analysis cannot be used to compare the two sCT generation techniques, due to the different sample size used in the analysis of each.

One reason for this variation may be due to differences in patient set-up during the CT and MRI scan, which may contribute to the HU deviations around the patient contour observed in Figure 3-5, despite the registration of the two scans. The need to co-register the CT and MRI scans in order for the MRI and sCT to be in the same frame of reference, introduces an uncertainty and is a limitation of this study. If patients were scanned in their immobilisation mask during the MRI scan, improvements may be seen in areas around tissue boundaries, such as the

skull. These currently observed HU deviations may be responsible for the deviations in the volume of the target receiving a high dose. Deviations in skull density may result in dosimetric issues for these patients, as radiotherapy beams will pass through these areas during treatment. It should be noted, however, that even if patients were scanned in the immobilisation mask during the MRI acquisition, it is likely that minor differences would still exist due to the CT and MRI being acquired during different scanning sessions.

All remaining changes in pass-fail rates were due to small changes in values close to tolerance, which is not unexpected. The only other large deviation observed was a 3% increase in the dose received by the right optic nerve for patient 12. In this case, the result was still in tolerance, however this deviation is larger than expected and again may be due to the issues discussed above.

The dose deviation for the 54 Gy PRVs was also analysed using gamma analysis. The pass rate was higher for global than local gamma analysis, which is to be expected as the local dose is likely to be lower than the prescribed dose. Again, generally good agreement was found, with an average of >95% of voxels passing the 2%/ 2mm global analysis and >91% the 2%/ 2mm local analysis for both sCT generation techniques. However, the low mean pass rate of <80% for the 1%/ 1 mm local analysis, again suggests that some dosimetric variation is present in these areas. The local pass rates for the PRVs are lower than the respective local pass rates for the PTV (see Table 3-5). This is most likely due however to the lower doses in the PRVs compared to the PTV.

Therefore although in general good dosimetric agreement was found between sCT and CT plans, the large deviation of a small number of metrics i.e. the volume of the brain receiving 105% and 107% of the prescribed dose, mean that these sCT generation methodologies cannot be currently implemented clinically. However, the results are promising and so it is likely that only minor adjustments would be needed to produce clinically acceptable sCTs. It is probable that if a change in the workflow is introduced and brain patients are scanned in their immobilisation masks during MRI scanning, that the sCT deviations around the

patient contour and at tissue interfaces will improve, and a clinically acceptable solution may be achieved.

As was stated previously, it was not the purpose of this study to compare the results of the multi-atlas patch-based model and the neural network model. However, similar issues could be seen in the results from both models, with HU errors present in the same anatomical locations, and with both exhibiting deviations in the PTV volumes which received a high dose. The common factor between the models is the training images and therefore it is likely that both models would be improved by higher quality training data i.e. by acquiring more MRI-CT atlases or by acquiring MRI scans of patients within their immobilisation masks. The inability to provide a further comparison of the two models is a limitation of this study, and is largely due to the low number of sCTs generated with the neural network model. An area of future work would be to test more patients using this model, once more patients have been consented to the study.

One limitation of atlas-based methods is the potential difficulty in producing high quality sCTs for patients with atypical anatomy. For brain patients, the main area of concern is for resected areas of the skull which may have a clinically significant dosimetric impact if they are not well represented in the sCT. In the CT slice in Figure 3-4 some areas of bone are missing in the skull. These areas are not well modelled in either of the sCT slices, with a slight improvement seen in the neural network generated sCT compared to the multi-atlas generated sCT. This inability to reproduce areas of missing skull accurately was also observed for other sCTs in the study and is caused by the resection not being evident on the MRI scan due to the bone appearing as a signal void. This does not appear to have had a significant dosimetric impact for any of these patients, however this does not mean that this will be the case for all patients with skull resections, particular as skull resections are generally around the same area as the PTV. Therefore this would also need to be an area of focus for future work. For example, accurate representation of these areas may be improved by using a patient cohort with more patients with large skull resections present.

The introduction of the new imaging acquisition protocol has improved the sCT results compared to those generated when using MRI scans which had not been optimised for MRI-only radiotherapy in the model. There are no areas of the brain which have been classified as skull tissue as observed in Figure 3-1 and, although there is still some blurring around the air cavities there appears to be a reduction in blurring throughout the sCT (see Figure 3-4 compared to Figure 3-1). This visual improvement can also been observed in the quantitative analysis performed, as the weighted average MAE has reduced from 173 HU to 130 HU, for the same multi-atlas technique. This MAE is comparable with those reported for brain sCTs in the literature for voxel-based techniques involving UTE sequences (94, 109, 110, 119). The generation of sCTs in this study however, did not require the use of any MRI sequences which were not routinely used in the radiotherapy clinic. The use of a larger training database in the new study, may also have improved the quality of the resulting sCTs. It is expected that the sCT quality will improve further if changes to the scanning protocol are implemented i.e. if patients are scanned in the radiotherapy treatment position within immobilisation masks.

3.3.5 Conclusions

The introduction of a new MRI scanning protocol, where patients were scanned on a head rest using a consistently large FOV including all relevant anatomy, improved the generated sCTs. These sCTs were found to be of a good quality for the majority of HU and dosimetric metrics tested. However the reported MEs were consistently low and large deviations in the volume of the PTV receiving a high dose were concerning when considering the technique for clinical implementation, particularly as this was found to cause some criteria to fail the dose assessment, when the original plan had passed. It is recommended that models including patients who have been positioned in their immobilisation masks during their MRI scans are tested. This is expected to improve the atlas registration and hence the quality of the sCTs.

3.4 Concluding remarks

The clinical acceptability of sCTs generated from two different models (a multi-atlas and a neural network model) were tested. This was done firstly using MR images acquired in a diagnostic position and then using MR images acquired with a new protocol, which was however still in-line with the existing patient workflow in the department. The new protocol was found to produce superior quality sCTs compared to those produced using non-optimised MRI sequences and protocols. These sCTs generally performed well across the quality metrics used for testing. However deviations observed in some clinical metrics mean that further improvements would be required if the technique was to be implemented clinically. This may be achieved through MRI scanning of patients within their immobilisation masks, increasing the number of patients in the training database and through further optimisation of sCT generation algorithms. The additional inclusion of patients with large skull resections in the training dataset should also be performed.

Chapter 4 Automatic contouring of brain structures in an MRI-only radiotherapy workflow

4.1 Introduction

4.1.1 Delineation of brain structures in radiotherapy treatment planning

In radiotherapy, delineation of organs at risk (OARs) is vital for treatment planning in order to assess the dose received by critical organs and to ensure that this does not exceed set tolerances. By lowering the dose that these critical organs receive, the probability of normal tissue toxicity is reduced.

CT is conventionally used in radiotherapy treatment planning for dose computation and is also the most commonly used imaging modality for delineating the target and OAR volumes (1, 17). However, as previously discussed in Chapter 1, soft tissue contrast is poor with CT imaging. Therefore it is preferable to perform target and OAR delineation by registering an MRI scan, which has a superior soft tissue contrast, with the CT image set to help guide these delineations. This is particularly pertinent in brain imaging where MRI is an ideal modality for contouring structures within the brain, such as the optic chiasm and the cochleae (1). As discussed in Chapter 1, MRI-only radiotherapy planning has further advantages including removing the co-registration error between the MRI and CT scans, and simplifying the radiotherapy pathway.

Delineating the target and OAR volumes has been reported to be the most time-consuming part of the radiotherapy planning process for clinicians, especially with modern radiotherapy planning where multiple image modalities are often consulted during the delineation session (61). Therefore automation of this process would enable a more efficient clinical workflow. Additionally, it is known that the quality of the delineated contours depends on the skill of the observer (138), and that even amongst experts, an inter-observer variation in the delineations will exist (61). Variation in OAR contouring can result in differences in the reported doses to these volumes, and potentially differences to the planned

treatments. Automatic contouring would reduce the inter-observer variability, resulting in more objective and consistent OAR contouring.

Commercially available automatic contouring software does exist, however it is uncommon for these to be used in routine clinical practice (138, 139). Where these have been incorporated into the clinic, they are often used to produce initial structures which can then be reviewed and adjusted by clinicians (139). Bauer et al., (140) argued that whilst the true advantage of automatic contouring lies in the resulting objectiveness of the delineations, the segmentation must also be sufficiently accurate when considered by the individual clinician. If this cannot always be achieved then the automatic segmentation should be seen as an interactive process where the contour can be modified after the initial generation (140). Therefore the advantage of automatic contouring is that it can improve the efficiency of the workflow and mean that contour delineation does not need to be completely carried out by hand, but rather only minor adjustments need to be performed.

The majority of commercially available automated segmentation techniques are based on CT imaging (139). Therefore a limiting factor to the quality of the generated contour is the need for improved soft tissue discrimination. This can be achieved through the use of MRI information in the segmentation process and, as such, it is desirable for automatic segmentation of OARs on MR images, to be developed (139). It has been shown that automatic segmentation using an atlas-based method is feasible for some structures in head and neck radiotherapy using MRI images (60). In the study by Wardman et al., (60) it was shown using quantitative metrics only that automatic contouring of the head and neck structures was equally or more accurate using the MRI compared to the CT. Therefore a study of automatic contouring of brain OARs, generated using MRI multi-atlases for MRI-only radiotherapy, appears warranted.

4.1.2 Automatic contouring within an MRI-only workflow

The multi-atlas patch-based synthetic CT (sCT) generation method tested in Chapter 3 can also generate automatic structure delineations based on the input

MR image. To achieve this, structures must be delineated on the training images, thus forming a binary structure multi-atlas associated with the MRI atlas. The method for the generation of the automatic contours (auto-contours) has been described in Chapter 2. In brief, the deformations used to register each of the MRI atlases to the incoming MR image are additionally applied to each of the structure atlases. The results are then weighted as before, in order to compute the final automatic structure delineations.

All automatically contoured structures reported in the original prostate multi-atlas model by the CSIRO group (the prostate, bladder, rectum and bones), achieved a DSC > 0.8 compared to expert contours and were close to the inter-observer level, which was concluded to be a high level of accuracy (85). The contours were however only analysed numerically, and no clinician assessment was performed.

4.1.3 Aim of the investigation

The aim of this study was to determine if the automatic OAR structure delineations generated from the multi-atlas patch-based model were suitable for clinical use. This was assessed through a comparison of the auto-contours with the inter-observer variation observed in manually drawn contours. The auto-contours were also individually evaluated by an experienced neuro-oncologist in order to assess whether they were clinically acceptable, unacceptable or whether minor manual adjustment to them would be needed prior to clinical use.

4.2 Method

4.2.1 Brain structure contouring

The OARs delineated in this investigation were chosen based on the clinical meningioma protocol at Leeds Cancer Centre. The meningioma protocol was used over other brain protocols as this involves contouring the greatest number of brain structures. The structures delineated were the brainstem, spinal cord, optic nerves (left and right), optic chiasm, pituitary gland, lacrimal glands (left and right), cochlea (left and right) and globes (left and right).

In order to generate an atlas training database it was necessary to delineate the structures on the images in the MRI atlas (further information regarding the acquisition of these images can be found in Chapter 3). A consultant neuro-oncologist (Clinician I) contoured the structures on each of the 16 patients in the MRI atlas. In order to determine the inter-observer variation when multiple experts manually contour the structures, two other expert clinicians (Clinicians II and III) also contoured a subset of 7 patients (patients 2 – 8). A subset of patients was used due to limited clinician availability, however these delineations meant that approximately half of the patient auto-contours could be assessed with respect to inter-observer variability. Patient 1 was also contoured by all three clinicians, however this patient was used by the clinicians as a training dataset and therefore was not included in further analysis.

In addition, the same structures were also contoured on the CT scans of each of the 16 patients in the atlas. Again, Clinician I contoured all the scans, and Clinicians II and III contoured the subset of patient scans (patients 2 – 8). Contouring of the structures on the CT scan is not necessary for the production of an auto-contouring model, however these delineations were carried out so as to provide an additional means of testing the acceptability of the auto-contours.

Prior to delineation, the contouring methodology for each of the structures was discussed between the clinicians in order to improve the consistency of the delineations. A consensus regarding the contouring methodology was decided upon for each structure. For the optic chiasm it was agreed that the optic radiation should be included in the delineation. The chiasm and optic nerve delineations should be connected, but be kept as separate structures. The optic nerves should not extend posteriorly beyond the optic canal. The transition between the spinal cord and the brainstem was defined as being located at the odontoid peg, as defined in Eekers et al., (141), and where the cerebellar tonsils could no longer be observed. The brainstem was classed as ending in the superior direction where cerebrospinal fluid could no longer be seen in front of the brainstem. The cochlea were defined as being positioned where the eighth cranial nerve ends on

the MR image, in line with recommendations by Scoccianti et al., (142). Imaging window levels were discussed and an optimum window level for delineating structures on the MR images was agreed upon. This window preset was created in the Oncentra Masterplan treatment planning system (Elekta AB, Stockholm, Sweden) for use by the clinicians during delineation.

The MR images were loaded into the Oncentra Masterplan treatment planning system in the imaging frame of reference, as described in Chapter 3. The clinicians then delineated the OAR structures in line with the decisions discussed above. Next, it was ensured that all the delineations had been labelled consistently, that no erroneous contours were present and the clinician names were changed to a numerical label. It was also checked that the previously drawn patient contour (described in Section 3.3.2.1) encompassed all the structure delineations. The delineations outlined by the first clinician, Clinician I, who had contoured the structures on all patients, were then exported to Mirada RTx (Mirada Medical, Oxford, UK), along with the MRI and CT scans. The CT scans were registered to the corresponding MRI scans and hence also brought into the same frame of reference as the MRI contours. The contours delineated by Clinician I on the MRI scans were shared with the CSIRO e-learning centre, Brisbane, Australia, along with the MRI and CT image sets, in order to produce a structure multi-atlas associated with the MRI multi-atlas. Using a leave-one-out approach, auto-contours associated with each incoming MRI and generated sCT scan were then produced for each structure by CSIRO, and were returned for analysis along with the corresponding sCT.

4.2.2 Automatic contour assessment

4.2.2.1 Comparison to expert manual contours

All delineated structures were exported from the Oncentra Masterplan treatment planning system and imported into ImSimQA (Oncology Systems Limited, Shrewsbury, UK). The auto-contours for each patient were also imported into this system. For each patient, Clinician I's MRI contours were selected as the reference contours. These were chosen firstly as these were the contours which

had been used to train the model and secondly as these structures had been delineated on all patients. As a leave-one-out approach was used to generate the auto-contours, using the Clinician I contours as the reference contours should not bias the results. The auto-contours were loaded as the set to be evaluated.

For each structure the dice similarity coefficient (DSC) (143) between the auto-contour and the Clinician I manual contour was computed. The DSC is defined as the size of the union between the evaluated auto-contour and the reference contour, divided by the average size of the datasets, and is calculated using Equation 4-1, where *R* is the reference volume, *E* is the evaluated volume and $(R \cap E)$ is the overlapping volume between the two. DSC values of 1 indicate that the volumes overlap completely, and values of 0 indicate no volume overlap.

$$DSC = \frac{2 \times (R \cap E)}{R + E}$$

Equation 4-1

The acceptability of the auto-contours was also assessed as a function of the volume of the structures. This was carried out by determining the volume of the structures outlined by Clinician I using the ImSimQA software. The average structure volume and the average DSC between the auto-contours and the Clinician I delineations across all patients was then determined for each structure.

The conformity index (CI) also describes the overlap between two volumes and is determined using Equation 4-2 (144), where $(R \cap E)$ is the volume of overlap between the two contours as before, and R_N and E_N are their non-overlapping volumes respectively. The CI is equal to 1 if the volumes match perfectly and 0 if there is no overlap.

$$CI = \frac{(R \cap E)}{R_N + (R \cap E) + E_N}$$

Equation 4-2

In order to determine whether the acceptability of the auto-contours varied across the axial slices of the imaged structures, the CI of the auto-contours with respect to the Clinician I manual contour on each slice was found using the ImSimQA software. The slice analysis was performed for structures which were, on average, larger than 1 ml in volume based on the Clinician I manual delineations (determined using the ImSimQA software). For each structure and patient, the ratio of the mean CI of the two outermost superior and inferior slices (four slices in total) to the mean CI across the central four slices was calculated. The slice gap of the MRI scan was 2 mm. Four central slices were analysed in order to be consistent with the number of outermost slices used. Through this calculation, it could be assessed whether the quality of the outermost slices was poorer than the quality of the central slices. The mean, maximum and minimum of these ratios across all patients was calculated for each analysed structure.

4.2.2.2 Comparison with respect to inter-observer variability

In order to assess the viability of the auto-contours, the variation of the auto-contour from a contour manually delineated by an expert should be determined as has been described above. It is likely that some degree of deviation from this contour will exist. However an inter-observer variation will also be present, even between experienced clinicians. Therefore an ideal way of assessing the auto-contours for clinical acceptability, is to determine if the variation between the auto-contour and the expert “gold standard” contour is within the inter-observer variation.

For each structure the CI, the mean distance to conformity (MDC) and the centre of gravity distance (CGD) between the auto-contour and the Clinician I manual contour were computed using the ImSimQA software. As the CI was included in this analysis, the DSC was not analysed due to similar information being obtained from both metrics.

The CI metric has already been described in this chapter and is useful in assessing the amount of overlap between the contours. No information, however, is given as to the degree of difference between the contours where they do not

overlap or to the directionality of the difference i.e. whether the evaluated contour is larger or smaller than the reference contour.

The MDC provides more information regarding the distance between the surfaces of the two contours. The MDC is defined as the mean distance of outlying voxels in the evaluated structure contour to the nearest voxels in the reference contour or, in other words, the average distance needed to move outlying voxels of the auto-contour to match the reference contour (145). The greater the MDC value, the greater the average distance needed to move the outlying contour voxels in order for them to match the reference contour. The MDC provides information regarding the deviation of the surface shape of the evaluated structures from the reference contours, however it does not determine the deviation between the centres of the structures.

The CGD describes the distance between the centres of gravity of the two structures. In ImSimQA a 3-dimensional nodal grid is used, where each node represents a 1 mm^3 volume. Both contours are transferred onto this grid and the distance between the nodes at the centres of the two contours is determined. The greater the CGD value, the greater the distance between the two centres of gravity. This provides information regarding the deviation in the anatomical location of the two contours.

These metrics were also computed with Clinician II and III's contours on the MRI dataset set as the evaluated contours, with Clinician I's contours still set as the reference. The metrics were also computed with Clinician III's contours set as the evaluated contours and Clinician II's contours set as the reference. This was carried out for all patients where three clinicians had performed contouring (seven patients in total). A diagram indicating the comparisons which were carried out can be seen in Figure 4-1.

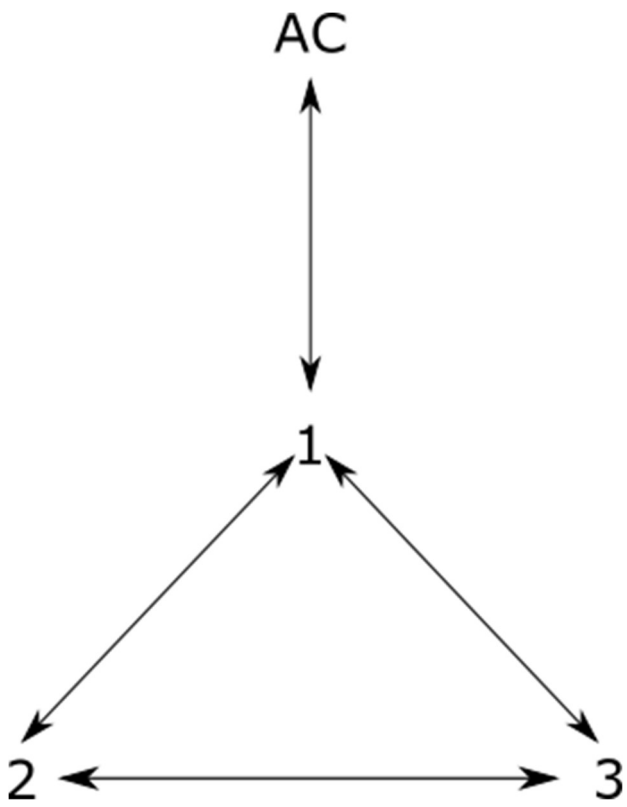

Figure 4-1 Schematic diagram showing the contour comparisons which were performed. AC represents the automatic contours generated from the multi-atlas model. 1, 2 and 3 are the contours manually delineated by Clinician I, II and III respectively. The conformity index, mean distance to conformity and centre of gravity distance were determined between these contours as indicated by the arrows.

For each similarity metric and structure, the mean of the inter-observer values was determined (i.e. the mean of the metric values between each of the clinician contours, as indicated by the triangle in Figure 4-1, across all patients). These were then compared to the metric values between the auto-contour and Clinician I contour.

These inter-observer comparisons using the CI metric were also carried out for the delineations performed on the CT dataset.

Additionally, for each structure, the percentage of patients where the CI between the auto-contour and the Clinician I delineation was greater than the lowest CI between inter-observer delineations on the MRI (the triangle in Figure 4-1) was

determined. This was to test whether the deviation of the auto-contour from an expert contour lay within inter-observer variability.

4.2.2.3 Clinical acceptability of automatic contours

Besides using quantitative metrics, the auto-contours were also individually assessed for clinical acceptability. A new case was created for each patient in Oncentra Masterplan containing the MR image set and the auto-contours. Clinician I then commented on the clinical acceptability of the auto-contours for each patient based on their clinical judgement. For each structure the auto-contour was either assessed to be clinically acceptable, unacceptable or requiring minor adjustments in order to be used clinically. The clinician also noted any additional comments for individual structures, for example if acceptability varied over different slices. The reason for this is that it was hypothesised that automatic contours may fail at their superior and inferior end slices, whilst other central slices may be acceptable.

4.3 Results

Three dimensional renderings of the auto-contours generated for a typical patient (patient 8) in two planes can be seen in Figure 4-2 along with the MR image.

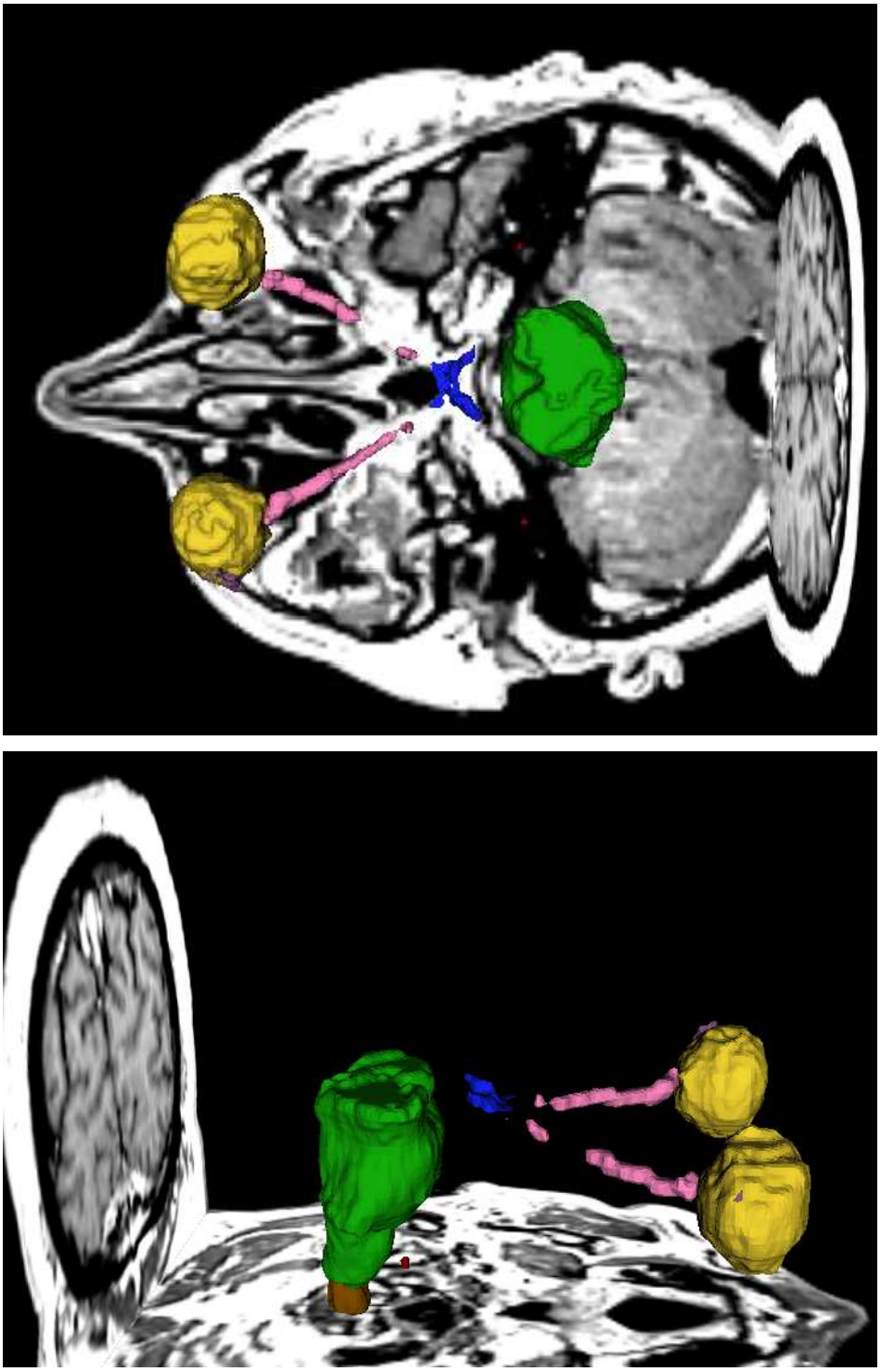

Figure 4-2 Three dimensional rendering of the automatic contours (brainstem in green, chiasm in blue, cochleae in red, globes in yellow, lacrimal glands in purple, optic nerves in pink, pituitary gland in orange, spinal cord in brown) generated in the 3D Slicer software (http://www.slicer.org) for patient 8. Slices from the patient MRI can also be seen. Two planes are shown. In the upper image the brainstem, optic nerves and chiasm, globes, lacrimal glands and cochleae are visible. In the lower image the spinal cord and pituitary gland can also be seen.

4.3.1 Automatic contour assessment with respect to expert manual delineations

The DSC between the auto-contours and the delineations drawn by Clinician I for each patient for the optical and other structures are shown in Figure 4-3 and Figure 4-4 respectively. Separate figures have been used to aid clarity, however both are to the same scale for ease of comparison. These figures show the variation in DSC across patients.

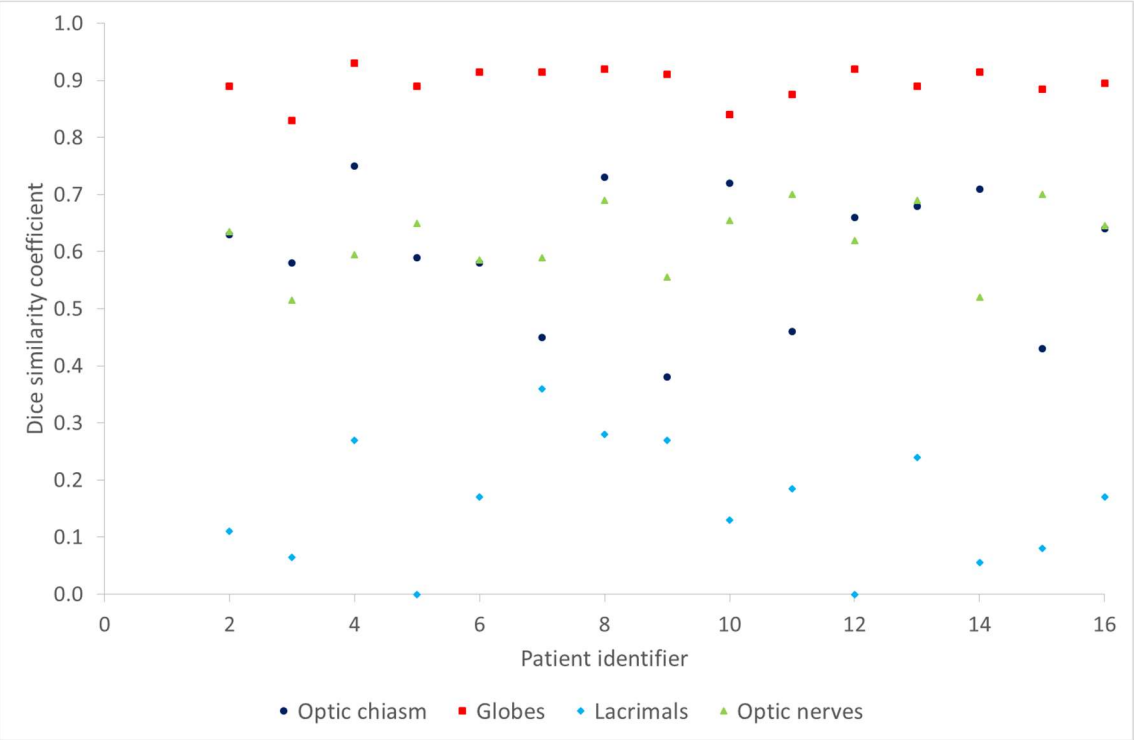

Figure 4-3 The Dice similarity coefficients calculated between the auto-contours and the delineations performed by Clinician I for the optic structures (chiasm, globes, lacrimal glands and optic nerves) for all patients. To aid visualisation, the mean of the left and right structures is shown for the globes and lacrimal glands. Patient one was excluded as this was a training case.

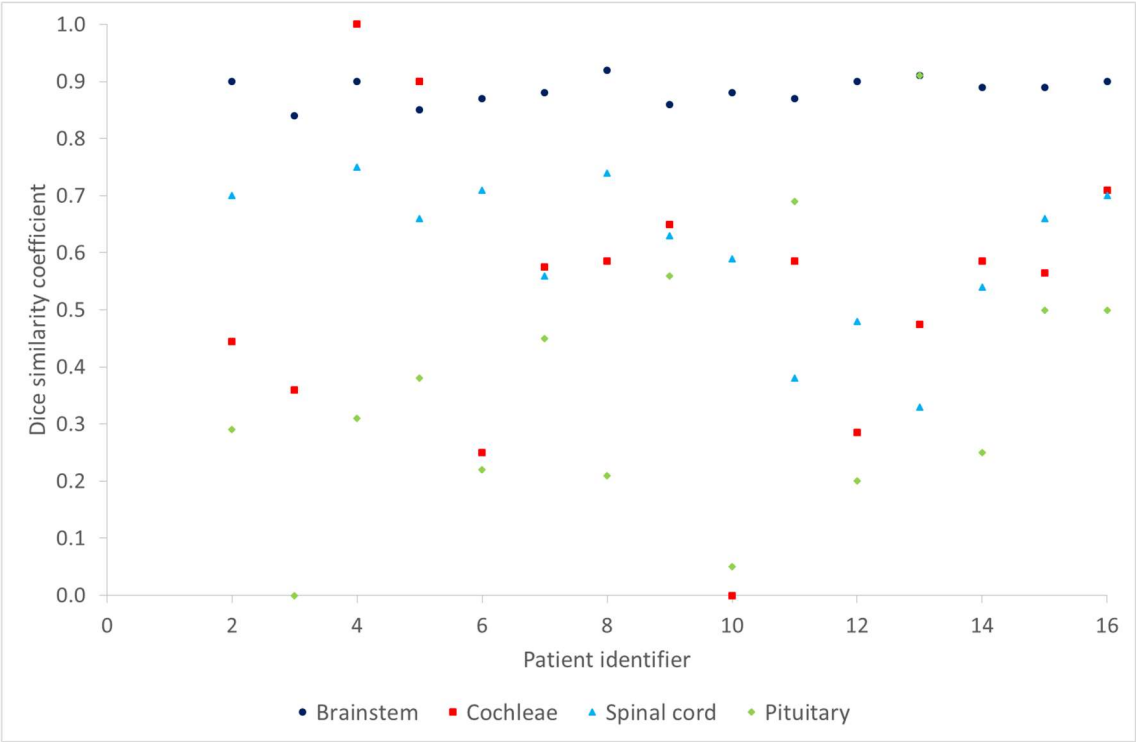

Figure 4-4 The Dice similarity coefficients calculated between the auto-contours and the delineations performed by Clinician I for the brainstem, spinal cord, cochleae and pituitary for all patients. To aid visualisation, the mean of the left and right cochleae is shown. Patient one was excluded as this was a training case. No spinal cord was present for patient 3 in the original MR image due to the inferior extent of the image, or in the auto-contours as expected.

The mean DSC values between the auto-contour and the delineations by Clinician I as a function of the mean volumes of the structures (delineated by Clinician I) can be seen in Figure 4-5.

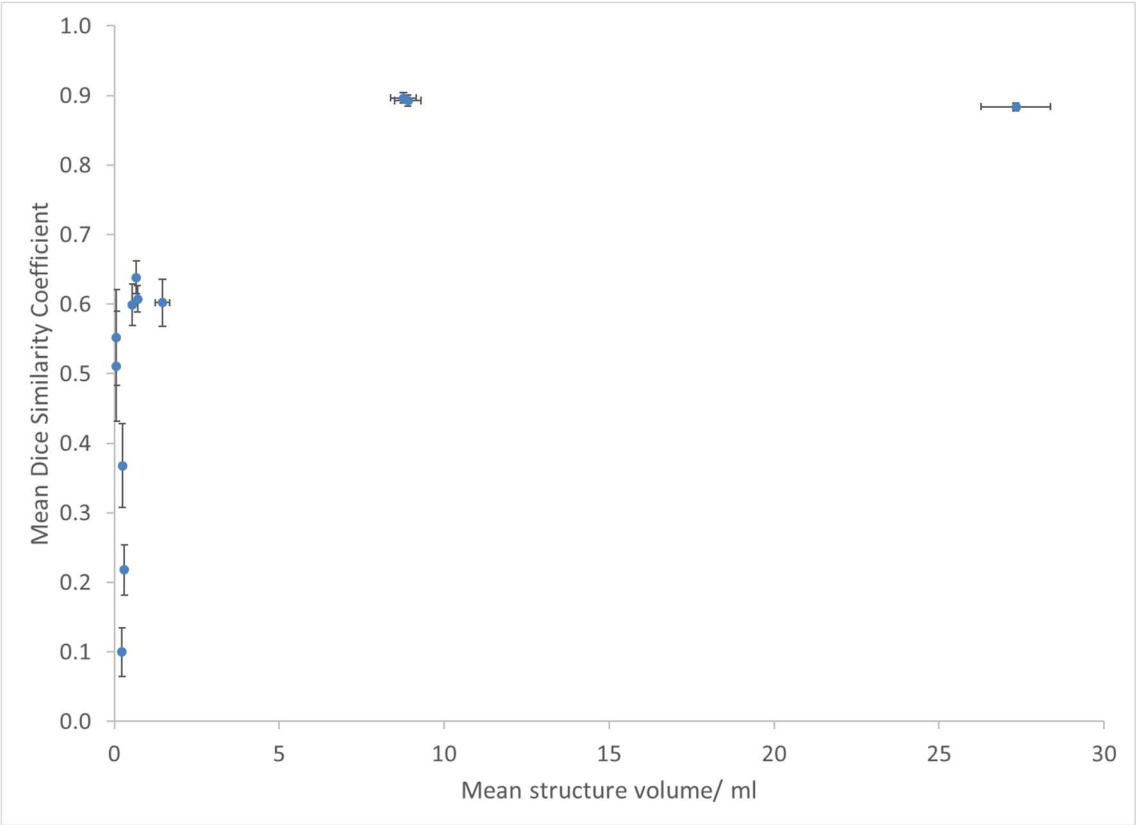

Figure 4-5 The mean Dice Similarity Coefficient (DSC) between auto-contours and Clinician I contours against the mean volume of structures contoured by Clinician I. The error bars represent the standard error of the mean.

The brainstem, spinal cord and left and right globes were found to have mean volumes of greater than 1 ml, considering all patient delineations contoured by Clinician I. For these structures, the ratios of the average CI over the most superior and inferior slices to the average CI over the central slices were determined. This metric determines the quality of the outermost slices of the structure relative to the quality of the centremost slices, so as to assess quality variation across the superior - inferior extent of the structure. The mean ratios across all patients for the four structures are shown in Figure 4-6. The range of ratios is also displayed. For most structures, patients 2 - 16 were included in this analysis. The only exception to this was for the spinal cord where only patients 2, 4, 5, 7, 9-11 were analysed. This is because for the remaining patients the small volumes would have led to the same slices being included in the central and outermost slice analysis.

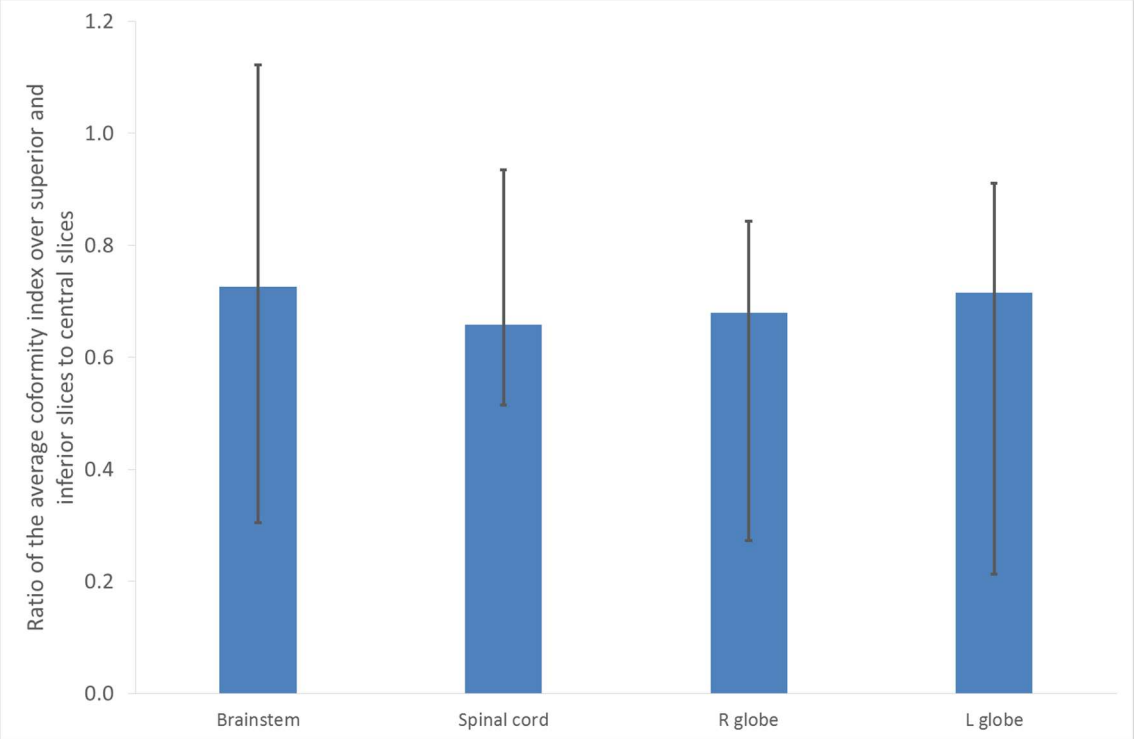

Figure 4-6 The bars represent the mean ratio over all patients of the mean conformity index over the most superior and inferior slices to the mean conformity index over the central slices. The error bars represent the range of ratios found over all patients. Only structures which were found to be, on average, greater than 1 ml in volume based on the Clinician I delineations have been analysed.

4.3.2 Automatic contour assessment with respect to inter-observer variability

The CI, MDC and CGD for the average inter-observer comparisons on the MRI (the comparisons illustrated by the triangle in Figure 4-1) and for the auto-contour and Clinician I MRI contour comparison were calculated for each patient structure. The average of these metric values for each structure over the 7 patients contoured by all three clinicians can be seen in Figure 4-7, Figure 4-8 and Figure 4-9 for the CI, MDC and CGD respectively. For the spinal cord, only 6 patients were included as this structure was not present on the MRI for one patient (patient 3) due to the inferior extent of the image. Due to time constraints, Clinician II was also unable to contour the spinal cord on patients 2, 4 and 5, so these have not been included in this analysis.

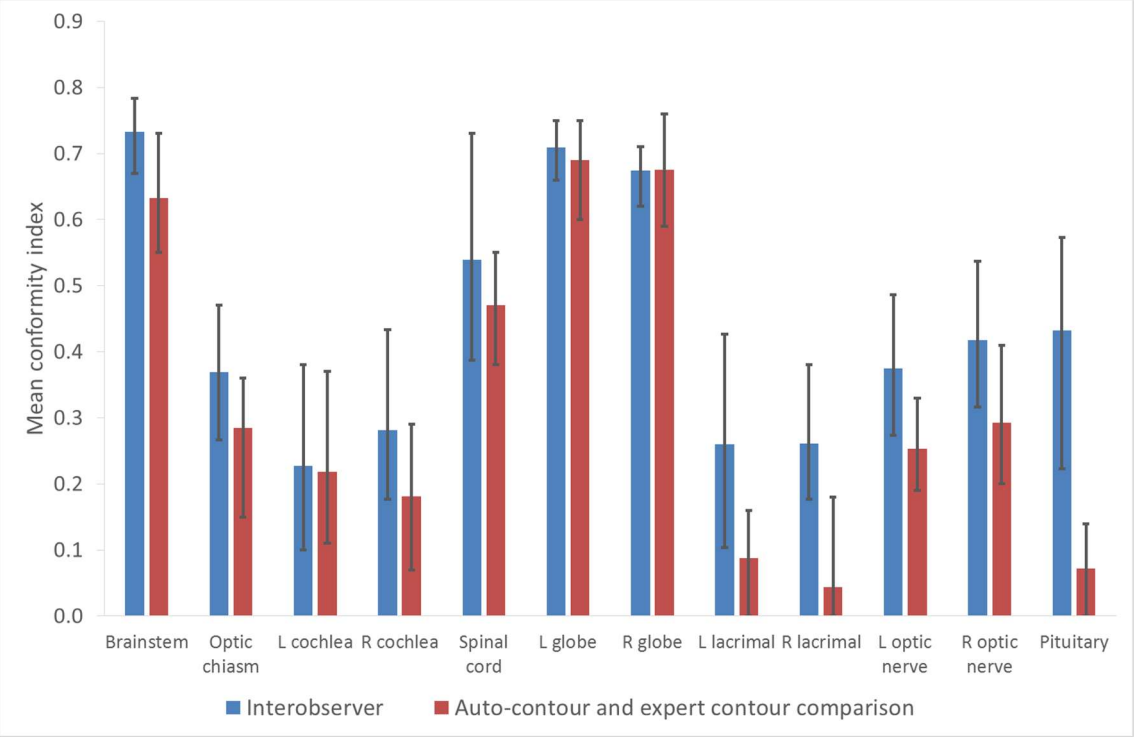

Figure 4-7 The mean conformity index (CI) for patients 2 – 8. For each structure, the mean inter-observer CI is shown alongside the mean CI between the auto-contour and Clinician I's manual contour. The errors bars represent the range of the mean inter-observer CIs over the seven patients and the range of the CIs between the auto-contour and Clinician I's manual contour over the seven patients respectively. The labels L and R represent left and right respectively.

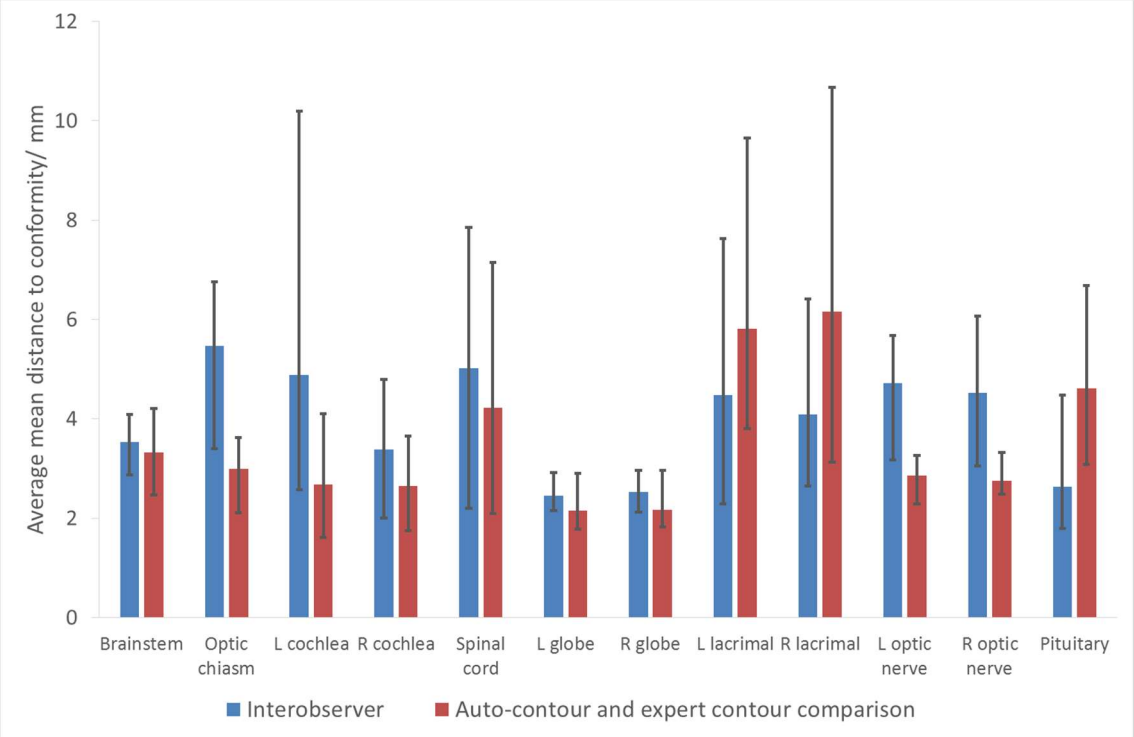

Figure 4-8 The mean of the mean distance to conformity (MDC) values for patients 2 – 8. For each structure, the mean inter-observer MDC is shown alongside the mean MDC between the auto-contour and Clinician I's manual contour. The errors bars represent the range of the mean inter-observer MDCs over the seven patients and the range of the MDCs between the auto-contour and Clinician I's manual contour over the seven patients respectively. The labels L and R represent left and right respectively.

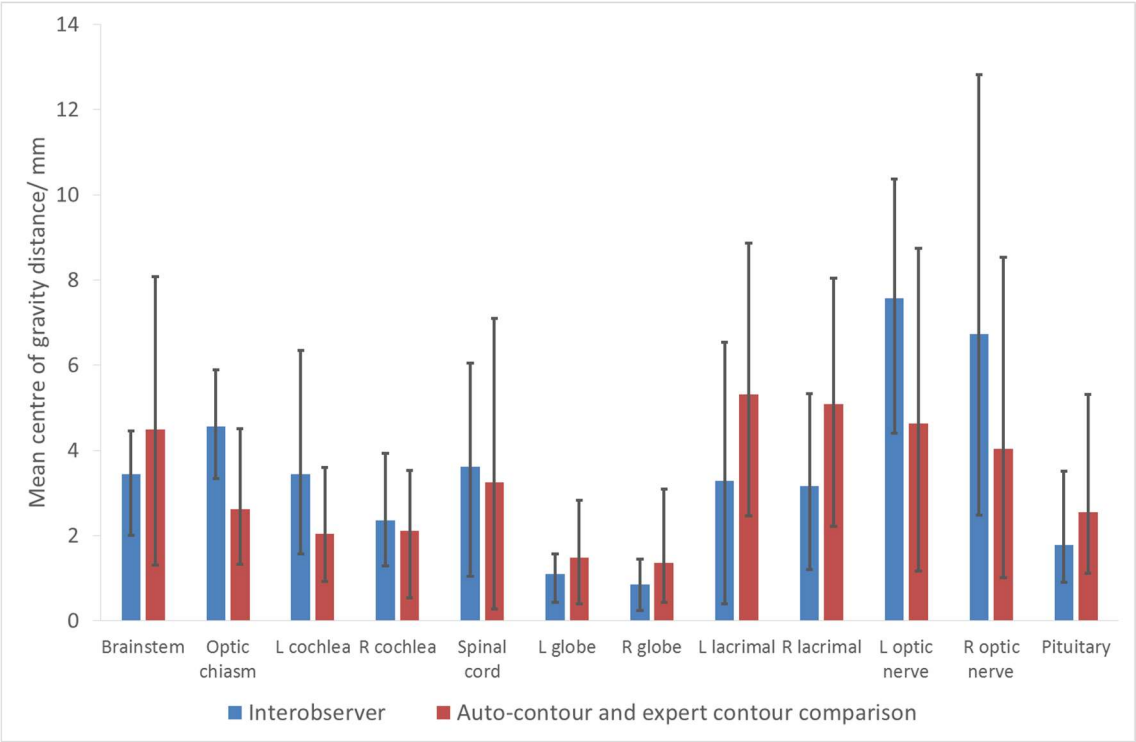

Figure 4-9 The mean centre of gravity (CGD) for patients 2 – 8. For each structure, the mean inter-observer CGD is shown alongside the mean CGD between the auto-contour and Clinician I's manual contour. The errors bars represent the range of the mean inter-observer CGDs over the seven patients and the range of the CGDs between the auto-contour and Clinician I's manual contour over the seven patients respectively. The labels L and R represent left and right respectively.

A comparison of the mean CIs between inter-observer delineations of the structures delineated on the MRI and CT scans can be seen in Figure 4-10. The range of CIs across all patients is also shown. Due to time constraints, Clinician II was unable to contour the spinal cord on patient 7 on the CT, so this has not been included in this analysis. It was ensured that for the spinal cord only inter-observer comparisons which were present on both the MRI and CT were included in the analysis.

The percentage of patients where the CI between the auto-contour and Clinician I contour was greater than the lowest inter-observer CI on the MRI for each structure is shown in Figure 4-11.

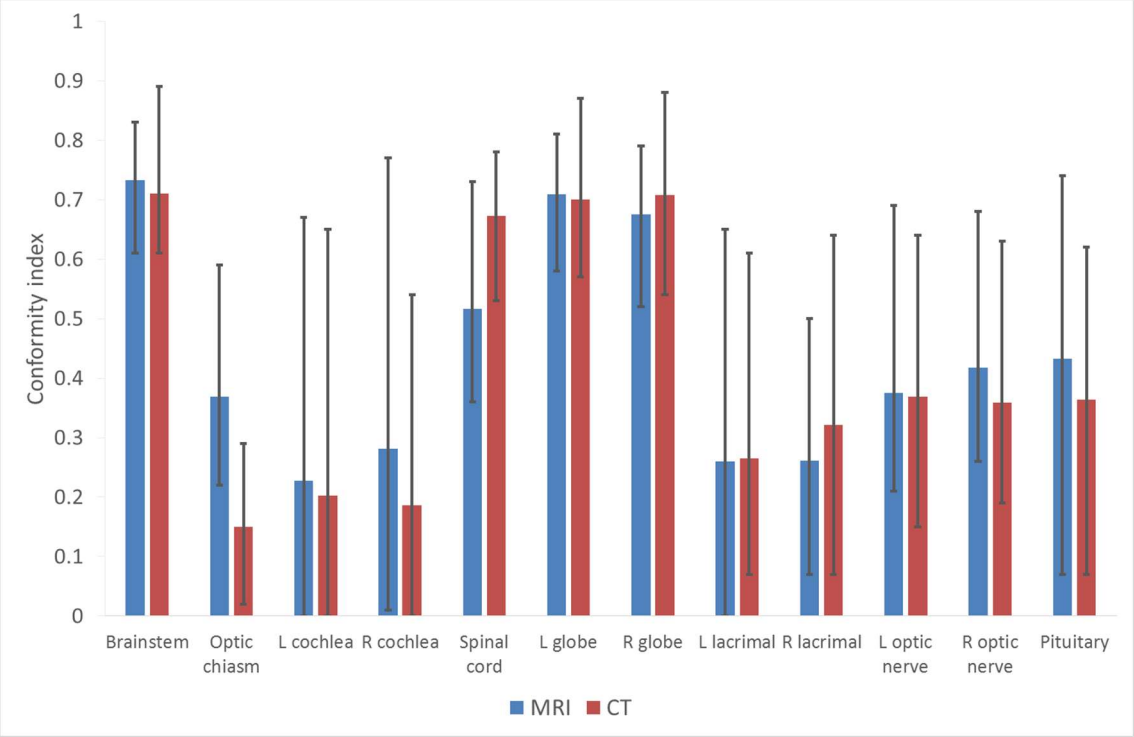

Figure 4-10 The mean conformity index for patients 2 – 8. For each structure, the mean conformity index between the inter-observer comparisons is shown for delineations on the MRI and CT. The error bars represent the range of CIs across all patients. The labels L and R represent left and right respectively.

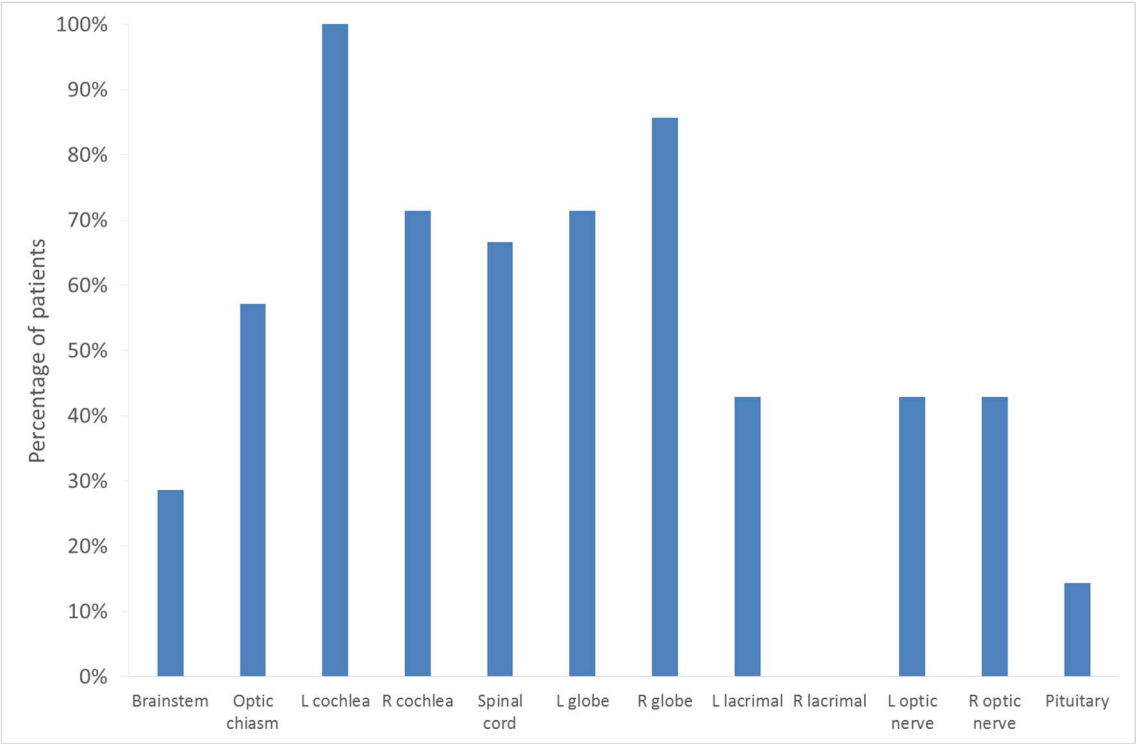

Figure 4-11 The percentage of patients where the conformity index between the automatic and manual contour was greater than the lowest conformity index between observers on the MRI. The results for all structures are shown. Due to limited delineations by Clinician II only patients 6-8 were analysed for the spinal cord.

4.3.3 Clinical acceptability of automatic contours

As well as performing quantitative analysis of the auto-contours compared to the inter-observer variability, the clinical acceptability of auto-contours was assessed qualitatively. The results of the clinician assessment can be seen in Figure 4-12. Each structure was either marked as clinically acceptable, in need of minor adjustment or not acceptable. In total 12 patients were included in this assessment with all structures being analysed for these patients. Patients 3, 7 and 12 were excluded from this analysis as the structure files could not be imported into Oncentra Masterplan. This will be further investigated and remedied in future work.

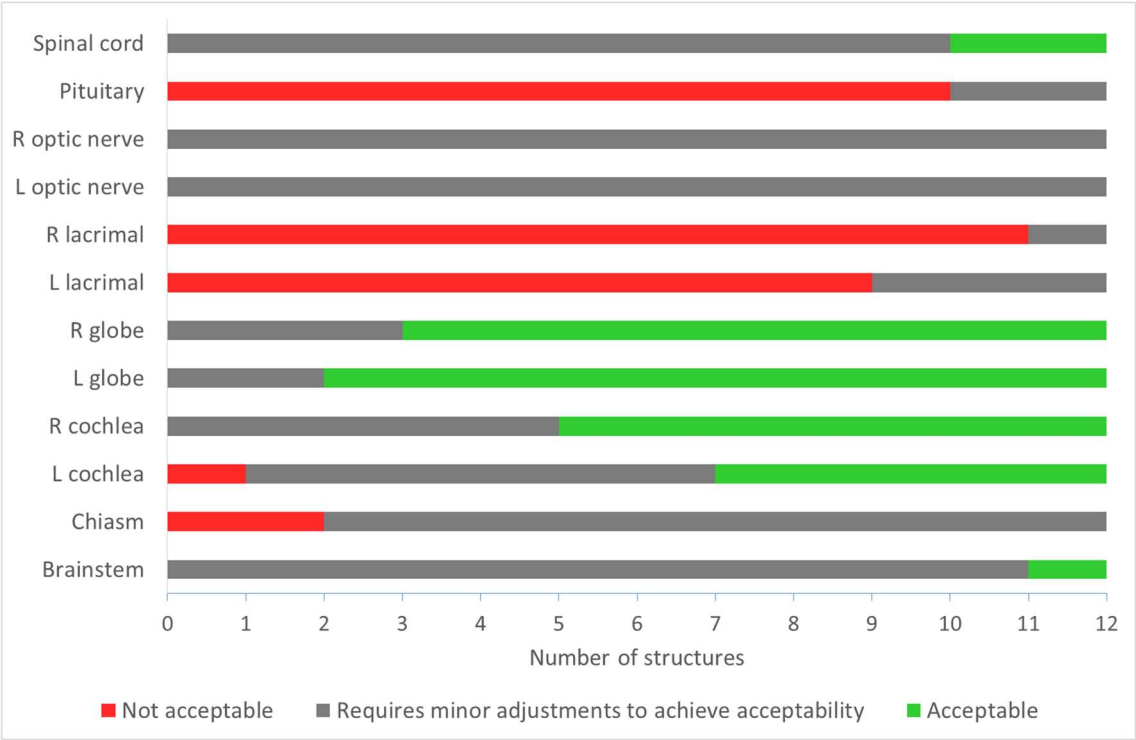

Figure 4-12 Automatic contour classification as either not acceptable (red), in need of minor manual adjustments (grey), or clinically acceptable (green). 12 patients were assessed for each structure. "L" and "R" represent left and right organs respectively.

The majority of spinal cord and brainstem auto-contours required minor manual adjustment. This however mainly concerned the most superior, and in some cases, inferior slices. It was noted however that for one spinal cord auto-contour the delineations were slightly too small across all slices. The auto-contours frequently appeared coarse at the surface, with rough edges, which would require smoothing. All optic nerve auto-contours required minor adjustments. The optic nerves appeared in the correct position and had the correct general shape, however there were often discontinuities in the contours and they required smoothing (see Figure 4-13 as an example).

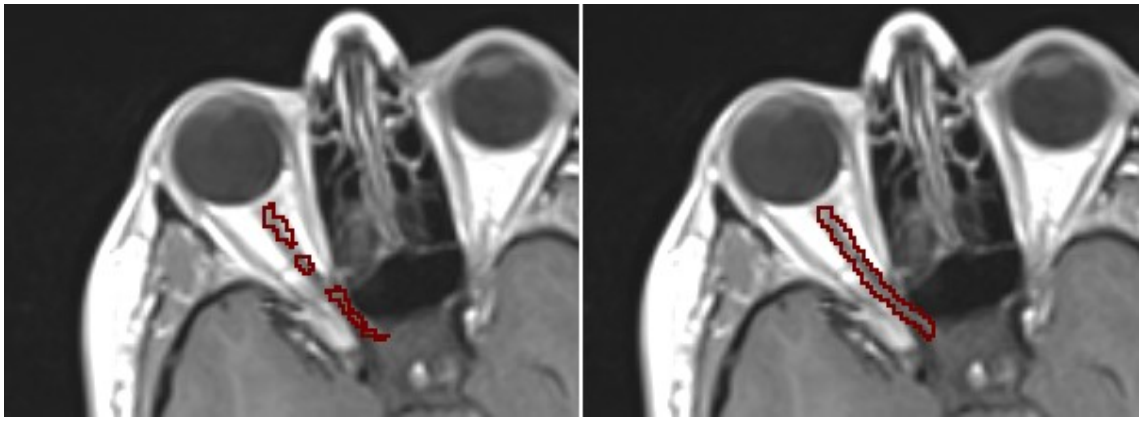

Figure 4-13 An axial slice through the MRI scan of patient number 6. The optic nerve contours are shown in burgundy for the automatically generated contour (left) and the manually drawn contour (right). These were classified as needing minor manual adjustments in order to be acceptable clinically due to the discontinuities in the contour and the need for smoothing of the contour.

The majority of globe auto-contours were classified as clinically acceptable although a slightly rough edge was noted. Five auto-contours, however, required minor adjustments. This was again due to the more uneven nature of the surface, which would require smoothing. Additionally some minor adjustments were needed for the superior and inferior slices. For one auto-contour it was found that the globe encroached on the maxillary sinus (see Figure 4-14).

Minor adjustments, which were required for 11 out of a total of 24 cochleae contours, were mostly related to the need for smoothing and the removal of small volume sections of the contour. For one contour, the inferior slice was misplaced. One auto-contour was found to be clinically unacceptable as it was missing a large section of the structure (see Figure 4-15).

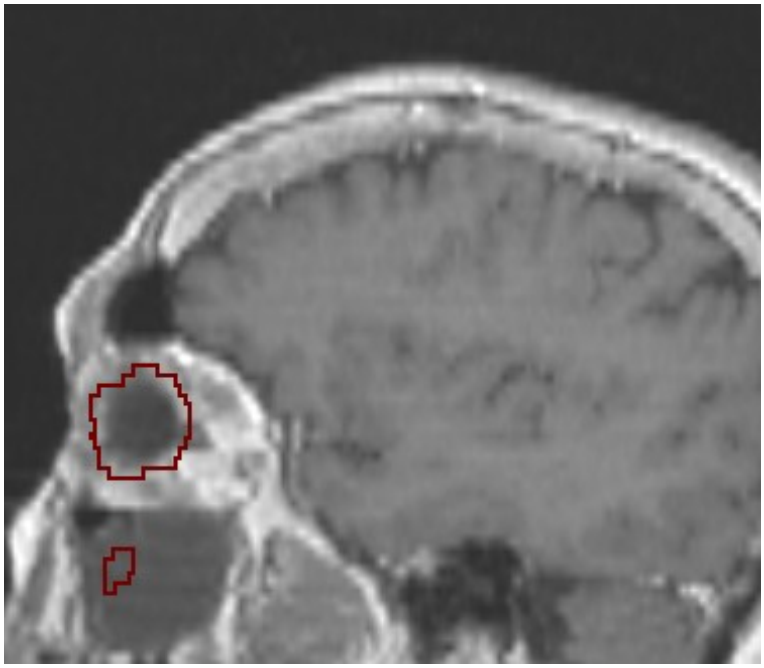

Figure 4-14 A sagittal slice through the MRI scan of patient number 10. The right globe auto-contour is shown in burgundy. The globe was accurately delineated, however the auto-contour was classified as needing minor adjustment due to the inclusion of an area of the maxillary sinus in the delineation. The uneven nature of the globe surface, which was also observed for other patients, can be seen.

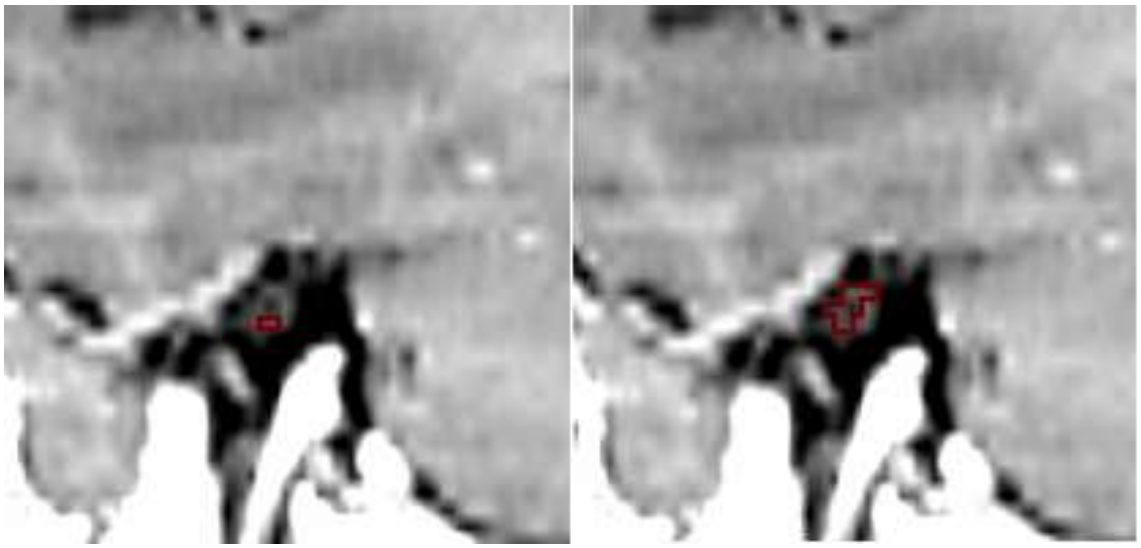

Figure 4-15 A sagittal slice through the MRI scan of patient number 6. The left cochlea contours are shown in burgundy for the automatically generated contour (left) and the manually drawn contour (right). The auto-contour was classified as clinically unacceptable as a large section of the structure was missing.

The chiasm auto-contours were found to require minor adjustments for 10 out of the 12 contours. This was related to the need for contours to be smoothed, as well as issues with the anterior sections of the contours (see Figure 4-16 as an example). Two contours were clinically unacceptable due to missing sections of the structure.

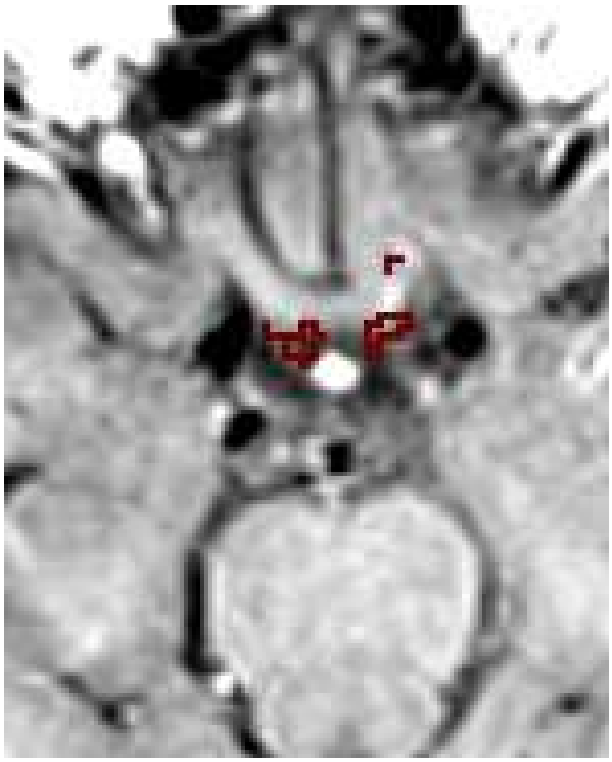

Figure 4-16 An axial slice through the MRI scan of patient number 11. The chiasm auto-contour is shown in burgundy. This auto-contour was classified as needing minor manual adjustments in order to be acceptable clinically due to the uneven and erroneous placement of the anterior section of the contour.

Generally the poorest quality auto-contours were with the lacrimal and pituitary glands where the majority of structures were clinically unacceptable. These small structures were found to consist of small volume “dots” when generated automatically, with frequent discontinuities in the contour, which were rough and in need of smoothing. They were also commonly in the wrong position and incomplete (see Figure 4-17 as an example).

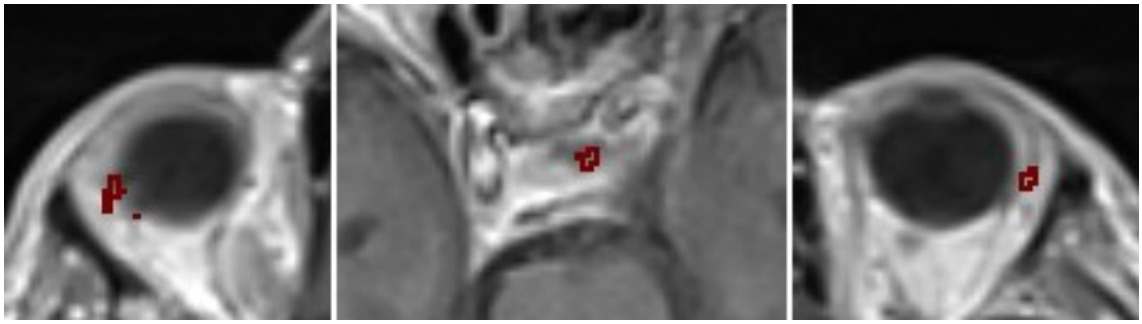

Figure 4-17 Axial slices through the MRI scan of patient number 9. The right lacrimal (left), pituitary (centre) and left lacrimal (right) auto-contours are shown in burgundy. These were classified as clinically unacceptable. The lacrimal gland auto-contours are uneven and discontinuous, with the contours being made up of small volume sections. The pituitary auto-contour is uneven and incomplete.

4.4 Discussion

The main aim of utilising automatically delineated OAR contours is to replace manually drawn expert contours, in order to improve time efficiency, as well as consistency between contours. In order to be able to do this, the contours must be as accurate as the expert contours i.e. the contours must be within the inter-observer variation of contours drawn by multiple expert delineators. As well as this, it must be shown that contours of consistently high quality can be produced for all patients. If these requirements are not met, time saving and improved inter-observer consistency may still be achieved by using automatic contours which only require minor manual adjustments.

The quality of brain structure auto-contours produced through this multi-atlas model are of variable quality. The lacrimal and pituitary gland auto-contours were consistently poor, with the mean, maximum and minimum of the CI, MDC and CGD metrics between the auto-contours and the Clinician I contours being of poorer quality than those of the inter-observer variations (see Figure 4-7, Figure 4-8 and Figure 4-9 respectively). The majority of these auto-contours were designated as being clinically unacceptable (an example can be seen in the rendering of the right lacrimal and pituitary in Figure 4-2 and in Figure 4-17).

It is noteworthy that the mean CI is also relatively low for the inter-observer agreement of the lacrimal glands both on the MRI and CT (less than 0.3 and 0.4 respectively), which indicates that the lacrimal glands are difficult to locate, even with expert manual delineation, particularly on the MRI. Therefore although it may be possible to improve the accuracy of these auto-contours through acquiring a larger patient cohort, it may be that these contours are not ideal for automatic MRI delineation due to the discrepancy and variation in the manual contour delineation with MR imaging. One reason that this may be the case is due to the lack of bone visualisation in T_1-weighted MRI imaging, which was used to guide the delineation in this study. The lacrimal glands are positioned partly enclosed by bone (146) and therefore this lack of information potentially leads to a larger inter and intra-observer variation in the delineation compared to CT, despite the increased soft tissue contrast provided by the MRI. Therefore an improvement may be found in both manual and automatic lacrimal MRI contouring if an ultra-short echo time MRI sequence was also acquired during the MRI imaging session, which could be used to visualise bone and hence aid contouring.

Conversely T_1-weighted MRI is the modality of choice for pituitary imaging (147). This is supported by Figure 4-10 which shows a higher mean and maximum CI between observers on the MRI datasets compared to the CT. There is a natural variation however in the size and shape of the pituitary gland in the population, for example it is typically larger in females and, in particular, females over the age of 50 (147). A contributing factor to the poor quality of the auto-contours may be a small sample size for each of these groups. A larger training cohort covering all groups adequately, may lead to an improvement in the pituitary auto-contours.

The lacrimal and pituitary glands are small in volume and therefore if an auto-contour needs a minor adjustment, it may be more time efficient to begin again. Therefore it seems to be the case that lacrimal and pituitary auto-contouring is only of benefit when clinically acceptable contours are consistently produced, as opposed to producing auto-contours which may need minor manual adjustments.

Difficulty in performing high quality auto-segmentation of small structures in the brain has been previously reported in the literature (148). This is also supported by the results in this study shown in Figure 4-5. The number of variables in this study is however limited and the results could be improved by the inclusion of a greater number of structures with a greater variety of volumes. This should be investigated in the future to further analyse the relationship between the two variables.

The quality of the auto-contours of the chiasm, another small structure, was highly variable across patients with DSC values between 0.38 and 0.75 found (see Figure 4-3). Patients with a poorer quality chiasm auto-contour, which required minor adjustments, needed smoothing of the contours. Erroneous placement of the surface of the contour was reported, particularly towards the anterior section, as shown in Figure 4-16. Two auto-contours were found to be unacceptable due to missing sections of the chiasm. Small structures can be difficult to contour automatically, especially if the position varies between patients. The chiasm position, as well as the slope of the structure (which slopes superiorly and posteriorly), can vary between patients (142). This means that when a particular optical chiasm is located differently, with a different slope, to those in the training data, the auto-segmentation algorithm will likely perform worse than expert delineators. As discussed in Chapter 2, this is a known issue with atlas-based models and it is possible that a machine learning technique would perform better for these structures.

MRI is recommended for delineation of the chiasm (141, 149). This is supported by the inter-observer analysis performed in this study, where the mean, maximum and minimum CI is greater for MRI delineations compared to those contoured on CT (see Figure 4-10). Potentially, however, the clinical MRI sequence currently used for radiotherapy planning is not optimum for auto-contouring of this structure. The chiasm is surrounded by cerebrospinal fluid (CSF) (142). It is therefore possible that if information from an MRI sequence such a Fluid Attenuated Inversion Recovery (FLAIR), which suppresses CSF, were also acquired and used in the training along with the T_1 sequence, the delineation of

the outer surface of the chiasm would improve. This however would need to be validated clinically.

The cochleae are also small structures and therefore the DSC is likely to be lower than for larger structures, which was indeed generally observed (see Figure 4-4). Despite this, half of the contours were found to be clinically acceptable and, as Figure 4-11 shows, a relatively large number of patients had a greater CI between the auto-contour and manual delineation than the lowest inter-observer CI.

Again, the cochleae required smoothing which may be able to be performed automatically after the auto-contour generation. Figure 4-15 shows the only cochlea auto-contour which was found to be unacceptable, which was due to missing a large part of the structure. For this patient however the variation in inter-observer delineation was high and was comparable to that observed with the auto-contour (the reported CI was 0.10 and 0.13 for the inter-observer and auto-contour delineation respectively). A larger training cohort with the inclusion of more patients which require complex contouring may improve the auto-contour here.

In terms of improving the efficiency of the workflow, larger volume structures such as the brainstem may be more useful to delineate automatically. The brainstem auto-contours were found to only need minor adjustments, with one contour being clinically acceptable. The high quality of these auto-contours is also evidenced by the high DSC (describing the overlap of the contours) of > 0.8 observed for all patients (Figure 4-4). The minor adjustments were found to be due to erroneous contours on the superior and inferior slices, with the central slices appearing more accurate. A mean ratio of 0.73 was found for the CI of the superior and inferior slices to the CI of the central slices of the brainstem (see Figure 4-6), also indicating a poorer contour on the outermost slices relative to the central slices. Performing minor adjustments of the superior and inferior slices is likely to save contouring time over manually delineating the entire brainstem, however this would require further testing.

Similar results were also found for the spinal cord with issues being reported for the superior and inferior slices. A mean ratio of 0.66 between the CI of superior/ inferior slices and central slices was found (see Figure 4-6), with all ratios being less than 1, which supports the clinical findings. The DSC was typically lower than that of the brainstem (see Figure 4-4), however the mean MDC and CGD were improved compared to the mean inter-observer values (see Figure 4-8 and Figure 4-9), so this may not be of clinical significance. As well as this, the CI of the auto-contour compared to the manual delineation was greater than the lowest inter-observer CI for 67% of patients (see Figure 4-11). The surface of the spinal cord auto-contours was uneven and required smoothing for some patients however. It may be possible for this to be performed automatically after the auto-contour has been generated, again improving contouring efficiency.

All optic nerves required minor adjustments due to the uneven nature of the surface contour and discontinuities present in within the auto-contour (see Figure 4-13 as an example). It is likely that applying a smoothing function and potentially using a larger training database will improve these auto-contours. The globes were generally well delineated, with metrics being similar between auto-contours and inter-observer variation and all DSC values > 0.8 compared to Clinician I's contours. These results agree well with those reported in the prostate study (85). This is to be expected as the globes are relatively large structures, with low variation in anatomy between patients. As with other large volume structures, Figure 4-6 suggests that improvement in the superior and inferior extents of the globes could improve the quality of these auto-contours even further. The erroneous placement of part of the right globe delineation shown in Figure 4-14 is unexpected and requires further investigation. It is possible that a larger patient training cohort would improve this contour.

4.5 Conclusion

This study has investigated the clinical acceptability of automatically delineated brain OAR structures on MRI generated using a multi-atlas method. Larger structures such as the brainstem, spinal cord, globes and optic nerves appear to

be delineated well, with only minor adjustments needed if any. Automatically delineating these larger structures would likely improve the efficiency of the workflow compared to manually contouring the entire structure, as well as reducing inter-observer variability. Smaller structures, particularly the lacrimal and pituitary glands, but also the cochlea and chiasm in some cases, were poorly delineated automatically. These contours could however be improved potentially by including a more diverse patient cohort and including information from additional MRI sequences.

Chapter 5 MRI geometric system distortion assessment

5.1 Measuring MRI geometric system distortions

5.1.1 Measurement requirements

As discussed in Section 1.4.2.3, it is essential that the geometric integrity of MR images is known across the full FOV of the scanner when the images are to be used for MRI-only radiotherapy. Therefore prior to implementation, the baseline geometric distortions should be mapped so as to ascertain the radial distance over which the distortions are acceptable for radiotherapy planning. The acceptable level of distortion is likely to be related to the type of radiotherapy treatment used (see Section 1.4.2.3), however a maximum distortion tolerance of 2 mm is commonly cited in the literature (6).

In order to clinically implement MRI-only radiotherapy, it is also necessary to ensure that the geometric system distortions continue to be within clinically acceptable limits after the baseline measurements have been acquired. It is however impractical to measure the geometric system distortions before every patient, due to time constraints on the MRI scanner. Therefore the necessary frequency of QA measurements needs to be determined before the technique can be implemented. In order to determine this frequency, measurements of geometric distortions over the full FOV should be taken over a period of time so as to determine the temporal stability of the MRI system distortions.

To measure geometric distortions over the full FOV, a large FOV MRI-compatible phantom is required. The phantom should contain internal markers or a grid which are visible both on MRI and CT images. In this way, the measured position of the structures for a particular MRI sequence can be compared to the true position of the structures in a CT image, the latter modality being geometrically accurate and the current gold standard in radiotherapy geometric positioning.

5.1.2 Phantom description

For this project, the GRADE phantom (Spectronic Medical AB, Helsingborg, Sweden) was used. The phantom, consisting primarily of foam, contains 1177 polyethylene glycol-filled spherical markers measuring 17 mm diameter and spaced approximately 50 mm from each other. Towards the edge of the phantom, 30 mm marker spacing was present in order to achieve an improved distortion assessment resolution in this region. These markers are visible with both MRI and CT imaging.

This phantom was chosen over other available phantoms, such as the CIRS large field of view distortion phantom (model 604) (CIRS incorporated, Norfolk, Virginia, USA) and the MagPhan RT Phantom (The Phantom laboratory, Greenwich, New York, USA). The reason for this selection was that the GRADE phantom design was customisable to fill the effective FOV of the MRI scanners used in the study. Additionally, the GRADE phantom was lightweight, and contained a large number of markers to aid analysis of geometric distortion.

When ordering the phantom a questionnaire stating our requirements was completed. The phantom was customised to fill an effective FOV of $45 \times 45 \times 45\ cm^3$ (the maximum FOV of the MRI scanners in this study). Additionally, at the development phase, a specific request was made that a curved anterior edge be used in the phantom design as opposed to the cubic phantom design which was offered at the time. The reason for this was to ensure that no area of the effective FOV was excluded from the measurements, which would be more likely to be the case with the cubic phantom due to the cylindrical shape of the bore of the MRI scanner.

For this study, distortion stability measurements were acquired on four MRI scanners, with differing field strengths, manufacturers and based at different centres. These scanners were a Siemens Prisma 3 T scanner (Siemens Healthineers, Erlangen, Germany) based at Leeds General Infirmary, Leeds, a Siemens Aera 1.5 T scanner based at Leeds Cancer Centre, Leeds, a 3 T GE Signa PET-MR scanner (GE Healthcare systems, Chicago, USA) based at the

University of Newcastle, Newcastle-upon-Tyne and a Siemens Espree 1.5 T scanner based at the Northern Centre for Cancer Care, Newcastle-upon-Tyne. Two phantoms were therefore used in this study due to the large distance between centres (one for the Siemens Prisma and Siemens Aera in Leeds and one for the GE Signa and Siemens Espree in Newcastle-upon-Tyne). Both phantoms matched the specifications described previously. Figure 5-1 shows an image of the phantom set-up on the Siemens Aera scanner.

Figure 5-1 Image of the large FOV GRADE phantom on the Siemens Aera scanner. The cross-hairs on the anterior section of the phantom were aligned with the MRI lasers. The curved anterior surface was requested in order to fill the effective FOV of the scanner.

5.2 MRI geometric distortion temporal stability analysis

5.2.1 Introduction

Various studies (150-152) have investigated the magnitude of MRI system distortions by using large FOV grid phantoms. There has been limited work

however assessing the temporal stability of geometric distortions over the full FOV.

Moerland (153) investigated field error stability over a period of 18 months on a 1.5 T Philips Gyroscan S15 (Philips Medical Systems, Best, The Netherlands). Four points were examined, each 15 cm from the isocentre, in the central coronal plane only. The field error was found to be stable within experimental errors over this period of time.

Ahmed et al., (154) assessed distortion changes within a limited FOV over 17 months on a Philips Intera 1.5 T scanner. It was found that the pattern of the geometric distortion within the scanned volume (for a base of tongue patient) remained the same over the duration of the study. Distortion was assessed within 3 sagittal planes (in the central plane and ±35 mm from the magnet isocentre), enclosing the main organs of interest for base of tongue tumour patients.

Mizowaki et al., (155) assessed distortion using a 27 cm-sided cube on a scanner with a maximum FOV of 35 cm, which did not have the capability of modern distortion correction techniques. Measurements were performed on a permanent 0.2 T magnetic MRI system, as opposed to a superconducting high field scanner which are now most commonly found in clinical settings. Distortion changes over both short and long term intervals were investigated. Their results indicated that distortion variations may become larger with time and the authors commented on the need for a regular QA programme which analyses the same sequences as those used for radiotherapy treatment planning, although the frequency of such QA was not discussed.

These studies demonstrate that preliminary work has been undertaken regarding temporal stability of MRI distortions. For the purposes of MRI-only radiotherapy treatment planning however, it is essential to perform investigations of the temporal stability of MRI geometric distortions in three dimensions over the full FOV using modern clinical MRI scanners with modern distortion correction techniques applied. Assessment of the temporal stability of these distortions on

a range of scanners, with different manufacturers and field strengths i.e. both 1.5 T and 3 T which are commonly used in radiotherapy, needs to be performed. Such work has not been previously reported in the literature, but will be carried out in this study. Here, measurement of system-induced distortions on four MRI scanners over a period of one year was performed, in order to evaluate the temporal stability of the systems.

5.2.2 Method

Large FOV GRADE phantoms were scanned on four MRI scanners at four institutions at regular intervals of approximately once a month over one year. The scanners were a 3 T Siemens Prisma, a 3 T GE Signa PET-MRI, a 1.5 T Siemens Espree and a 1.5 T Siemens Aera.

For each scanner and sequence, 5 test-retest measurements were also acquired. Between each test-retest measurement the phantom was set up by removing and replacing it and then re-aligning the cross-hairs of the phantom with the lasers.

5.2.2.1 MRI scanning protocol

Scanning was performed using 2D multi-slice fast spin echo and 3D fast gradient echo sequences. The sequences were adapted from those recommended by the phantom manufacturer, with adjustments made on a per-scanner basis in order to produce an acceptable image quality within time and machine constraints (full details of the MRI parameters for each scanner and sequence can be seen in Table 5-1). The scanning was performed in the axial plane. Frequency-encoding was performed in the left – right (x) direction and phase-encoding in the anterior – posterior (y) direction. Slices were acquired in the superior – inferior (z) direction. An illustration of the x, y and z dimensions can be found in Figure 1-2. For the chosen sequences, it was ensured that the resulting image quality and marker visibility was acceptable, meaning that distortions could be assessed and that the entirety of the phantom was sampled. The 3D distortion correction provided by the MRI manufacturer was applied to the acquired scans.

Sequence	2D Fast Spin Echo				3D Fast Gradient Echo			
Scanner	**Siemens Prisma**	**GE Signa**	**Siemens Espree**	**Siemens Aera**	**Siemens Prisma**	**GE Signa**	**Siemens Espree**	**Siemens Aera**
Field of view/ mm	500 x 500	500 x 500	450 x 450	500 x 500	500 x 500	500 x 500	450 x 450	500 x 500
Acquisition matrix	512 x 512	512 x 512	448 x 448	512 x 512	512 x 512	512 x 512	448 x 448	512 x 512
Slice thickness/mm	4.0	2.5	2.5	4.0	1.95	1.0	1.0	1.95
Slice number	128	200	128	128	256	508	256	256
Slice gap/ mm	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0
Echo time/ repetition time/ ms	94/ 15000	96/ 15000	17/ 11560	94/ 15000	1.58/ 4.4	1.3/ 3.8	2.39/ 4.81	1.58/ 4.4
Flip angle/°	130	130	130	130	20	20	20	20
Pixel bandwidth/ HzPixel^{-1}	391	250	385	391	490	250	510	490
Signal averages	1	1	1	1	1	1	1	1
Receive coils	Integrated body	Integrated body	2 flex and spine	Integrated body	Integrated body	Integrated body	2 flex and spine	Integrated body
Acquisition time/ s	1020	1666	1077	1020	576	1244	553	576

Table 5-1 MRI sequence parameters used on each scanner. Flip angle refers to the refocusing flip angle for the spin echo sequence.

5.2.2.2 CT scanning protocol

CT scans (120 kVp, 32 mA, 1 s exposure time) of the phantoms were acquired in order to ensure that the marker spheres were intact and to provide a reference image for the distortion calculations. Scanning was performed with a Siemens Somatom scanner for the first phantom (used with the Siemens Aera and Siemens Prisma MRI scanners) and a Siemens Sensation scanner for the second phantom (used with the GE Signa and Siemens Espree MRI scanners). A FOV of 650 x 650 mm (512 x 512 pixels) was selected and the full length of the phantom was imaged ensuring that the entirety of the phantom was scanned. A slice width of 2 mm was used. CT phantom scans were also acquired at the end of the measurement period to confirm that no changes in the phantom structure had occurred. This was achieved visually as well as by calculating subtraction images of the two scans. A transaxial CT slice and MRI slice through the first phantom can be seen in Figure 5-2 and Figure 5-3 respectively.

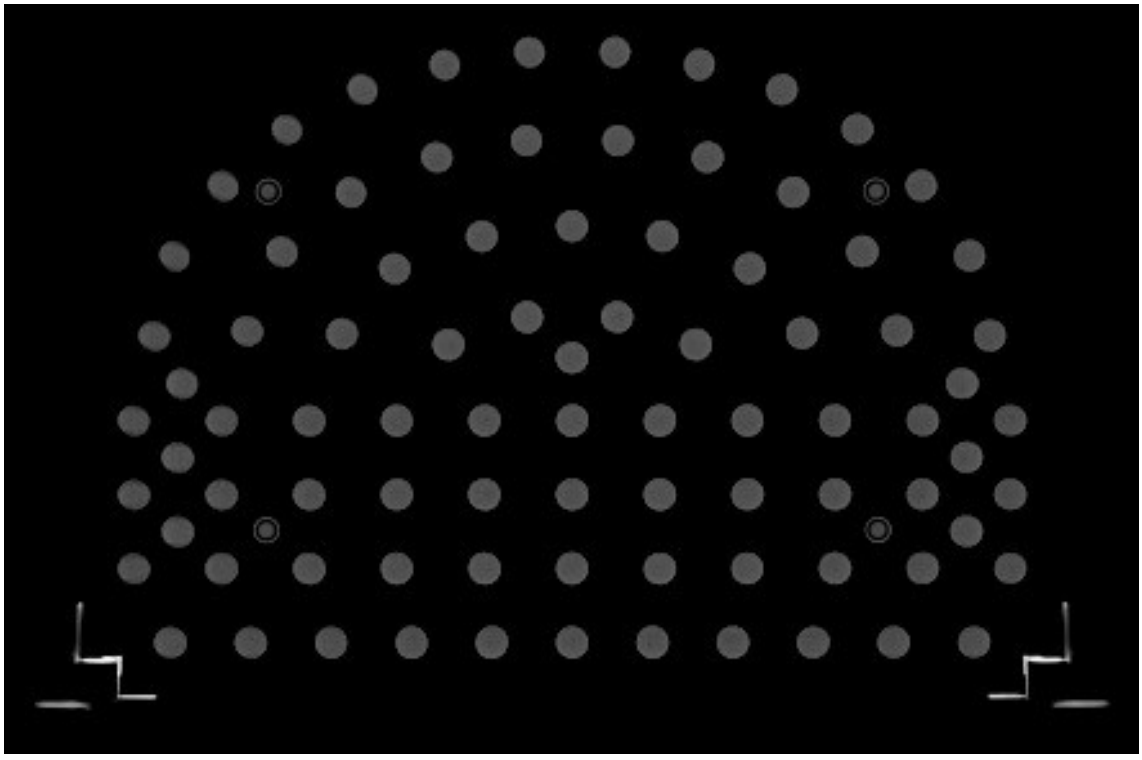

Figure 5-2 A CT slice through a large FOV GRADE phantom. Contrast agent filled spheres can be seen throughout the phantom. It should be noted that 4 screws can also be seen in this image (visible as circles with a concentric ring around them).

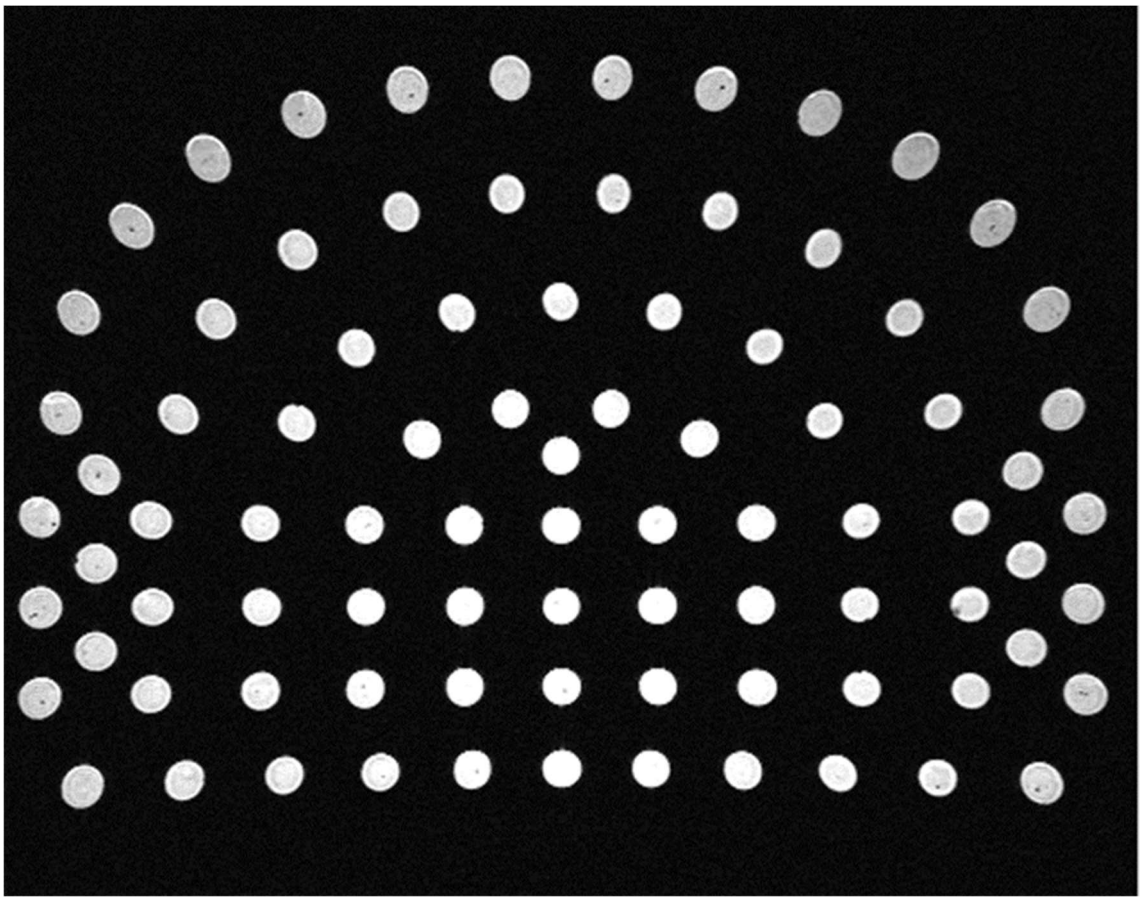

Figure 5-3 An MRI slice through a large FOV GRADE phantom. Three-dimension distortion correction (provided by the scanner manufacturer) has been applied to the acquisition.

5.2.2.3 Temporal stability analysis

Results were analysed using the Spectronic Medical GRADE evaluation software (Spectronic Medical AB, Helsingborg, Sweden). This software returns the centroids of the scanned marker positions as x, y and z co-ordinates with reference to the phantom isocentre, as well as the corresponding modelled marker positions (based on the CT images of the phantom). During analysis, the original CT scans are interpolated to a 1 x 1 x 1 mm resolution.

The distortion of the markers i.e. the absolute distance between the centroids of the markers in the MR image and the corresponding marker positions in the phantom model, as well as the radial distance to isocentre, were calculated for each identified marker. This was performed through an in-house script written in

Python(x,y) (version 2.7.10.0, https://python-xy.github.io). In-house scripts in this language were also used for all further analysis in this section unless otherwise stated.

5.2.2.3.1 Comparison of monthly and test-retest distortion variation

For each scanner and sequence, the range and standard deviation (SD) of the distortions of each marker in the image over the 5 test-retest measurements were calculated (termed $R_{Tr,M}$ and $SD_{Tr,M}$ respectively, where the notation Tr stands for test-retest measurements and M represents each marker and ranges from 1 to the number of markers identified). The same was also calculated for an equal number of measurement points acquired over 5 sequential months (termed $R_{5,M}$ and $SD_{5,M}$ respectively). These sequential months were normally the first 5 monthly measurements, apart from if a larger than average measurement gap fell within this period, in which case a later measurement period was used. The two results were compared for markers identified in all scans for each scanner and sequence. It was determined whether there was a statistically significant difference between the two using a Wilcoxon signed-rank test (where $p<0.05$ was considered significant). The Wilcoxon signed-rank test was performed in the SPSS Statistics software (IBM Corp., Armonk, New York, USA, Version 21).

If any statistically significant difference in range and SD was observed for a scanner and sequence the same comparison was performed for distortions in the three orthogonal directions, x, y and z in order to determine if there was any directional-dependency in the high distortion variation (as different gradients are used in different directions). Here the z dimension is defined as the direction parallel to the main magnetic B_0 field. The x and y dimensions are perpendicular to this in the in-plane directions (see Figure 1-2).

5.2.2.3.2 Identification of highly deviating distortion points

For each monthly measurement, it was determined if any individual marker distortions were greater than their respective mean baseline distortion from the test-retest measurements for that scanner and sequence plus or minus 3 $SD_{Tr,M}$.

According to Chebyshev's inequality, it is unlikely that a distortion value will differ from the mean distortion value by more than 3 SDs (with a probability of less than 11.1 %)(156). Therefore this assessment allows determination of any marker distortions which vary more than expected from their baseline distortion measurement. Chebyshev's inequality assumes that $SD_{Tr,M}$ and the mean test-retest distortion value are equal to the standard deviation and mean of the total population of distortion measurements respectively, which is unlikely to be the case. It is however a useful criterion for identifying higher than expected marker distortion values.

As an additional criterion, it was determined if any individual marker distortions were greater than 1 mm (the resolution of the interpolated CT scans) from their baseline mean test-rest value. Markers which met both of these criteria were recorded for each month. The distance-to-isocentre values for these markers were also recorded.

5.2.2.3.3 Additional tests of distortion stability over a year

Histograms of the distortions were plotted for every measurement over a year for all scanners and sequences. As well as plotting histograms of the marker distortions, histograms of the distortions in the x, y and z dimensions were also produced. For each histogram the kurtosis and skew were calculated. The kurtosis of a histogram is a measure of the combined weight of the tails compared to the remaining distribution, whilst the skewness describes the symmetry of the distribution with a larger absolute value indicating a greater asymmetry (157). Fisher's definition of kurtosis was used in this study meaning that a normal distribution would have a kurtosis value of 0, as well as a skewness value of 0. The mathematical formulas (157) for the calculation of skew, *sk*, and Fisher's definition of kurtosis, *kt*, can be found in Equation 5-1 and Equation 5-2 respectively. Here, d_i, is the distortion value of the *i*th data point in mm, where there are a total of *N* data points. $\bar{d}$ and sd are the mean and standard deviation of the distortion values respectively, both with units of mm.

$$sk = \sum_{i=1}^{N} \frac{\left(d_i - \bar{d}\right)^3 / N}{sd^3}$$

Equation 5-1

$$kt = \sum_{i=1}^{N} \frac{\left(d_i - \bar{d}\right)^4 / N}{sd^4} - 3$$

Equation 5-2

The kurtosis and skew values were tracked over the year in order to assess variation of the shape of the distortion histograms. The SD and mean of the skew and kurtosis values over the entire measurement period were determined for each scanner and sequence.

In order to assess how distortions can vary with time, for each marker the maximum variation in distortion relative to the median distortion value over the year was calculated for each scanner and sequence. These values were investigated as a function of distance-to-isocentre. The median value was chosen for this investigation rather than the mean in order to ensure that outliers did not adversely affect the results.

5.2.3 Results

The mean distortion for radial distance-to-isocentre values above and below 200 mm over the measurement period can be seen in Table 5-2. 200 mm was chosen as a value of clinical interest based on the size of the patient contour on head and neck scans. In addition, the mean range and SD of the distortion values for each scanner and sequence over this period are also shown in Table 5-2. The results of the Wilcoxon paired-samples tests can also be seen (the corresponding p-value is given for each scanner and sequence). Here the range and SD of marker distortions were compared for measurements over 5 months and over 5 test-retest measurements.

Scanner and Sequence	Mean distortion (for DI ≤ 200 mm)/ mm	Mean distortion (for DI > 200 mm)/ mm	Mean distortion range/ mm	Mean SD of distortions/ mm	Wilcoxon p-value	
					Range	SD
Prisma 2D	0.6	2.8	0.9	0.3	**<0.01**	**<0.01**
Prisma 3D	0.6	1.8	0.8	0.2	0.52	0.71
Signa 2D	0.5	3.1	1.2	0.4	**<0.01**	**<0.01**
Signa 3D	0.5	3.1	0.8	0.3	0.41	0.73
Espree 2D	1.2	3.5	1.2	0.4	0.26	0.61
Espree 3D	0.8	1.9	1.1	0.3	**0.03**	0.09
Aera 2D	0.7	9.0	2.0	0.7	**<0.01**	**<0.01**
Aera 3D	0.6	2.2	0.8	0.3	**<0.01**	**<0.01**

Table 5-2 The mean distortion for distance-to-isocentre (DI) values of ≤200 and >200 mm, as well as the mean distortion range and standard deviation (SD), for all scanners and sequences over the approximately yearlong measurement period. The p-values (values considered significant are highlighted in bold) of the paired samples Wilcoxon tests comparing the marker ranges and standard deviations between 5 test-retest measurements and 5 monthly measurements are also shown.

The mean distortion for distance-to-isocentre values of less than 200 mm varied between 0.5 - 0.8 mm for most scanners and sequences over the year. The only exception was for the shorter bore Siemens Espree, where a mean distortion of 1.2 mm for the 2D sequence was found. At greater distance-to-isocentre values (above 200 mm), mean distortion values between 1.9 and 3.5 mm were generally observed. The greater mean distortion for the Siemens Aera 2D sequence of 9.0 mm is a concern, however it should be noted that distortions of this magnitude are not observed below 235 mm distance-to-isocentre.

For the Siemens Espree, no statistically significant difference was observed for either the range or SD of distortions between 5 measurements acquired over a period of months and the test-retest measurements for the 2D sequence. This was also true for the SD of distortions for the 3D sequence. The p-value of 0.03 when comparing the distortion ranges for the 3D sequence suggested that a statistically significant difference between the two could not be ruled out.

For the Siemens Prisma and GE Signa no evidence of a statistically significant difference for either the range or SD was observed for the 3D sequence when comparing the monthly measurements to the test-retest measurements. However for the 2D sequence, both marker distortion range and SD were significantly larger over 5 monthly measurements compared to the test-retest measurements on both scanners.

For the Siemens Aera scanner, a statistically significant difference in distortion range and SD between the 5 test-retest and 5 monthly measurements were observed for both the 2D and 3D sequences.

The significant difference, with monthly variation being larger than variation over the test-retest measurements, was observed in all dimensions for the Siemens Prisma and GE Signa 2D sequences and for the Siemens Aera 2D and 3D sequences. The only exceptions to this were in the x dimension for Siemens Prisma 2D sequence and in the z direction for the Siemens Aera 3D sequence.

The maximum and minimum percentages of marker distortions which were outside the 3 $SD_{Tr,M}$ and 1 mm limits of their respective mean baseline values over the monthly measurements can be seen in Table 5-3. The maximum failure rate was greater for the 2D sequences compared to the 3D sequences across all scanners. In general, there was also less variation for the 3D sequences. No trend was seen in the percentage failures with time and the variation appeared random.

An example of a distortion histogram for a typical month can be seen in Figure 5-4. This histogram has been binned into 1 mm bin sizes, due to the tolerance of 1 mm typically being used to assess distortion acceptability (see Section 1.4.2). The mean and SD of the kurtosis and skew values of the distortion histograms over all monthly measurements for each scanner and sequence can be seen in Figure 5-5 and Figure 5-6 respectively. The results for histograms of distortions in the x, y and z dimensions are also presented.

Scanner	Sequence	Minimum percentage failure/ %	Maximum percentage failure/ %
Siemens	2D	0.4	6.1
Prisma (3T)	3D	0.5	2.0
GE Signa (3T)	2D	0.5	13.0
	3D	0.0	2.1
Siemens	2D	0.0	5.9
Espree (1.5T)	3D	0.2	2.6
Siemens Aera	2D	7.5	11.5
(1.5T)	3D	3.1	7.5

Table 5-3 The maximum and minimum percentages of markers failing the criteria of being within 3 SDTr,M or 1 mm from their mean baseline distortion value over all the months for each scanner and sequence.

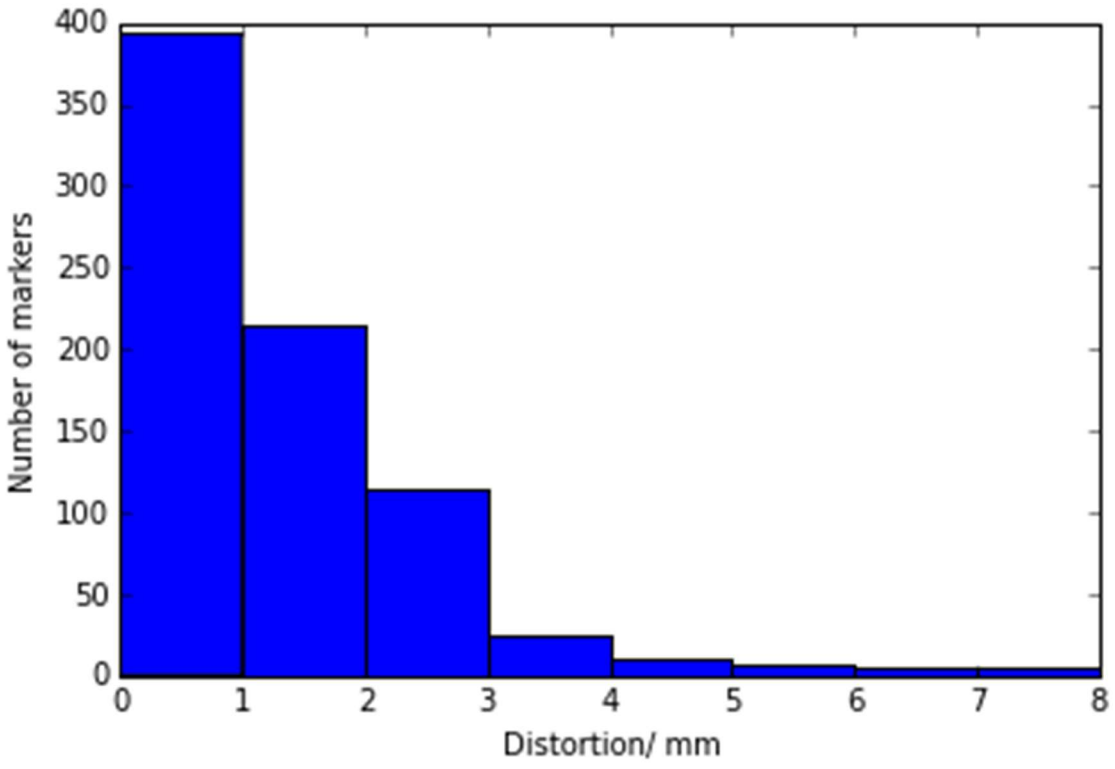

Figure 5-4 The histogram of distortions for one month on the Siemens Prisma scanner (for the 3D sequence).

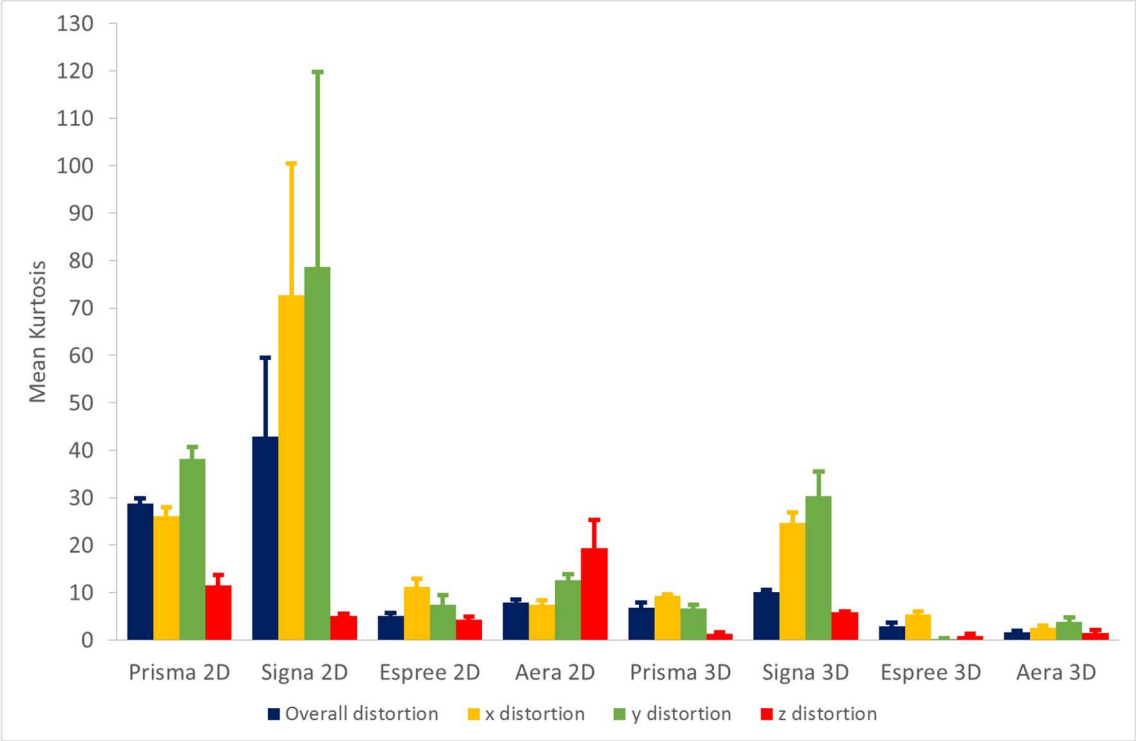

Figure 5-5 The mean kurtosis for each scanner and sequence. The error bars (beginning at the top of the column) are equivalent to one standard deviation of the kurtosis values.

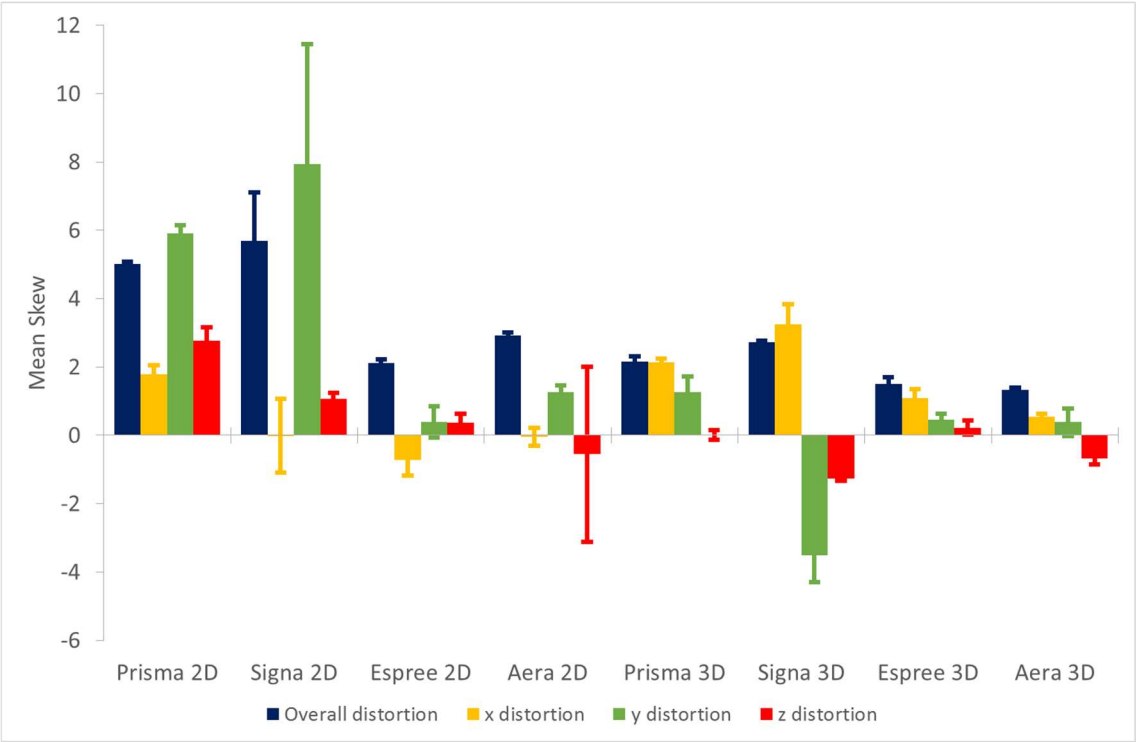

Figure 5-6 The mean skew for each scanner and sequence. The error bars (beginning at the non-zero value of the column) are equivalent to one standard deviation of the skew values.

It was found that the distortion histograms for the GE Signa 2D sequence had a higher than expected (2 SD above the average) mean kurtosis value, implying a heavy histogram tail and a higher than average number of outlier values. This was also true of the mean kurtosis values for distortion histograms in the x and y direction. In addition, the SDs of the kurtosis and skew values over the monthly measurements were also greater than expected. The only exception to this was in the z direction.

For the Siemens Prisma 2D sequence distortion histograms, the mean of the skew values was greater than expected in the z direction, suggesting a trend towards larger distortions. For the Siemens Aera 2D sequence distortion histograms, the SDs of both the kurtosis and skew values in the z direction were greater than expected, implying greater variation in these values. This was also true for the mean value of the kurtosis in the z direction.

Maximum deviations from the median marker distortion values were below 2 mm, apart from for the GE Signa 2D sequence, Siemens Prisma 2D sequence and the Siemens Aera results (Figure 5-7). The maximum deviation in distortion from the median value over the yearly measurements was found to generally increase with distance-to-isocentre.

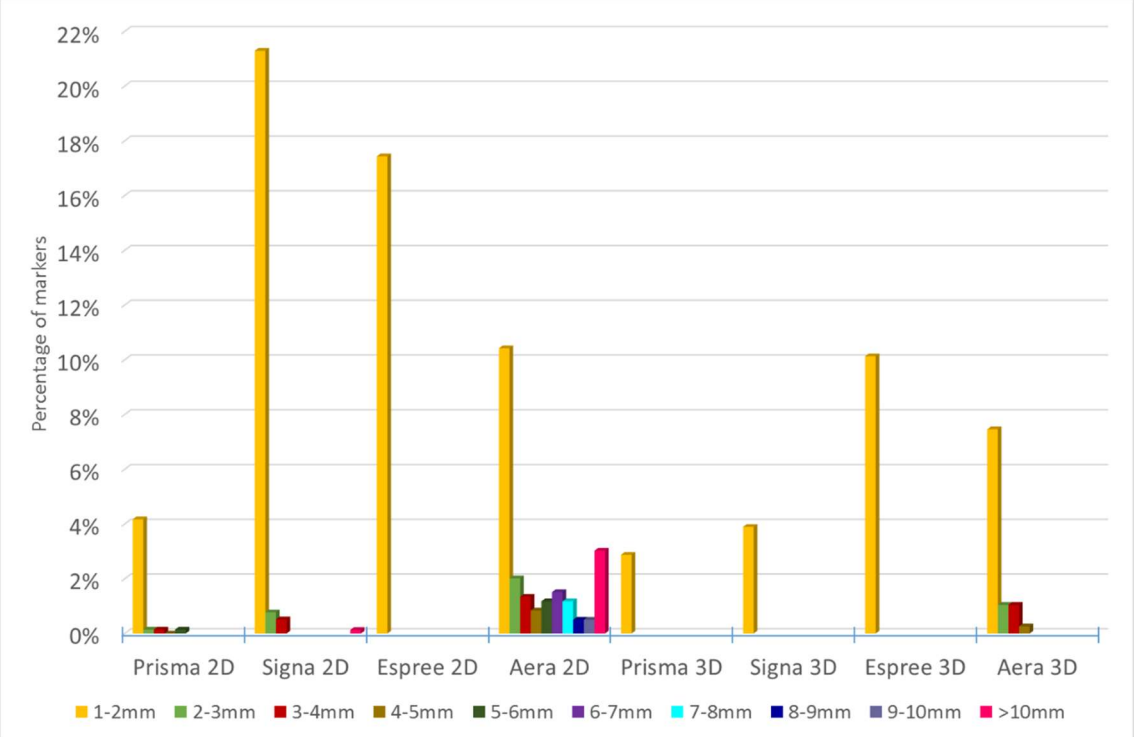

Figure 5-7. The maximum deviation from the median distortion over a year for each marker. The percentage of markers falling within each category of maximum deviations (in millimetres) is shown for each scanner and sequence. Maximum deviations below 1 mm are not shown, to aid visualisation.

5.2.4 Discussion

The results presented in this study regarding distortion magnitude agree with previous findings in the literature. Torfeh et al., (152) found mean distortions of less than 1 mm for radial distances up to 200 mm which was the case for the majority of scanners and sequences (see Table 5-2). Mean distortions of up to 4 mm (in-plane) were seen at greater radial distances (152). Paulson et al., (158) commented that although 3D vendor distortion correction improved results,

distortions larger than 2 mm persisted at large FOVs, agreeing with the findings presented in this work.

The mean distortion ranges were generally between 0.8 and 1.2 mm, however the Aera 2D sequence showed a greater mean range of 2.0 mm over the measurement period. It should be noted however that when one month in the middle of the measurement period was excluded from the results, the mean distortion range was reduced to 1.2 mm, in line with other results. This suggests that the results from this month may be an anomaly and examination of the scan indeed showed a slight twist in the phantom set-up. The average SD of marker variation was low with most scanners and sequences having values of less than 0.4 mm.

The main aim of this study was to determine the temporal variation of multiple MRI scanners over the full FOV. Other studies have been undertaken in the literature regarding temporal variation of MRI distortions, however mostly using smaller FOV phantoms.

Some studies observed a small number of points within a limited FOV, reporting a stable distortion variation. For example, Moerland (153) investigated field error stability by examining four points, each 15 cm from the isocentre, in the central coronal plane only over a period of 18 months. The field error was found to be stable within experimental errors over this period of time. Mah et al., (159) selected 4 randomly distributed points in a grid phantom in each of the cardinal planes and recorded their locations over 3 months, reporting mean SDs of 0.7, 1.1 and 1.2 mm in the superior-inferior, anterior-posterior and medial-lateral dimensions respectively. Using a small FOV 2D peg phantom, less than 2 mm variations in distortions were recorded over 8 months.

Others carried out a more extensive analysis, again reporting stability, however still within a limited FOV. Ahmed et al., (154) found that the pattern of geometric distortion within the scanned volume (for a base of the tongue patient) remained the same over the 17 month duration of the study. Distortion was assessed within

3 sagittal planes (in the central plane and ±35 mm from the magnet isocentre), enclosing the main organs of interest for base of tongue tumour patients.

Price et al., (160) assessed distortion stability in the 3 cardinal planes over 6 months using a 2D phantom with dimensions 36 x 43 x 2 cm^3 which contained 255 capsules. Good temporal stability was found, with 95% of voxels within 0.6 mm of their baseline measurements.

Baldwin et al., (161) analysed distortion stability over a 26 x 26 x 20 cm^3 volume of a 3T MRI scanner. As a test of reproducibility, the distortion was measured over several months and errors less than the pixel dimension were reported. The authors acknowledged however that a larger volume phantom would be necessary in order to investigate distortion of clinically relevant volumes.

In this study distortion stability has been assessed over the full FOV in order to ensure that clinically relevant volumes have been included. Evidence has been found which supports the temporal stability of MRI distortion reported in these smaller FOV studies. The 1.5 T Siemens Espree 2D and 3D sequences, the 3 T GE Signa 3D sequence and the 3 T Siemens Prisma 3D sequence showed no evidence of a significant difference in distortion SD between monthly measurements and test-retest measurements. This was also true for the distortion range in these cases, apart from for the Siemens Espree 3D sequence where a significant difference could not be ruled out. This apparent temporal stability was also observed in the other investigations in this work. All maximum deviations from the median distortion values were less than 2 mm. The SD and mean values of the kurtosis and skew of the monthly distortion histograms showed no results that were greater than expected. For these scans, the number of markers which failed to be within 3 $SD_{Tr,M}$ or 1 mm from their respective mean test-retest measurement values was also relatively low.

For these scanners and sequences it therefore appears unnecessary to perform MRI distortion QA on a monthly basis and a more infrequent QA measurement frequency seems justified, for example on a biannual basis.

There is however literature reporting evidence of MRI distortion instability and which have pressed the need for regular QA when the images are used for radiotherapy. Mizowaki et al., (155) assessed distortion using a 27 cm-sided cube on a scanner with a maximum FOV of 35 cm, which did not have the capability of modern distortion correction techniques. The magnet was a permanent 0.2 T magnet. Distortion changes over both short (8 day) and long term (6 month) intervals were investigated. Their results indicated that distortion variations may become larger with time and the authors commented on the need for a regular QA programme which uses the same sequences as those used for radiotherapy treatment planning.

Kapanen et al., (162) measured geometric distortions on a 1.5 T MRI scanner for the purposes of commissioning MRI-only radiotherapy. An 18 cm spherical grid phantom and the surface of a 17 cm diameter spherical uniform phantom were used to analyse distortion measurements. Visually no differences were observed in the distortions measured using the grid phantom (at a 3 month interval), however a significant time trend was seen in the aberration of the geometrical distortion for the surface of the uniform phantom (for measurements taken daily over a year). The authors concluded that for MRI-only radiotherapy, regular distortion measurements should be carried out in order to ensure the consistency of the static and gradient fields (162).

The 1.5 T Siemens Aera 2D and 3D sequences, the 3 T GE Signa 2D sequence and the 3 T Siemens Prisma 2D sequence in this study showed evidence of distortion variation. For these scans a statistically significant difference was observed in marker distortion range and SD when comparing the monthly measurements to the test-retest measurements, with measurements over a period of months showing a greater variation. This was also found to be generally true for the distortion results in all three dimensions. For the Siemens Prisma 2D and Siemens Aera 3D sequences however this was not found to be the case for the x and z dimensions respectively. This suggests a temporal stability in

distortions in these directions, potentially due to the use of different gradients, and shows the importance of analysing distortion results in separate dimensions.

Damyanovich et al., (163) analysed variations in overall distortion as well as in separate dimensions using a small FOV 10 x 11 x 17 cm^3 phantom for the purposes of stereotactic radiosurgery. Measurements were taken on a 1.5 T and 3 T scanner over 32 months. As we observed in some cases, deviations were found to vary across the different components, with the largest in the z dimension, attributed to the poorer gradient linearity in this direction.

These scans (1.5 T Siemens Aera 2D and 3D sequences, the 3 T GE Signa 2D sequence and the 3 T Siemens Prisma 2D sequence) also had some large (> 2 mm) variations in monthly distortion measurements relative to the median distortion value. It should be noted however that these occurred at large distance-to-isocentre values (> 220 mm), so may not be clinically relevant for all patients and anatomical sites, for example this would not be a clinical issue for brain or head and neck cancer patients.

These scans had a relatively high number of markers failing the criteria of being within 3 $SD_{Tr,M}$ or 1 mm of the baseline test-retest mean value for some or all of the monthly measurements. These high failure rates were not always consistent across time however. The GE Signa and Siemens Prisma had 2 and 3 months respectively with relative low failure rates (below 2 %). The results for the Siemens Aera are more consistent, suggesting a poorer scanner distortion stability, however some monthly variation can be seen. The observed higher failure rates for these scanners support initial results suggesting that the range and SDs of monthly distortions were higher than for the test-retest measurements.

The results of these 2D scans showed comparatively larger variation when analysing the distortion histograms. The GE Signa exhibited this greater variation in the kurtosis and skew values across all dimensions, except in the z direction, suggesting instability in the shape of the histogram distortions and therefore

random variation in the distortion field over time. These scans also showed a comparatively larger mean kurtosis value in the same dimensions, indicative of heavy histogram tails with a large number of outliers. This suggests a large spread of distortions for these dimensions. The Siemens Aera 2D scans showed the same results but in the z direction only. The Siemens Prisma 2D sequence histograms showed a comparatively greater mean skew value in the z direction, meaning the histograms were more skewed than expected towards positive and larger distortions.

The observed distortion variation for these scanners and sequences, possibly caused by minor gradient instability, suggests that more frequent QA measurements may be needed if these sequences were used clinically, for example on a monthly basis. This investigation has shown occasional differences in distortion stability results in the z dimension (parallel to the main B_0 magnetic field) compared to the perpendicular in-plane dimensions. This may be a result of a difference in gradient design or shimming in this dimension (150). It is important that where temporal distortion instability is noted, it is investigated whether there is a dimensional dependency involved.

This work has shown that the temporal distortion stability of MRI fields can vary between scanners and sequences. As a result of this, it is recommended that for each MRI scanner to be used for MRI-only radiotherapy:

- The QA frequency should be determined based on the variation observed over a period of months compared to initial test-retest measurements for relevant clinical sequences.
- In order to identify when discrepancies occur, it is useful to:
    - Track the variation in distortion histograms over time,
    - Track the deviations of markers from their median distortion values over time.

Only one of each scanner type was tested in this study and the results are specific to the individual scanner tested and may be partly related to specific installations,

for example helium level and shimming at commissioning. It is not possible therefore to comment on the generalisability of these results to the same scanner type in different centres. Additionally, due to different machine constraints, there was some variation in the sequence parameters, such as bandwidth, between different scanners. Therefore whilst the results from this study are useful in analysing the distortion variation of an individual scanner over time (where the parameters were kept constant), it was not the aim, nor indeed possible from these results, to discuss differences in variation between scanner types and manufacturers.

It should be noted that this investigation has been solely focused on MRI system-induced distortions. It is also necessary to understand the effects of patient-induced distortions on individual MR images. Steps should be taken in order to ensure that MRI sequence parameters are chosen to minimise the effects of these distortions i.e. through appropriate selection of bandwidths. For QA related to patient-induced MRI distortions, anthropomorphic phantoms would be required. This is discussed more in Chapter 6.

5.2.5 Conclusion

The temporal stability of four MRI scanners with differing field strengths and manufacturers was assessed. The temporal stability of the MRI geometric distortions varied across scanners and sequences, with some demonstrating a good temporal stability. Others, however, showed more variation over the year, suggesting that more frequent distortion QA would be needed when using these MR images for radiotherapy purposes. It is therefore necessary prior to implementing an MRI-only radiotherapy pathway to determine the temporal distortion stability of the individual MRI scanner for clinical MRI sequences. This indeed is also true for any situation where it is essential to know that the geometric integrity of the MR image is clinically acceptable across the full FOV at any time.

5.3 Investigation of GRADE phantom susceptibility

5.3.1 Introduction

At boundaries between tissues of different densities, magnetic susceptibility effects may occur which can induce inhomogeneities in the B_0 field, as discussed in Section 1.4.2.2. This is not only the case for human tissues, but also for phantoms containing transitions between different materials, as well as at phantom-air boundaries at the outer edge of the phantom. In order to ensure that these effects did not substantially influence the measurement of system distortions it was necessary to measure the effects of the magnetic susceptibility of the phantom.

It is known that in MRI scanners the geometric system distortions predominately arise from gradient-nonlinearities, rather than inherent inhomogeneities in the main B_0 field (44). Magnetic susceptibility effects will however add to this inherent inhomogeneity in the field. It can be tested whether the distortion caused by the total B_0 inhomogeneity (including both system and phantom-induced distortions) is small compared to the distortion caused by gradient non-linearities. This would test the hypothesis that phantom susceptibility is not a significant cause of the distortions in the phantom measurements.

The magnetic field homogeneity, H_B of the scanner can be assessed using the method described in Chen et al., (164) where H_B (measured in parts per million (ppm)) is defined using Equation 5-3. As the magnetic field inhomogeneity only affects measurements in the readout direction (165), x, it is only necessary to calculate the inhomogeneity in this direction. In Equation 5-3, $x' = x + \Delta B_0(x,y)/G_x$ is the x position which has been affected by the distortion created by $\Delta B_0(x,y)$, the magnetic field inhomogeneity at the point (x, y), and where G_x is defined as the linear field gradient in the x direction (164). If two scans are acquired with different bandwidths, termed BW_1 (Hz) and BW_2 (Hz) respectively then the difference between the measured distortion in each of the scans in the x direction $(x_1' - x_2')$ (mm) is proportional to H_B. If the image FOV in the x direction, FOV_x (mm) and the magnetic field strength B_0 (T) are known, then H_B

can be calculated. In Equation 5-3 $\gamma = \gamma/2\pi$, where γ is the gyromagnetic ratio, and is equal to 42.576 MHz.T^{-1} (54).

$$H_B = \frac{BW_1 \times BW_2 \times (x_1' - x_2')}{\gamma \times B_0 \times FOV_x \times (BW_2 - BW_1)}$$

Equation 5-3

5.3.2 Method

MRI measurements of the GRADE phantom were acquired using a 2D multi-slice spin echo sequence on the 3 T Prisma scanner, with frequency encoding performed in the right-left (x) direction. Similar to the method used in Walker et al., (54) images of the phantom were acquired using this sequence, whilst varying the bandwidth. Spectronic Medical AB recommend that a bandwidth of approximately 390 Hz.pixel^{-1} be used for phantom measurements using this sequence. A lower bandwidth of approximately one quarter of this value (107 Hz.pixel^{-1}) was used for these measurements, as well as a higher bandwidth, near the maximum value, of 751 Hz.pixel^{-1}.

The distortion of each marker was determined as in Section 5.2.2 for measurements acquired at the two bandwidths. The difference between the distorted x-coordinates for the 751 and 107 Hz.pixel^{-1} bandwidths was calculated for each of the markers. By substituting this difference, $(x_1' - x_2')$, into Equation 5-3, the field inhomogeneity, H_B, at each marker position was determined. It should be noted that the bandwidths were converted to units of Hz by multiplying the values with the number of pixels in the readout direction, prior to substitution into Equation 5-3.

The offset, *o*, (mm) that this level of inhomogeneity would cause at the normal measurement bandwidth, BW, of 391 Hz.pixel^{-1} was then calculated at each marker position using Equation 5-4. Here, f_R is the resonant frequency of hydrogen at 3 T and equal to 127.7 MHz (166) and w is the pixel width in the x direction and equal to 0.9766 mm.

$$o = \frac{H_B \times f_R \times w}{BW}$$

Equation 5-4

Reference images were also acquired at a bandwidth of 391 Hz.pixel^{-1} using the same methodology as described in Section 5.2.2.1. The x component of distortion at each marker position was determined using the Spectronic software as outlined in Section 5.2.2.3.

The maximum and mean values of the absolute marker offset caused by field inhomogeneities and the measured absolute distortion in the x direction across all marker positions were calculated. It was also determined whether the absolute offset values were significantly smaller than the absolute x distortion values using a Wilcoxon signed rank test (a result of $p < 0.05$ was considered to be significant).

The difference between the absolute distortion in the x direction and the absolute inhomogeneity offset at each marker position was also determined. The maximum, minimum and mean of these differences across all marker positions were calculated. If any positions were found where the absolute offset was > 1 mm larger than the absolute x distortion, the position of the marker was noted.

5.3.3 Results

In total 445 markers were identified by the software. Table 5-4 shows the calculated maximum and minimum of the field inhomogeneity offset values and x distortion values across all marker positions.

	Maximum value/ mm	Mean value/ mm
Offset due to field inhomogeneities	2.2	0.3
Distortion in the x direction	15.0	0.5

Table 5-4 The maximum and mean values of the offset due to field inhomogeneities and the distortion in the x direction.

When comparing the offset caused by the field inhomogeneity with the measured distortion in the x direction, a p-value of <0.001 was found between the two, indicative of a significant difference. It was found that the difference was due to significantly smaller absolute offset values compared to the absolute distortion values in the x direction.

The mean difference between the absolute distortion in the x direction and the absolute offset due to field inhomogeneity across all marker positions was 0.3 mm. The maximum and minimum differences were 14.9 mm and -1.9 mm respectively. Here, a positive value indicates that the absolute x distortion was greater than the absolute offset and a negative value indicates that the absolute offset was larger than the absolute x distortion. 7 out of the 445 markers were found to have a difference of < -1 mm i.e. where the absolute offsets were greater than 1 mm larger in magnitude than the absolute x distortions. These were all positioned at distances-to-isocentre of ≥ 240 mm.

5.3.4 Discussion

It has been shown that the offset resulting from the inhomogeneity in the B_0 field (both inherent and induced by the phantom) was significantly lower than the measured distortion in the x-direction. This was found when frequency encoding was performed in the left-right direction, as was the case for all measurements in the temporal stability study. This is also supported by the larger maximum and mean values of the x distortion compared to the offset (see Table 5-4), as well as the mean difference between the two at each marker position, which demonstrated that, on average, the absolute x distortion was greater than the absolute inhomogeneity offset.

The large maximum difference between the two (of 14.9 mm) shows that the offset due to field inhomogeneities can be dwarfed by the distortion in the x direction. A large (> 1mm) difference where the absolute offset was greater than the absolute x distortion was found for 7 marker positions, however these were all positioned at distances-to-isocentre of ≥ 240 mm. This is likely to be caused

by the inherent inhomogeneity of the magnetic field, which is more significant at large radial distances-from-isocentre (4).

Although the B_0 inhomogeneity resulting from the magnetic susceptibility of the phantom was not separated from the inherent inhomogeneity in the magnetic field, it has been shown that the measured distortion is not dominated by either of these two effects, except at very large distances-to-isocentre. It is likely therefore that the measured distortion in Section 5.2 is predominantly caused by non-linearity of the gradient fields.

5.4 Performing validation of external software related to the GRADE phantom

5.4.1 Aim

The Spectronic Medical AB software has been used in this study in order to determine the centroids of markers in the CT and MR images. These centroid positions have then been used in further analysis of the distortions. The aim of this study is to determine the accuracy of this centroid placement in order to validate the Spectronic Medical AB software.

5.4.2 Method

Spectronic Medical AB software returns the position of the centroids of markers identified in the undistorted CT image, as well as in MR images, giving the x, y and z positions of each marker centroid in millimetres. The positions are given as distances from the magnetic isocentre.

In order to verify the accuracy of these centroid positions, the returned values were converted to positions within the image, where the voxel assigned as (0,0,0) is positioned at the top, left corner of the first slice of the image sequence. This transformation was performed using knowledge of the isocentre location in the image.

For one set of measurements on the Siemens Prisma and Siemens Aera scanners, a sample of image slices were selected at random, which all corresponded to the approximate centre of the markers in the z direction. The position of the centre of the visible markers in these slices was then identified and recorded using the ImageJ software (v1.49) (167, 168). The image slice being analysed was first converted to a binary image within the software. The ultimate eroded points of the Euclidian distance map (169) were then found. The Euclidian distance map finds the distance of each pixel with intensity 1 from the nearest pixel with intensity 0. The ultimate eroded points are calculated as the maxima of the distance map within each binary particle. In this way, the centre of each marker circle in the binary image was determined. By placing a cross-hair at each of these points within the binary image, the x and y co-ordinates were located. As an additional check, a cross-hair was also placed at the visual centre of the original markers (prior to the image being converted into binary) to ensure that no gross error had occurred during the image conversion.

The absolute difference between the transformed centroid positions output by the software and the in-house detected marker centres were then determined for each of the slices. It was not expected that exact agreement would be found between the two, as, due to time constraints, the image analysis found the centre of the marker on each slice, rather than the centroid of the marker sphere. However the analysed slices were the closest to the centre for the majority of the markers and therefore the difference was expected to be small and of the order of millimetres and therefore acceptable as an initial form of validation.

A CT dataset was also analysed using the above methodology.

5.4.3 Results

An overlay plot of the two marker positions for a slice approximately 10 cm from isocentre for the Prisma scanner can be seen in Figure 5-8. For this slice, no detected centre was more than 1.2 mm from the returned centroid position in the x or y directions, which was deemed acceptable.

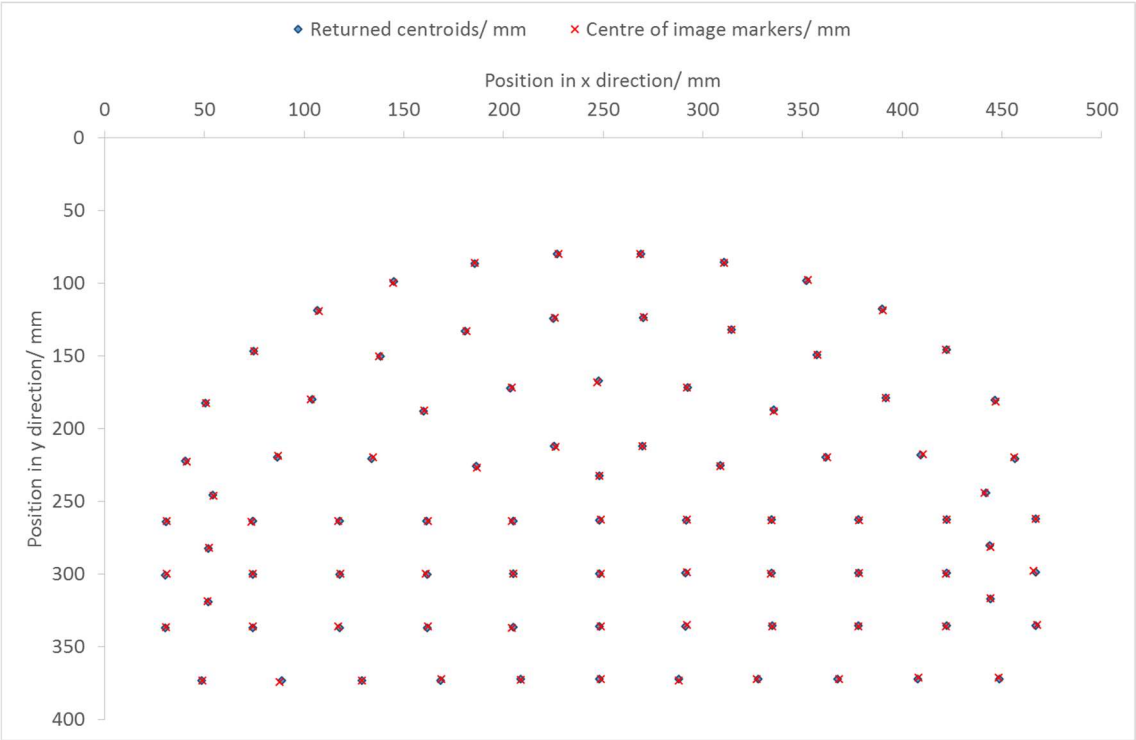

Figure 5-8 The position of the returned centroids of the markers from the Spectronic Medical AB software, converted to the image plane, and the position of the centres of the markers in the image slice which were identified in-house, for a slice positioned approximately 10 cm from isocentre on the Siemens Prisma scanner.

A slice approximately 14 cm from isocentre in the opposite z direction was also analysed. Here all marker centres were found to be within 2.1 mm of their respective recorded centroid positions in the x and y directions, which was accepted within the given uncertainties.

A slice approximately 190 mm from isocentre was analysed in a similar way. This was the furthest z slice analysed in the dataset. The x and y co-ordinates of all markers with a radial distance-to-isocentre of less than 299 mm were found to have a difference of less than 2 mm from the recorded centroid values. However some large deviations up to 19 mm in individual dimensions were recorded for some spheres at distances-to-isocentre above 299 mm. All these spheres were located in the most outer ring of the phantom. This was also checked visually and a discrepancy between the identified centroid positions and the visual centre of the sphere on the image was observed, confirming the previous findings. An

overlay plot of the two marker positions for this slice can be seen in Figure 5-9. The discrepancy between the two for the outermost ring of markers is evident.

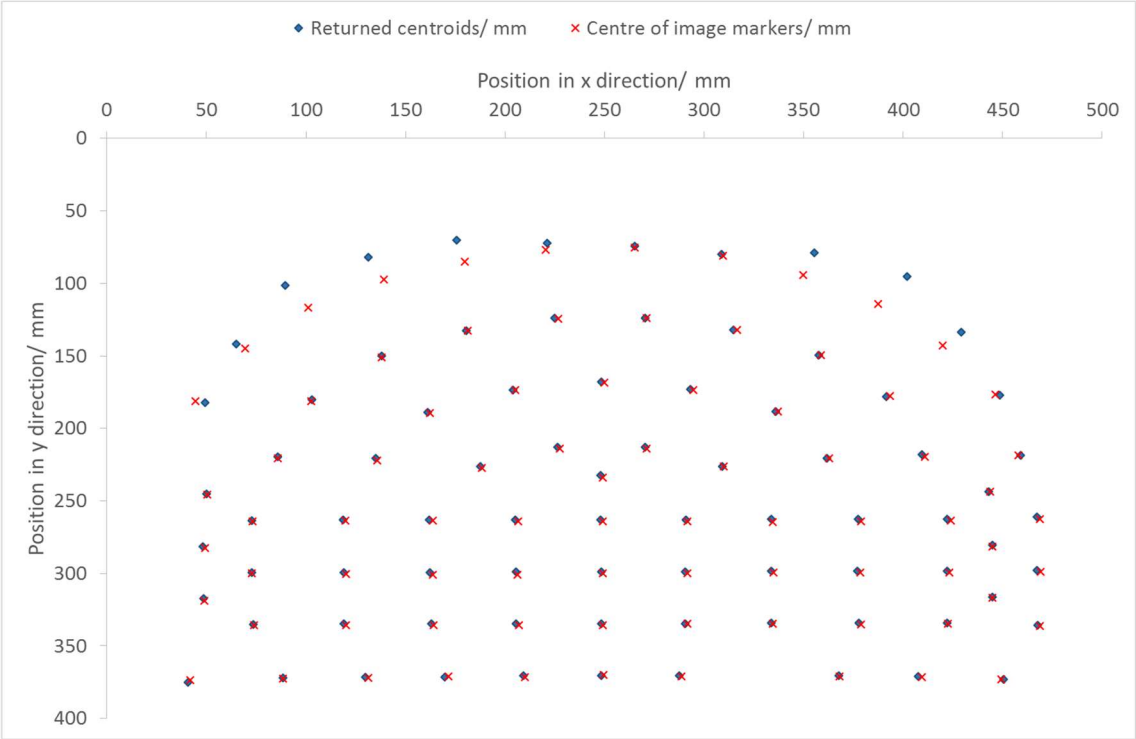

Figure 5-9 The position of the returned centroids of the markers from the Spectronic Medical AB software, converted to the image plane, and the position of the centres of the markers in the image slice which were identified in-house, for a slice positioned approximately 19 cm from isocentre on the Siemens Prisma scanner.

The investigation was also carried out for one dataset of the Siemens Aera scanner. For a slice approximately 5 cm from isocentre, the difference between the x and y co-ordinates of the identified centres and the reported centroids were within 2 mm for all markers. However, similarly to the Siemens Prisma scanner some large differences (greater than 3 mm) were observed for markers with radial distance-to-isocentre values greater than 250 mm. In some extreme cases, for the outermost ring of spheres at large distances-to-isocentre, variations of up to 36 mm in one direction were recorded.

No such discrepancies were observed when the centroids of the markers in the reference CT scan were compared to the identified centres.

5.4.4 Discussion

Whilst the position of most spheres appeared to be identified correctly, for slices further from the isocentre, some spheres at large distances-to-isocentre, located in the outermost ring of spheres, indicated a discrepancy between the identified centre of the spheres and the returned centroids. These results were discussed with Spectronic Medical AB.

The software should exclude markers where the signal to noise ratio is too low to detect the centroid of the marker accurately. However, it appears that the software failed to identify and exclude some of these markers, due to the similar signal between markers and the background towards the edges of some of the images at greater distances from isocentre.

As a result of the findings of this work, Spectronic Medical AB are currently working on modifying the software.

5.4.4.1 Implications of these findings

It has been found that spheres at large radial distances-to-isocentre (> 250 mm) can have inaccurate reported centroid values. This has been reported to Spectronic Medical AB and, as a result, the software is currently being modified to fix this issue. As only a minority of spheres are affected it is not expected that this will impact on the results reported in Section 5.2, however in order to test this, the effect of removing all markers with a distance-to-isocentre greater than 250 mm was investigated for the MRI scanners used in this study. By removing these markers, any erroneous centroid measurements in the outermost sphere of markers for slices at larger z dimensions will be removed. However this would mean that distortion over the full scanner FOV would not be assessed which would not allow MRI-only planning for some patients, for example for a large pelvis patient.

The p-values of the Wilcoxon signed rank test for the ranges and SDs of marker distortions over five months and five test-retest measurements were re-calculated

with markers with distances-to-isocentre greater than 250 mm removed. A comparison of the p-values with these markers excluded and included can be seen in Table 5-5. The results for the Siemens Espree are not present as no analysed markers are positioned at greater than 250 mm distance-to-isocentre for this scanner.

Scanner and Sequence	Wilcoxon p-value including all markers		Wilcoxon p-value for markers with DI <250 mm	
	Range	SD	Range	SD
Prisma 2D	**<0.01**	**<0.01**	**<0.01**	**<0.01**
Prisma 3D	0.52	0.71	0.35	0.43
Signa 2D	**<0.01**	**<0.01**	**<0.01**	**<0.01**
Signa 3D	0.41	0.73	**0.04**	0.25
Aera 2D	**<0.01**	**<0.01**	**<0.01**	**<0.01**
Aera 3D	**<0.01**	**<0.01**	**<0.01**	**<0.01**

Table 5-5 Comparison of the p-values of the Wilcoxon signed rank test for all markers and when only markers with distance-to-isocentre (DI) values less than 250 mm are included in the analysis. Results in bold and printed in red indicate that a significant difference was found.

It can be seen from Table 5-5 that the majority of results are unaffected by excluding these markers, with no difference in the significance of the findings. However for the 3D sequence on the GE Signa a significant difference in the range of the markers was reported, which was not the case previously. This result is unexpected, as if any erroneous centroid measurements from outer markers were removed, it would be expected that distortion variation would reduce, as well as the difference between monthly and test-retest measurements. For this sequence the difference in ranges was further investigated, but only including markers with distance-to-isocentre values of less than 200 mm. For this scenario, a p-value of 0.31 was found, in-line with the previously reported p-value for all markers, indicative of no significant difference. Therefore it appears likely that this difference was not caused by erroneous markers, but by a true larger variation of marker distortions with distances-to-isocentre between 200 and 250 mm.

Based on these results, it appears that the potential inclusion of any erroneous centroid positions from the outermost sphere of markers would be unlikely to influence the results of the temporal stability analysis in Section 5.2. Indeed the results in Table 5-5 show that even when all markers with distance-to-isocentre values greater than 250 mm are removed, there is still a variation in the temporal stability of distortion between scanners. However in order for the technique to be clinically implemented, it is necessary for either the outermost sphere of marker distortions to be analysed, or for the clinical FOV to be limited to within a radial distance-to-isocentre of 250 mm, which would be acceptable for brain patients.

5.5 Concluding remarks

This study utilised the GRADE distortion phantom. This is a lightweight, large FOV phantom with internal markers covering the full FOV of the four MRI scanners used in this investigation. As requested for the study, the phantom was designed with a curved, rather than straight, anterior edge to ensure that the full MRI FOV was sampled. The magnetic susceptibility of this phantom was found to be negligible at radial distances-to-isocentre of less than 240 mm. In this study, an error was identified in the Spectronic Medical software regarding the analysis of the position of the outermost distortion markers, however this was not found to influence the overall results of the temporal stability distortion analysis.

This chapter has shown that it is difficult to provide a blanket recommendation on the required frequency of QA of MRI geometric distortions for an MRI-only radiotherapy workflow. This is due to variations of the temporal stability of the MRI geometric distortions between different scanners and sequences. It is therefore necessary prior to implementing an MRI-only radiotherapy pathway to determine the temporal stability of the MRI geometric distortions on individual MRI scanners for clinical MRI sequences.

Chapter 6 MRI and CT compatible head and neck phantom

6.1 Introduction

In order to clinically implement MRI-only radiotherapy, end-to-end testing of the workflow should be performed, in addition to the MRI distortion QA discussed in Chapter 5. End-to-end testing of the pathway is commonly used in radiotherapy to ensure that the entire workflow is operating as expected and it should be carried out whenever a new technique is brought into clinical practice to ensure the fidelity of the overall system delivery (55). This is normally performed using a phantom which represents the patient and which can be used for both imaging and dosimetry. For testing an MRI-only radiotherapy workflow, the phantom would need to appear geometrically similar to the patient anatomy, and representative of a patient when scanned using both CT and MRI modalities. The latter should be true when using clinically relevant protocols and sequences.

The aim of this study was to design an anthropomorphic phantom which could be used for the end-to-end testing of an MRI-only radiotherapy workflow for brain patients. In order to ensure that the phantom would be as useful as possible, the phantom was also designed to be suitable for testing a head and neck MRI-only workflow.

To construct the anthropomorphic phantom this project was carried out in collaboration with Leeds Test Objects (LTO) Ltd. (Boroughbridge, United Kingdom). The company agreed to construct a phantom which met the specifications and designs which I created. Prior to meeting with the company, I produced the phantom specifications, initial phantom designs and a review of potential phantom materials. Once materials had been manufactured by LTO, I tested these using clinical MRI and CT scanners, and analysed the results in-house in order to guide the final selection of phantom materials. I created the final anthropomorphic phantom design by using tissue segmentations produced in-house using a patient CT scan.

6.2 Initial phantom design

6.2.1 Phantom specification

The primary purpose of the phantom was the end-to-end testing of an MRI-only radiotherapy workflow. The phantom was designed as a head and neck anthropomorphic phantom, which would allow the workflow to be tested for head and neck, as well as brain, patients. In order for the phantom to be suitable for this purpose, the head, neck and also the superior section of the thorax including the top of the lungs was included in the phantom design.

To perform end-to-end testing of the entire workflow, the phantom needed to appear anatomically correct when using MRI imaging. These images would be used for creating a sCT of the phantom, which would be used as a reference image for radiotherapy planning and dose calculations. The actual delivered dose from the radiotherapy plan would be measured at positions within the phantom, to ensure the accuracy of the planned dose.

The phantom would also ideally be scanned using cone-beam CT (CBCT) imaging. This means that the phantom should also appear anatomically accurate when imaged with CT. The shifts needed to match the CBCT image of the phantom in its position on the linac with the reference MRI or sCT scan of the phantom would be compared to the results when using the CT as the reference image. The phantom would either be set up accurately or with introduced known errors to simulate problems in the patient set-up.

Additionally this phantom could also be used to compare dose distributions when radiotherapy planning is performed on the sCT and CT of the phantom. These could also be compared to the actual delivered dose. The geometrical difference between the CT and sCT of the phantom could be determined by registering the two and comparing the HU intensity of the two image sets. These tests would also form part of the MRI-only radiotherapy QA programme, although not technically forming part of an end-to-end QA test. These tests require the

phantom to be compatible with CT imaging, adding another incentive for the phantom materials to appear similar to tissues on CT as well as MR images.

To determine the necessary compartments of the phantom, head and neck anatomy was studied using CT and MRI patient scans. The phantom must appear sufficiently similar to a patient that a sCT could be generated by imaging the phantom. It is especially pertinent that high density bone and air cavities are accurately modelled. The technical ability to manufacture small and intricate sections of the head and neck was discussed with LTO during the phantom design phase.

In order to accurately represent the head section of a patient the phantom should contain surrogates for the skull, brain tissue, teeth, as well as air cavities. Using a single brain tissue is an approximation of a human brain which contains both grey and white matter. However in order to reduce the complexity of the prototype a single material will be used to represent the brain. This is not expected to affect the quality of the sCT, however if it is found that this in unacceptable, a more complex brain structure will be considered in future phantom versions.

Some patients who are treated at LCC have dental fillings which induce artefacts in MRI and CT imaging. Therefore it would be useful to a have a dental filling present in the phantom to represent these patients. Ideally this would be removable in order to also test the workflow for patients without dental filling artefacts.

For the head and neck region the spinal vertebrae should be modelled. The bony structure should also include the shoulders. This is important for the assessment of magnetic susceptibility artefacts (see Section 1.4.2) during MRI imaging of head and neck patients. Studying these transitions in tissue densities inside a phantom should be an aspect of performing QA of the sCT generation pathway for head and neck patients.

Another type of artefact which should be assessed during sCT generation QA is the chemical shift artefact. This typically arises in the human body due to the chemical shift between water and fat protons, as previously discussed in Section 1.4.2. Therefore ideally this would be modelled in the phantom by including a volume of an adipose tissue surrogate surrounded by the water equivalent tissue.

In order to assess the dosimetry of the radiotherapy plans, it would be useful to measure the actual dose within the patient. One way of achieving this would be to create spaces in the phantom for dosimetric measurement chambers. Space for placing films for measuring planar dosimetry would also be useful. Decisions regarding the type of chambers to use were made after discussions with medical physics experts.

There are also some general requirements that the phantom should fulfil. Firstly it should be ensured that MRI receive coils can be placed as close to the phantom surface as possible. This would aid in achieving a high signal to noise ratio in MR images, and mimic the clinical set-up. To do this, no filling caps should be placed where a coil may be positioned. For a reproducible position to be maintained when imaging, isocentre markers should be present on the phantom. Ideally, there should be two sets; one for a brain, and one for a head and neck set-up. This again makes the scanning of the phantom more similar to scanning a patient. Additionally, the phantom should possess the ability to lie in a stable position on a flat top couch. This again allows a reproducible phantom set-up to be achieved.

Initial designs of the ideal phantom were produced. The resulting schematic design can be seen in Figure 6-1, Figure 6-2 and Figure 6-3 in three planes. The head includes brain material surrounded by the skull. Air cavities are present in the head and neck, and the spinal column is modelled. Teeth are included within the head phantom and there is also a removable dental filling. Shoulder bones and the superior part of the lungs are also included. A spherical volume containing adipose surrogate is present. There is room for two dosimetric chambers, one in the superior section of the phantom, and one in the inferior. These should fit semi-flex dosimetric chambers which are commonly used in

radiotherapy departments. The size of these chambers was found using manufacturers specifications, and was confirmed with physical measurements of the chambers. Film measurements can also be performed within a cross-section of the brain. The phantom is contained in a skin surrogate representing the dermis. All remaining areas inside the phantom would be filled with a soft tissue equivalent material. In this design, the shoulder and neck sections are separate from the head section of the phantom. The purpose of this would be to not have to use the bottom section if only the brain workflow were being tested.

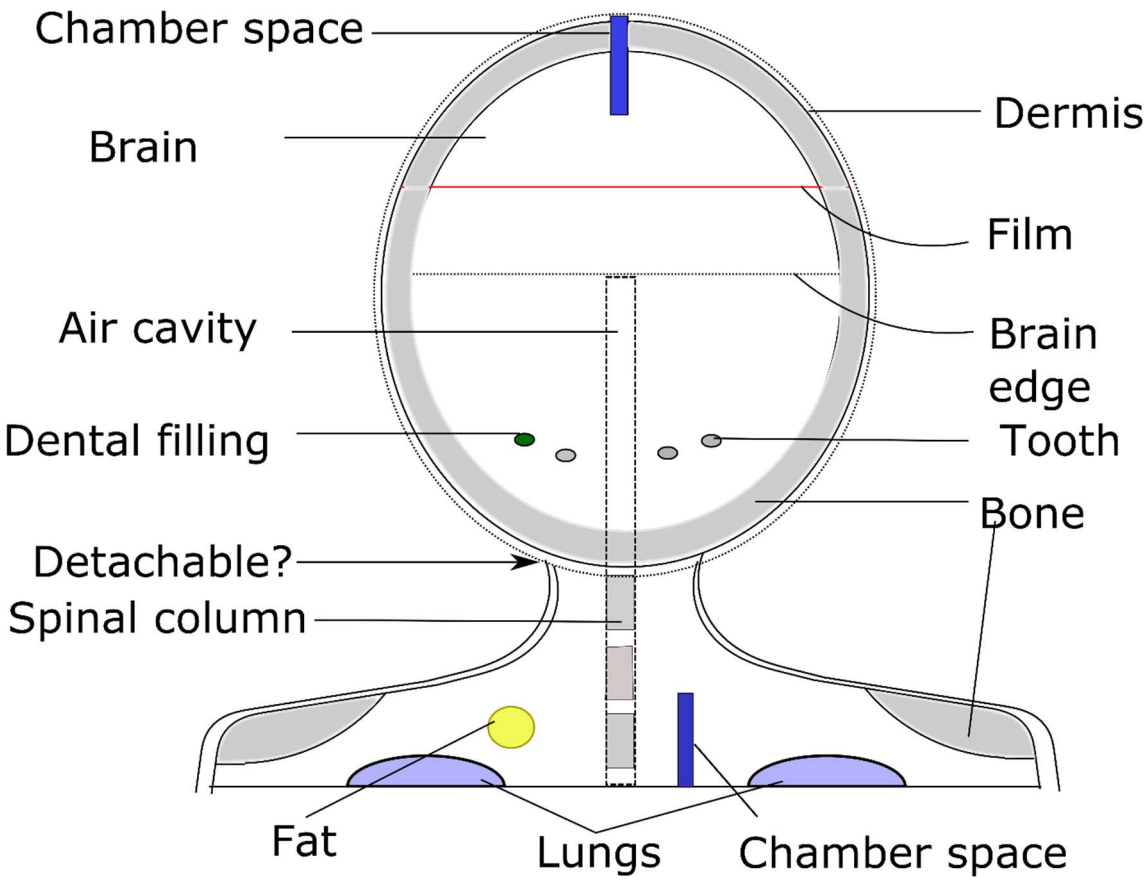

Figure 6-1 Coronal cross-sectional image of the initial phantom design. This schematic diagram is for illustrative purposes only in order to show the components which should be present in the phantom. The image is not to scale. The position of the air cavity is indicated by dotted lines, however it would not be present in this slice. The remainder of the phantom (unlabelled sections) would contain soft tissue equivalent material.

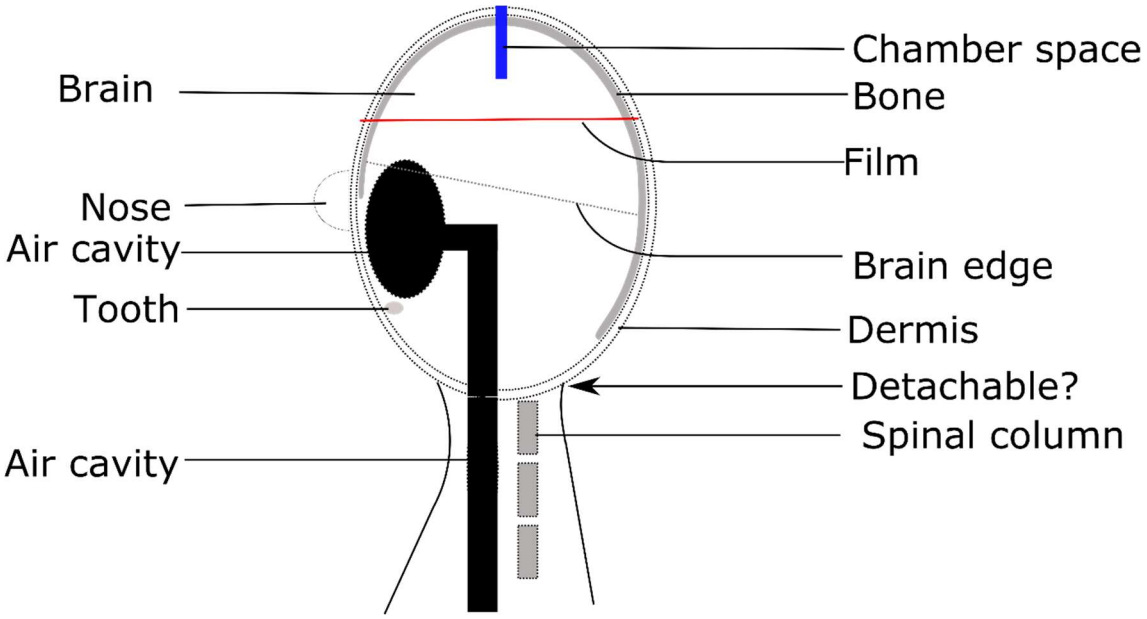

Figure 6-2 Sagittal cross-sectional image of the initial phantom design. This schematic diagram is for illustrative purposes only in order to show the components which should be present in the phantom. The image is not to scale. The remainder of the phantom (unlabelled sections) would contain soft tissue equivalent material.

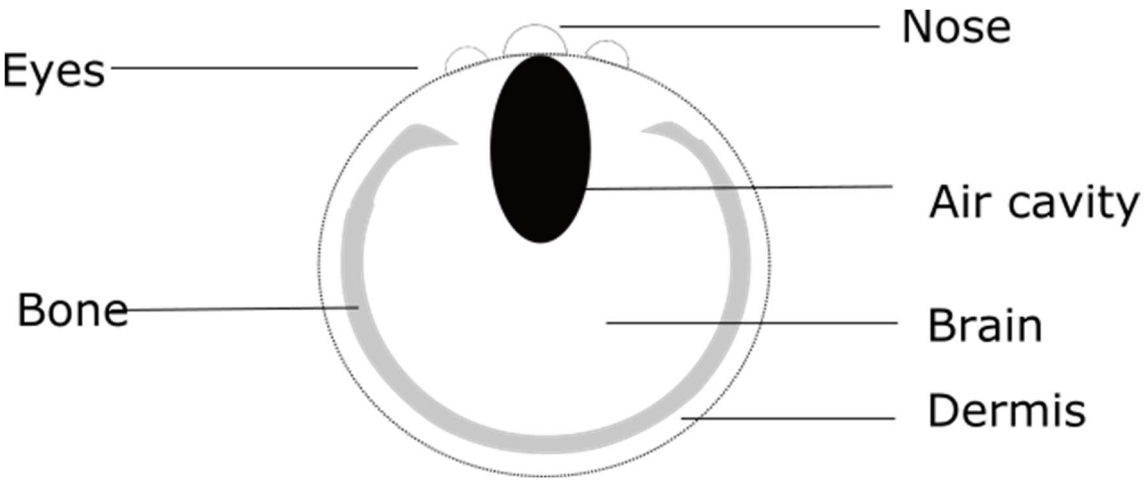

Figure 6-3 Axial cross-sectional image of the initial phantom design. This schematic diagram is for illustrative purposes only in order to show the components which should be present in the phantom. The image is not to scale. The remainder of the phantom (unlabelled sections) would contain soft tissue equivalent material.

6.2.2 Identification of potential tissue surrogate materials

The phantom should be constructed with tissue surrogates which have similar properties to those of the tissues of interest on both CT and T_1 and T_2-weighted MRI images. In order to determine which materials should be used a literature review was performed to identify potential phantom materials. These were then sourced and measurements were carried out to determine the suitability of the materials to act as a surrogate for each tissue.

6.2.2.1 Bone

It has been suggested (56) that gypsum (calcium sulfate dihydrate) bandages may be used as a substitute for the high attenuation cortical bone. Gypsum has been reported to have CT numbers of up to 1200 HU at 120 kV and also has an absence of signal in MR imaging, similar to cortical bone. Niebuhr et al., (56) produced a 3D-printed hollow bone case where the number of layers of gypsum applied was varied to create the desired CT number distribution. The gypsum was found to require the addition of a clear lacquer in order to avoid water absorption.

The inner section of bone is highly inhomogeneous due to a mixture of tissue types (56). For older patients, signal intensity in the MRI is dominated by fatty compositions. For this reason, Niebuhr et al., (56) chose Vaseline (petroleum jelly) as a base for this part of the bone in a pelvis phantom, as it has similar CT and MRI properties to adipose. Dipotassium phosphate (K_2HPO_4) was then added to simulate the bone. Brain and head and neck cancer patients have a different age demographic to pelvis patients and therefore this mixture may not be appropriate for this study, as brain cancers, for example, occur relatively frequently at younger ages (170).

Dipotassium phosphate diluted in water was used in another study to simulate cranial bone (171). The concentration of the solution can be adjusted to match bone electron densities in the CT images. However it was found to be an

inadequate surrogate for bone in MR imaging, due to the resulting high signal intensity. Gypsum or epoxy resin were suggested instead.

There is also the potential for the bone to be represented by a ceramic material. This has the advantage of being able to be 3D printed. As a solid material this should have similar MRI properties to bone with both having an extremely short T_2. The CT properties of the ceramic and the similarity to bone would have to be assessed if it were to be used as a surrogate.

6.2.2.2 Brain

Gallas et al., (171) used a gel containing agarose powder and water to simulate brain tissue. Variation of the concentration of agarose in the gel enabled differing T_1 and T_2 contrasts. Agarose concentration however had little effect on the CT number of the gel (agarose is a tissue equivalent composition meaning it may be suitable to act as a brain surrogate). The authors in this study found a 5% concentration of agarose in distilled water was a good brain tissue surrogate. The longevity of an agarose solution may be a concern (57). It is possible to add a preservative in order to extend its lifetime, however the solution would still have to be replaced regularly.

LTO have used polyvinyl alcohol cryogel (PVA-c) previously in tissue-equivalent ultrasound phantoms. A cryogel is formed by freezing and thawing an aqueous solution of PVA (58), resulting in chemical changes. The tissue-equivalence of this material in ultrasound and MRI is well-established in the literature, and it has been used to create homogeneous brain phantoms for imaging with these modalities (59). It has been reported that the number of freeze-thaw cycles performed during the manufacture of the PVA-c affects the T_1 and T_2 properties of the material (172), for example Surry et al., (59) reported T_1 and T_2 values in the range of 718 - 1034 ms and 108 – 175 ms respectively for a 10% aqueous PVA-c solution over 1 to 4 freeze-thaw cycles at 1.5 T. The similarity of the CT properties of PVA-c aqueous solution to brain tissue would have to be confirmed, however as it is primarily water-based it is likely to be approximately equivalent.

6.2.2.3 Adipose

Vegetable oil has been suggested as a suitable surrogate for adipose tissue for both MRI and CT imaging (56). If this was placed within a water filled cavity, chemical shift artefact could be investigated, as discussed in Section 6.2.1.

6.2.2.4 Dental Filling

The most common dental fillings are made of amalgam (173), a mixture of mercury, silver, tin, copper and zinc. Mail et al., (174) combined a PTW water phantom with a slab of high density dental filling material (containing 50% mercury, 25% silver, 14% tin, 8% copper and 3% trace metals) in order to investigate the impact of dental fillings on treatment planning and dose delivery. There is the potential to use this mixture in the phantom as a dental filling.

6.2.2.5 Teeth

The literature search did not identify any currently used normal dental surrogates for CT or MRI imaging. On inspection of CT and MRI images of patients, it appears that the CT and MRI numbers of the teeth appear similar to the cortical bone in the skull. Therefore gypsum or a ceramic material may be suitable.

6.2.2.6 Spine

Cerebrospinal fluid in the ventricles is similar to water (171). Gallas et al., (171) therefore used distilled water as a surrogate for this tissue. The remainder of the spinal cord is made of white and grey matter and therefore an agarose solution or PVA-c gel may be a suitable surrogate here. Gypsum bandage or ceramic may be suitable for mimicking the bony part of the spinal vertebrae.

6.2.2.7 Lung

The literature review did not identify any CT and MRI lung surrogate materials. As a starting point the lungs could be simulated as air cavities.

6.2.2.8 Phantom casing

The material for this depends to an extent on the way the phantom is manufactured. Epoxy resin (solid water) may be a suitable outer casing material (171). This is used in many radiotherapy phantoms as it exhibits similar dosimetric properties to water in megavoltage photon beams (175). MRI and CT images of the material would need to be taken to assess its suitability as an outer casing material and to ensure that it is not significantly different from the skin of a patient.

LTO use a photopolymer, which has similar properties to thermoplastic polymers, i.e. the material is strong, durable and water resistant, for 3D printing. These properties are ideal for the casing of the phantom as it will not need to be replaced and it will be resistant to the inner water-equivalent materials.

6.3 Evaluation of tissue surrogate materials

MRI and CT properties of PVA-c, vegetable oil, solid gypsum, gypsum powder and a ceramic material were acquired in order to investigate their suitability as tissue surrogate materials.

6.3.1 Measurements

6.3.1.1 Brain surrogate material

Different aqueous solutions of PVA-c were prepared by LTO. A concentration of 10% was tested based on the study by Surry et al., (59) where this concentration had been used to create a simple MRI and ultrasound tissue-equivalent brain phantom. A 5% solution was also prepared for comparison. Each of these solutions were prepared using 2, 3, 4 and 5 freeze-thaw cycles, so that in total 8 PVA-c solutions were prepared in separate cylindrical containers for assessment.

In order to assess the brain tissue equivalence of the solutions, MRI scans were acquired on a Siemens Prisma 3 T MRI scanner. In the MRI scanning session

the solutions were scanned in a 20 element head and neck coil, with immobilisation pads used to reduce the chance of potential vibration. Using the head coil meant that a high signal to noise ratio was achieved due to its proximity to the solutions.

T_1 and T_2 mapping were performed in order to quantitatively assess the similarity of the solutions to brain tissues. T_1 mapping was carried out using a method originally developed for mapping the relaxation times of the myocardium, specifically the Modified Look-Locker Imaging (MOLLI) sequence (176), which is an inversion recovery method. This method allows the T_1 to be mapped rapidly and the maps are generated automatically by the scanner software. T_2 mapping was performed using a Fast Low Angle Shot (FLASH) sequence with a T_2 magnetisation preparation module, a sequence also developed for mapping the relaxation times of the myocardium. This module involves an initial 90º pulse followed by a series of 180º pulses, before a -90º pulse, tipping the magnetisation back to the z-axis, but with reduced longitudinal magnetisation. Varying the duration of this preparation delay (in this case 0, 30 and 55 ms) and fitting to an exponential decay signal model allowed determination of the T_2 relaxation time.

Additionally, a single slice T_1-weighted spin echo sequence, a single slice T_2-weighted turbo spin echo sequence and a multi-slice FLAIR sequence were also acquired. These sequences are clinically used for brain imaging at LCC and therefore the purpose of imaging with these sequences was to visually assess the appearance of the solutions when scanned with typical clinical MRI sequences.

Regions of interest (ROIs) were drawn within each of the imaged solutions in the T_1 and T_2 maps and the mean and the standard deviation of the values found. These were compared to the T_1 and T_2 values of white and grey matter at 3 T from literature in order to determine which solution was the best match and suitable to be used as a brain tissue equivalent material.

CT scans of the solutions were acquired on a Siemens Sensation scanner (120 kVp, 138 mA), which is routinely used for scanning radiotherapy patients at LCC. The solutions were placed within a water tank to simulate scanning within a patient. The mean HU within ROIs for each of the samples was determined. For three typical radiotherapy brain patients, 10 ROIs within the brain tissue were also created (30 ROIs in total) and the mean HU value within each found. The HUs of the samples were then compared to those within the brain CT scans.

6.3.1.2 Bone surrogate material

CT and MRI scanning was also performed of the gypsum powder, solid gypsum, and a ceramic material provided by LTO with a stated density of 1.75 gcm^{-3}. Two forms of gypsum were used. Firstly in a high purity form which is typically used to produce hard casts requiring fine detail reproduction (177). The second was typically used as dental plaster for the production of dentures (178). Solid (manufactured in-house) and powdered forms of each were produced and tested. In the MRI scanning session, only the MRI imaging sequences, rather than the mapping sequences, were carried out for these materials, to confirm the lack of signal on conventional MRI sequences. For comparison of the CT images, 10 ROIs were drawn within the skull of 3 patients and for each the mean HU value was recorded.

6.3.1.3 Adipose surrogate material

CT scanning was performed of vegetable oil and Vaseline. MRI scanning of these materials were not carried out in this investigation due to constraints on the MRI scanner. The MRI properties of these materials will be assessed in future analysis. ROIs were drawn in the CT images, and the mean HU was compared to that of adipose, as stated in the literature.

6.3.2 Results

6.3.2.1 Brain surrogate material

Table 6-1 shows the mean T_1, T_2 and HU numbers for each aqueous solution of PVA-c.

Concentration of PVA/ %	Number of freeze-thaw cycles	Mean T_1/ ms	Mean T_2/ ms	Mean HU number/ HU
5	**2**	1743 ± 16	149 ± 6	12.0 ± 13.2
5	**3**	1748 ± 32	153 ± 9	9.3 ± 11.9
5	**4**	1651 ± 62	142 ± 5	5.0 ± 11.9
5	**5**	1575 ± 47	128 ± 11	11.4 ± 10.9
10	**2**	1253 ± 15	90 ± 4	25.0 ± 11.3
10	**3**	1156 ± 24	85 ± 3	19.9 ± 11.3
10	**4**	1090 ± 20	77 ± 2	30.2 ± 10.7
10	**5**	1056 ± 17	80 ± 6	29.7 ± 12.1

Table 6-1 The mean T1, T2 and Hounsfield unit number of each of the aqueous solutions of PVA-c. The values for 5% and 10% concentrations of the solutions for freeze-thaw cycles of 2 - 5 are shown. The errors represent the standard deviation within the region of interest.

The mean HU values for white and grey matter from the measurements of patient CT scans can be seen in Table 6-2.

	Patient 1	Patient 2	Patient 3	Mean
Grey matter	32.7 ± 1.2	36.4 ± 2.4	36.1 ± 1.8	35.1 ± 1.1
White matter	11.2 ± 0.3	11.1 ± 1.5	13.3 ± 2.3	11.8 ± 1.0

Table 6-2 The mean Hounsfield units of white and grey matter for three typical brain patients. The error represents the standard error of the mean.

6.3.2.2 Bone surrogate material

The mean HU numbers of gypsum powder and solid gypsum for both forms of gypsum, as well as that of the ceramic, can be seen in Table 6-3. None of the materials were found to have signal using conventional MRI sequences.

Material	Mean HU value/ HU
Ceramic	973.5 ± 10.2
High purity gypsum powder	785.7 ± 77.4
Gypsum powder	533.3 ± 13.9
High purity solid gypsum	1110.6 ± 30.4
Solid gypsum	625.2 ± 5.5

Table 6-3 The mean Hounsfield unit value for each of the bone surrogate materials imaged. The error represents the standard error of the mean.

The mean HU value within the patient skull measurements was found to be 1107.8 ± 75.9 HU. The range of measurements i.e. the minimum and maximum HU values over all ten ROIs was 456 HU to 1641 HU.

6.3.2.3 Adipose surrogate material

The mean HU of the vegetable oil and Vaseline were -133.1 ± 0.6 HU and 2.3 ± 10.6 HU respectively.

6.3.3 Selection of materials

6.3.3.1 Brain surrogate material

T_1 values of white and grey matter typically vary between 700 and 1000 ms and 1400 and 1700 ms respectively at 3 T (179). The T_2 values are approximately equal to 70 ms and 100 ms for white and grey matter respectively at 3 T (179). Using the results from Table 6-1 it appeared that the 10% solution was the closest match with T_1 and T_2 values being between those for grey and white matter. The HU values of the 10% solutions, particularly at the higher number of freeze-thaw cycles, are similar to those observed for grey matter which makes up the main component of the brain. It was decided to proceed with using the 10% solution,

with four freeze-thaw cycles to produce the brain phantom. This solution was the closest to the HU value of the grey matter and all MRI parameters were in-between the grey and white matter values.

A larger volume of this solution was prepared to ensure that a homogeneous material could be produced. A CT slice through this volume can be seen Figure 6-4. The volume appeared homogeneous throughout. The mean HU was found to be 22.4 ± 1.0 HU which was acceptable as it measured between the white and grey matter HU values.

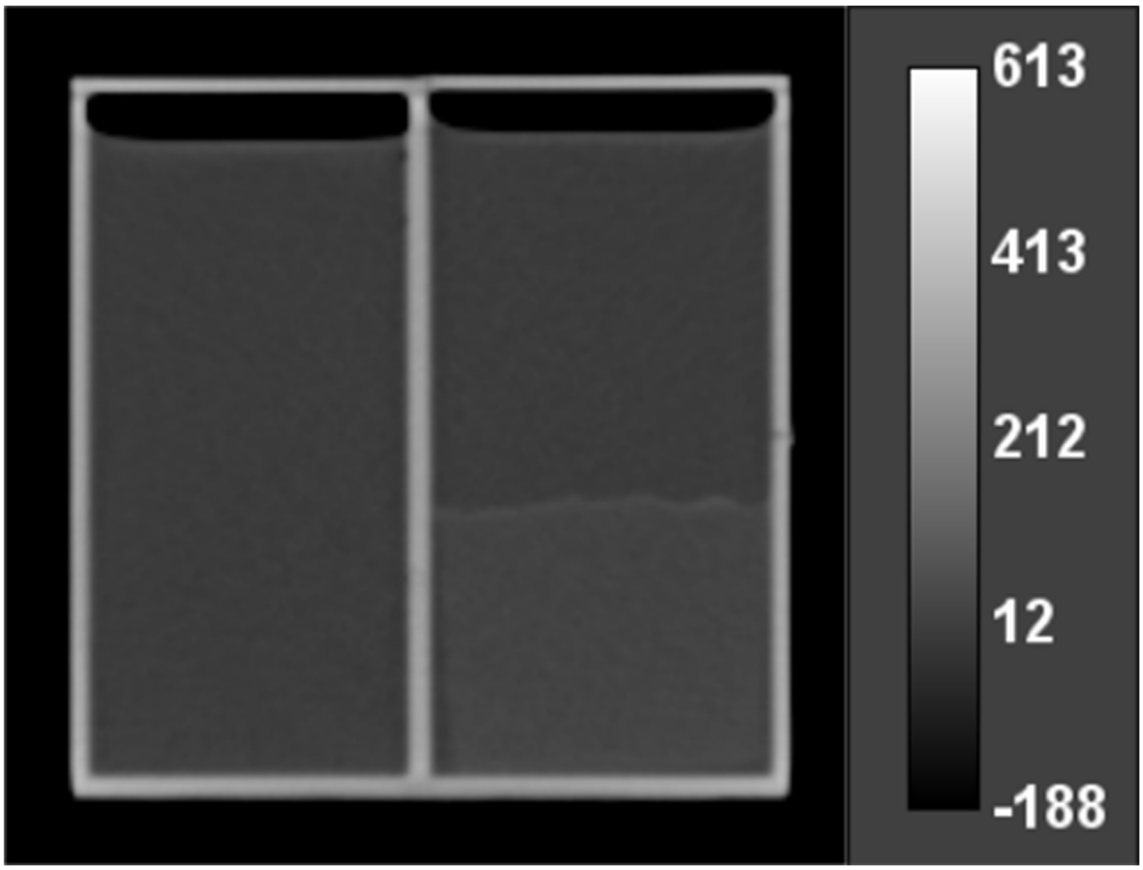

Figure 6-4 CT scan of the PVA-c 10% aqueous solution, produced using four freeze-thaw cycles. The PVA-c is positioned in the section in the bottom right of the phantom. The remaining solution in the phantom is water. The scale in HU is shown to the right of the image.

6.3.3.1.1 Update to the selection process

Between these measurements and the phantom construction, the freezing process carried out by LTO was modified as a new freezer was used and

therefore the measurements were repeated. The PVA-c (10% solution, 4 freeze-thaw cycles) was formed into a brain shape using a mould. MRI and CT measurements of this brain phantom were carried out before the mould was fitted in the head and neck phantom. Again the T_1, T_2 and HU values of the phantom were measured. These were compared to the previous findings, as well as to the white and grey matter values. MRI and CT images were also acquired and checked for uniformity.

The mean T_1 and T_2 recorded within the brain phantom were 1395 ms and 106 ms respectively. These were both longer than previously measured, most likely due to differing amounts of water in the phantom, due to the difference in the freezing process caused by using a different freezer. This is also supported by an observed difference in the mean HU number of the phantom which was found to be equal to 18.1 HU. However the mean T_1 and T_2 values were similar to that of grey matter, and the HU value was between those of white and grey matter and therefore the properties were accepted. A T_2 weighted image of the brain phantom can be seen Figure 6-5. This exhibited good homogeneity, as did the CT image of the phantom. The chamber space can be seen in the T_2 image. This can be filled with a PVA-c plug.

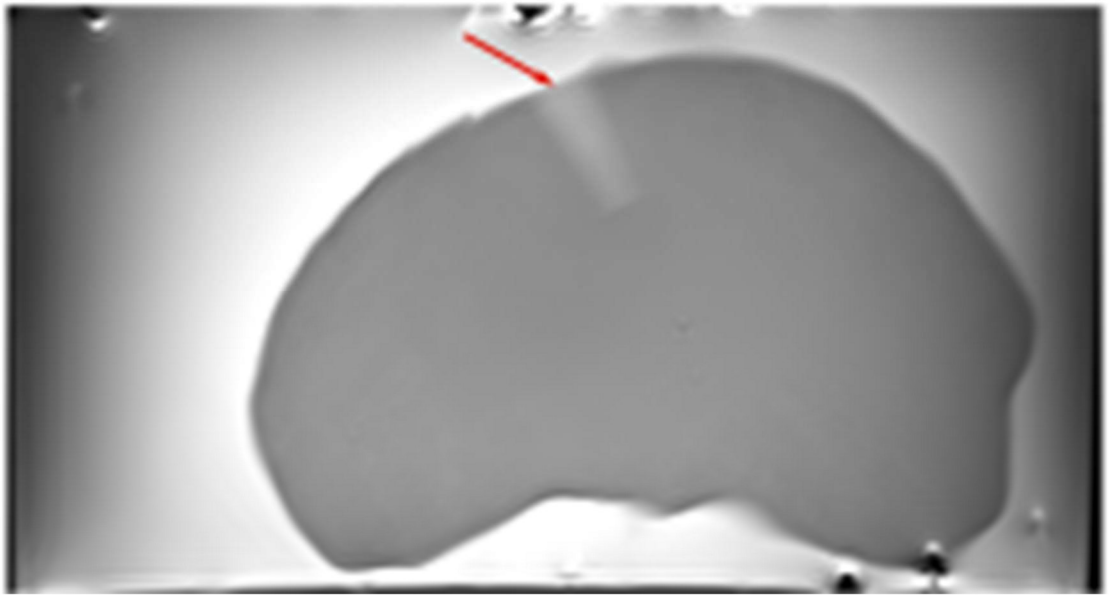

Figure 6-5 A single shot T_2-weighted turbo spin echo image of the brain phantom. A red arrow demonstrates the presence of a dosimetric chamber cavity. The brain phantom is surrounded by ionised water within a plastic casing.

6.3.3.2 Bone surrogate material

Compared to the patient reference values, the gypsum powders appear to have HU values which are too low (see Table 6-3). The high purity gypsum powder was also found to be highly inhomogeneous, as indicated by the observed variation in HU value. Layers were observed in this material and therefore it is unsuitable for inclusion in the phantom. The solid gypsum in its low purity form also had too low a mean HU value to be used as a bone surrogate. Both the high purity solid gypsum and the ceramic appear to have mean HU values close to the mid-range of the reference skull HU values. Due to the ease of manufacturing into a skull with a 3D printer, as well as the latter having no issues with water absorption, the ceramic was used in the manufacturing process for the skull.

LTO produced a skull prototype using a standard design using this ceramic. Acquired coronal CT slices through this phantom can be seen in Figure 6-6. This figure demonstrates the complexity that can be achieved when this material is used for printing.

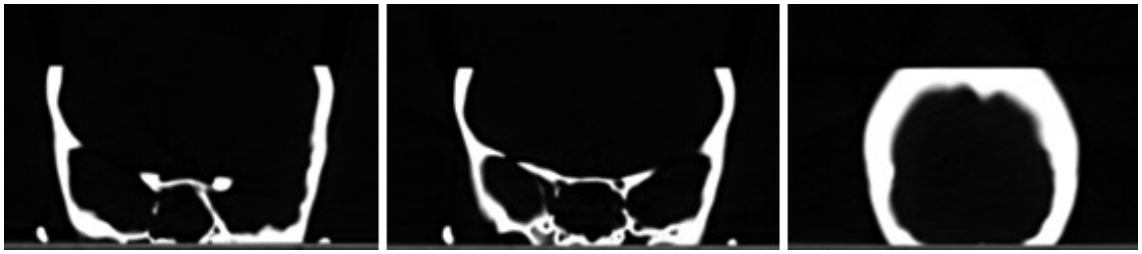

Figure 6-6 CT slices through a prototype skull phantom made from ceramic.

6.3.3.3 Adipose surrogate material

The HU range of human adipose is known to be between -150 and -50 HU (180). The vegetable oil appears to be a suitable surrogate with a mean HU value of -133.1 HU. Vaseline appears to be a poor substitute as a sole material, as the mean HU is too high at 2.3 HU.

6.4 Creation of the phantom structure

In order to perform end-to-end testing of an MRI-only radiotherapy pathway, the phantom must appear sufficiently anatomically similar to a real patient head and neck. An anonymised CT scan of a typical head and neck patient treated at LCC was selected. The bone, air cavities, including the lungs, eyes, brain, brainstem and spinal cord, as well as the patient contour were delineated in the Oncentra Masterplan treatment planning system. For the bone contouring it was ensured that cartilage and contrast in the CT image were not included in the delineation. Care was taken when delineating the patient contour so that no immobilisation devices, such as the mask, were included in the delineation. Apart from the body contour, it was checked that there was no overlap between any of the contours. The structure file and CT image were then exported from Oncentra Masterplan.

The radiotherapy structure files were then converted to stereolithography (STL) files to be in line with the phantom manufacturing process requirements. This was performed by firstly converting the exported radiotherapy structure file to separate nifti files for each structure using the open source software Smilx (Australian e-Health Research Centre, CSIRO, Australia, ITK version 4.7.0). These files were

then converted to individual surface meshes in an STL format using the ITK-SNAP software (version 3.6.0, www.itksnap.org) (181). The 3D surface meshes for each delineated structure can be seen in Figure 6-7, Figure 6-8, Figure 6-9, Figure 6-10, Figure 6-11, Figure 6-12 and Figure 6-13. The STL files were then shared with LTO, along with the original CT scan, in order for the phantom to be constructed using these structures as a guide.

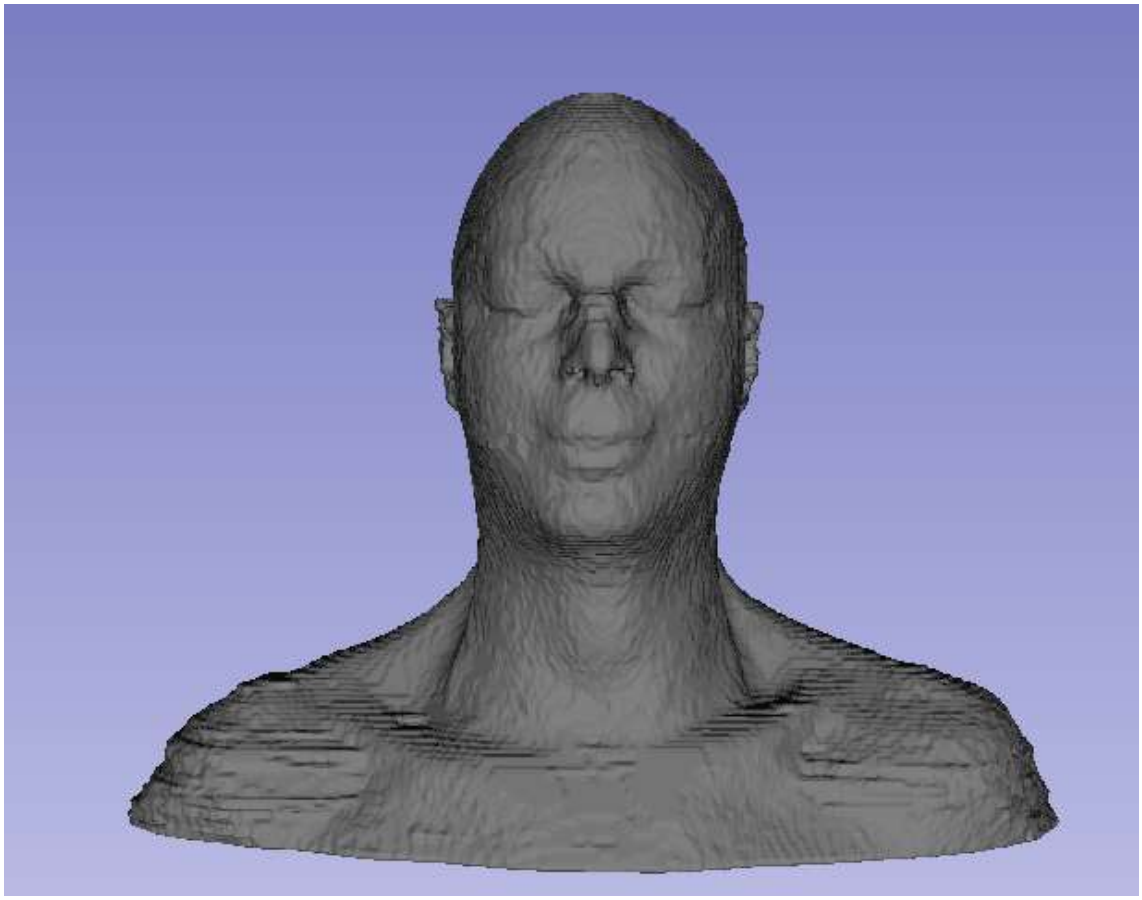

Figure 6-7 Image of the 3D structure mesh which was created as a basis for the external contour in the phantom.

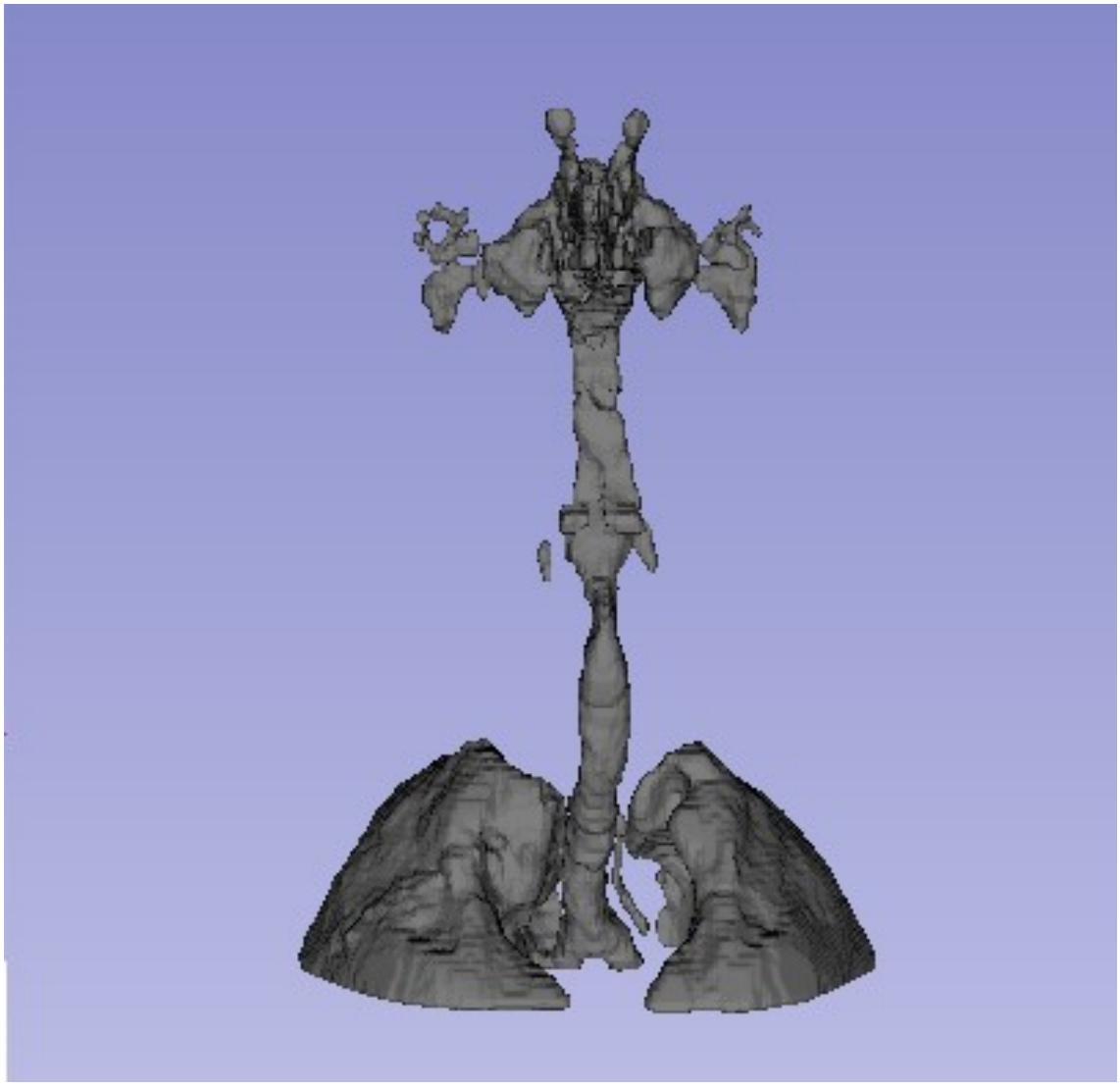

Figure 6-8 Image of the 3D structure mesh which was created as a basis for the air and lungs in the phantom.

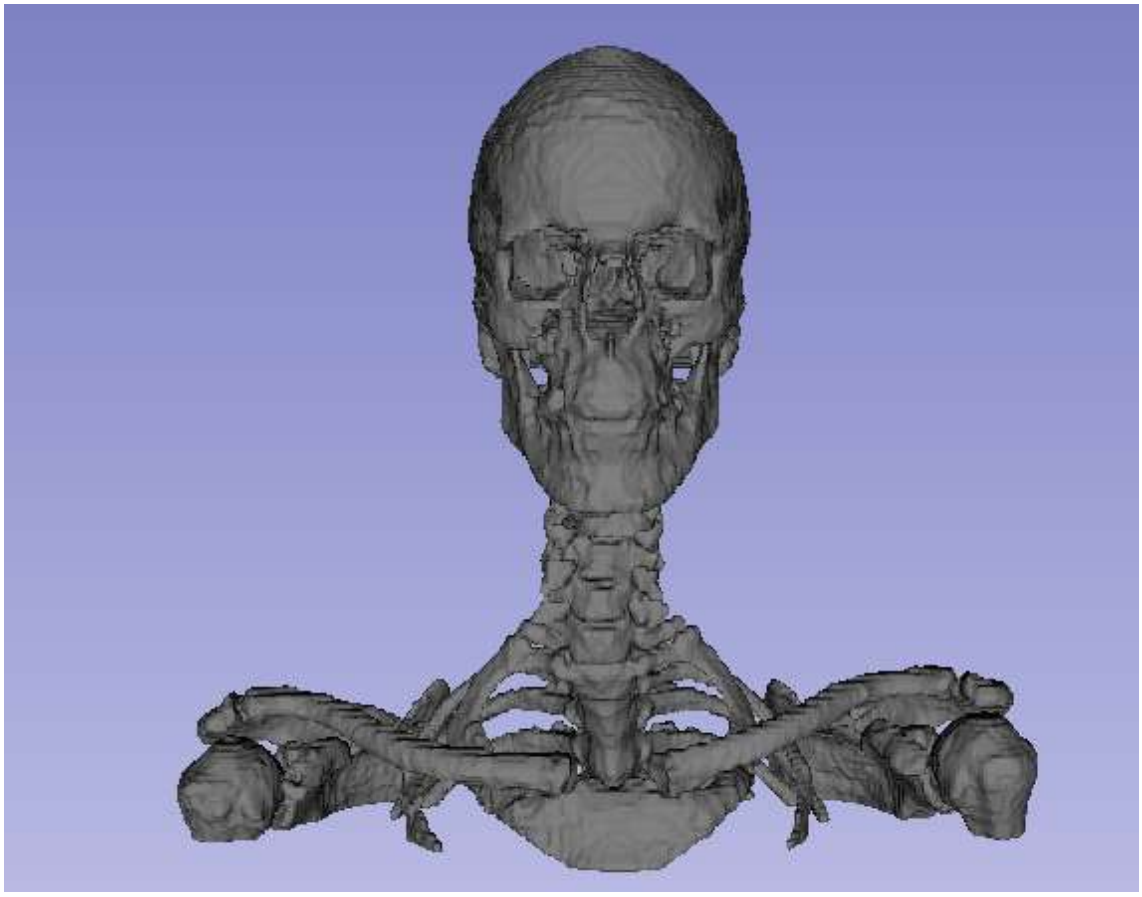

Figure 6-9 Image of the 3D structure mesh which was created as a basis for the bone in the phantom (coronal view).

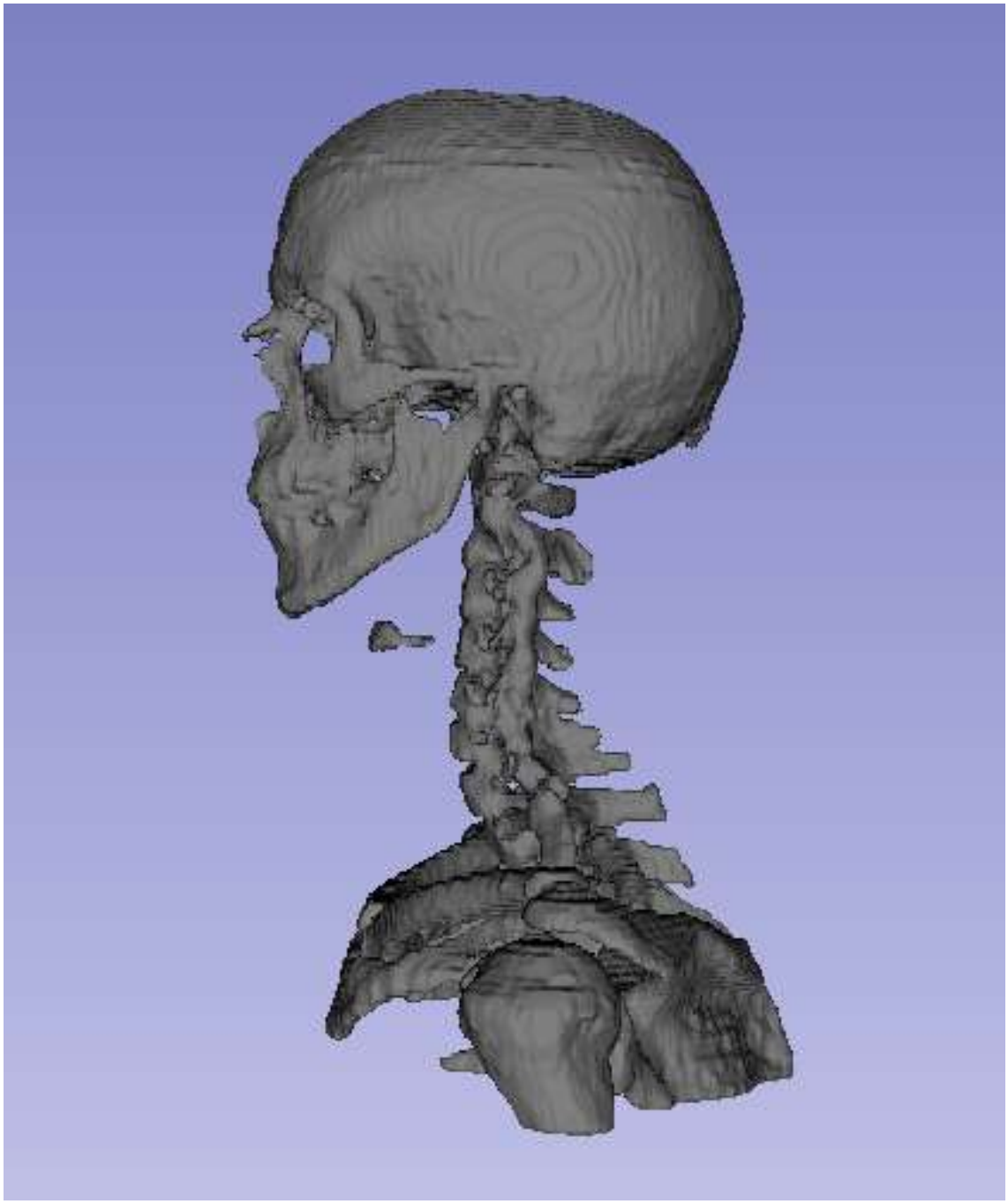

Figure 6-10 Image of the 3D structure mesh which was created as a basis for the bone in the phantom (sagittal view).

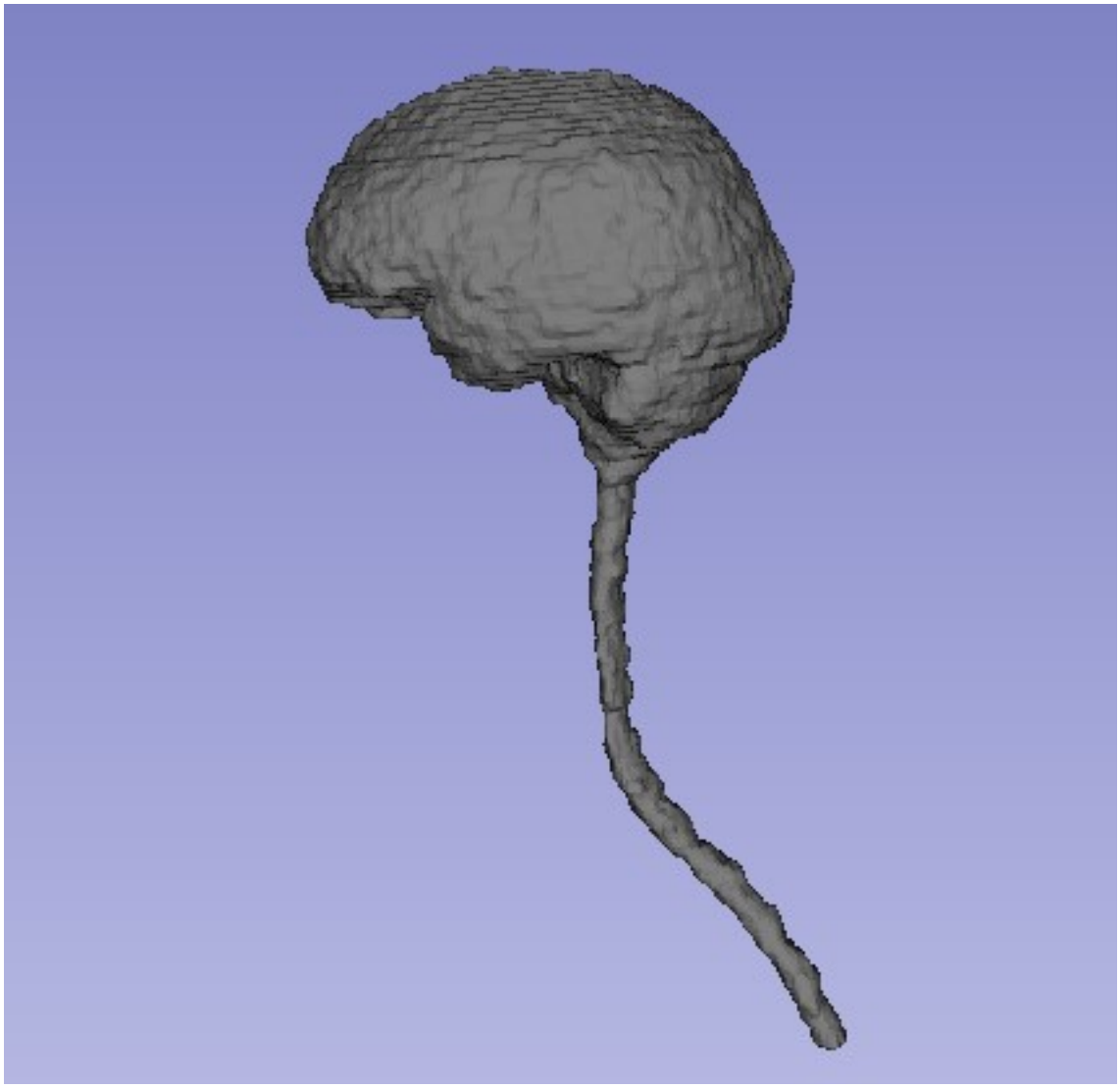

Figure 6-11 Image of the 3D structure mesh which was created as a basis for the brain, brainstem and spinal cord in the phantom (sagittal view).

Figure 6-12 Image of the 3D structure mesh which was created as a basis for the brain, brainstem and spinal cord (in green) and the eyes (in grey) in the phantom (coronal view).

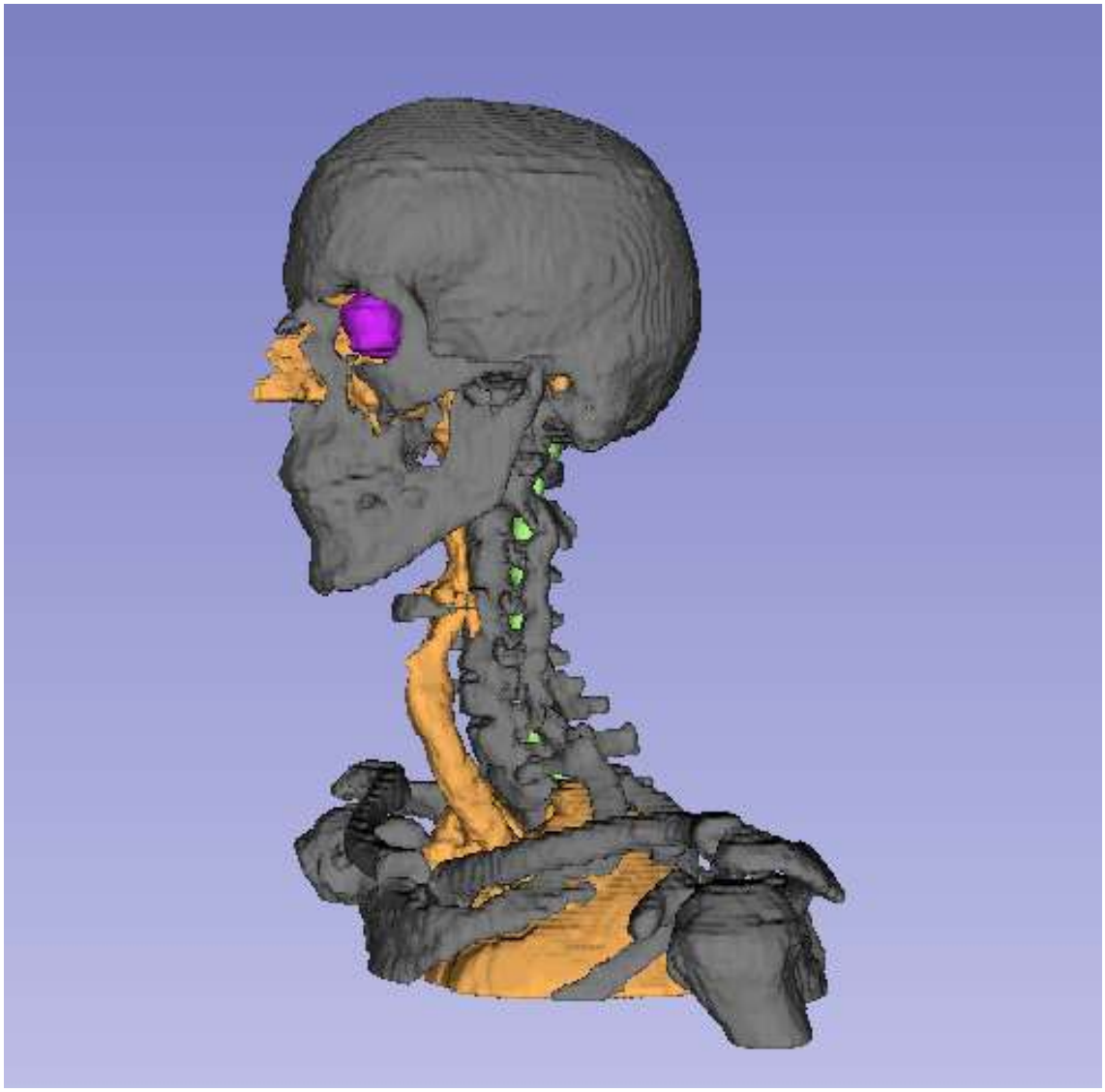

Figure 6-13 An image of all structure meshes apart from the external contour (bone represented in grey, air in orange, eyes in purple and the brain, brainstem and spinal cord in green).

Physical measurements of a semi-flex dosimetric chamber were taken at LCC. The circumference was measured to be 7.0 mm and the length was measured as 49.7 mm. These measurements were used in the production of a space for a semi-flex chamber within the casing and brain of the phantom. A plug should be produced composed of PVA-c, which can be inserted into the gap in the brain when dosimetric measurements are not being carried out. The air cavity in the thorax can also be used for chamber measurements. Film measurements were not considered in this phantom prototype, due to manufacturing constraints.

6.5 Prototype phantom design

Using the 3D structures meshes described in Section 6.4, components of the phantom were 3D printed. The outer casing was produced based on the patient outline using the photopolymer. The bone was printed based on the bone structure using the ceramic material. It should be noted that this bone structure includes the teeth, which were also produced using the ceramic material. Air cavities were created in the casing based on the delineated air structures. The phantom was built as one structure in this prototype, rather than as two detachable sections, to make the initial manufacturing process simpler.

A mould was used to create the brain based on the brain delineation shown in Figure 6-11. PVA-c was used for this with a 10% concentration and produced using four freeze-thaw cycles. Due to difficulties in the manufacturing process it was decided that the brainstem and the spinal cord would be excluded from this phantom prototype. This however is not expected to greatly affect the ability to produce sCTs based on the phantom images due to the similarity of water with cerebrospinal fluid. The spinal vertebrae will still be present.

The remainder of the phantom, including the eyes, was filled with water in this prototype. This has similar electron density to soft tissue and hence should appear approximately tissue equivalent on CT, however it is unlikely to be tissue equivalent when imaged with MRI. This will be tested for both modalities and if necessary the solution will be changed.

Images of the phantom whilst under construction can be seen in Figure 6-14.

Figure 6-14 The phantom under construction. Half of the outer casing can be seen. The inferior part of the skeleton is also present within the casing. The top section will be attached once the brain is fitted. A cavity in the superior section of the outer casing where a dosimetric chamber could be inserted can be seen. No filling caps are present in positons where the MRI coils would be used.

6.6 Prototype phantom measurements

A photographic image of the prototype phantom can be seen in Figure 6-15. The phantom largely matched the design outlined in Section 6.4, however some deviations were identified. Firstly, some components were missing from the phantom prototype. A cavity in the superior section of the phantom had been created which could fit a semi-flex dosimetric chamber (see Figure 6-14), however a PVA-c plug which could be inserted into the cavity during phantom scanning had not been provided. This did not cause problems for the initial assessment of the phantom with MRI and CT imaging, however the generation of a sCT image of the phantom was not possible without this plug.

Figure 6-15 Photograph of the completed phantom prototype. Yellow filling holes on the right shoulder can be seen. Within the phantom the skeleton, lungs and brain can be seen.

Additionally, no adipose surrogate was present in this prototype. A dental filling was also not present and all teeth were instead made from ceramic. These omissions made production of the first phantom prototype less technically challenging, but the addition of these components will be included in future

phantom versions. The initial phantom design also included a space for a second dosimetric chamber in the inferior section of the phantom in the form of the trachea. However this would not be possible using this prototype as the bottom section of the phantom is sealed (see Figure 6-16). Therefore if this is deemed useful, for example for head and neck measurements, this will be included in a future phantom version.

Figure 6-16 Photograph of the phantom, showing the inferior section which is sealed in this prototype.

6.6.1 MRI scanning

MRI scanning was performed on a Siemens Prisma scanner at 3 T. The phantom was scanned within the head coil described in Section 6.3.1 (see Figure 6-17) with an 18 channel flexible array coil over the shoulder region in order to achieve a high SNR. As with previous measurements, scanning was performed using a T_1-weighted spin echo, a T_2-weighed turbo spin echo and a FLAIR MRI sequence thereby replicating the sequences typically used in radiotherapy brain scanning at LCC. A sagittal slice through the FLAIR image of the head section of the phantom can be seen in Figure 6-18.

Figure 6-17 The prototype phantom set-up within the MRI head coil. A coil array was also added later to cover the shoulder section of the phantom.

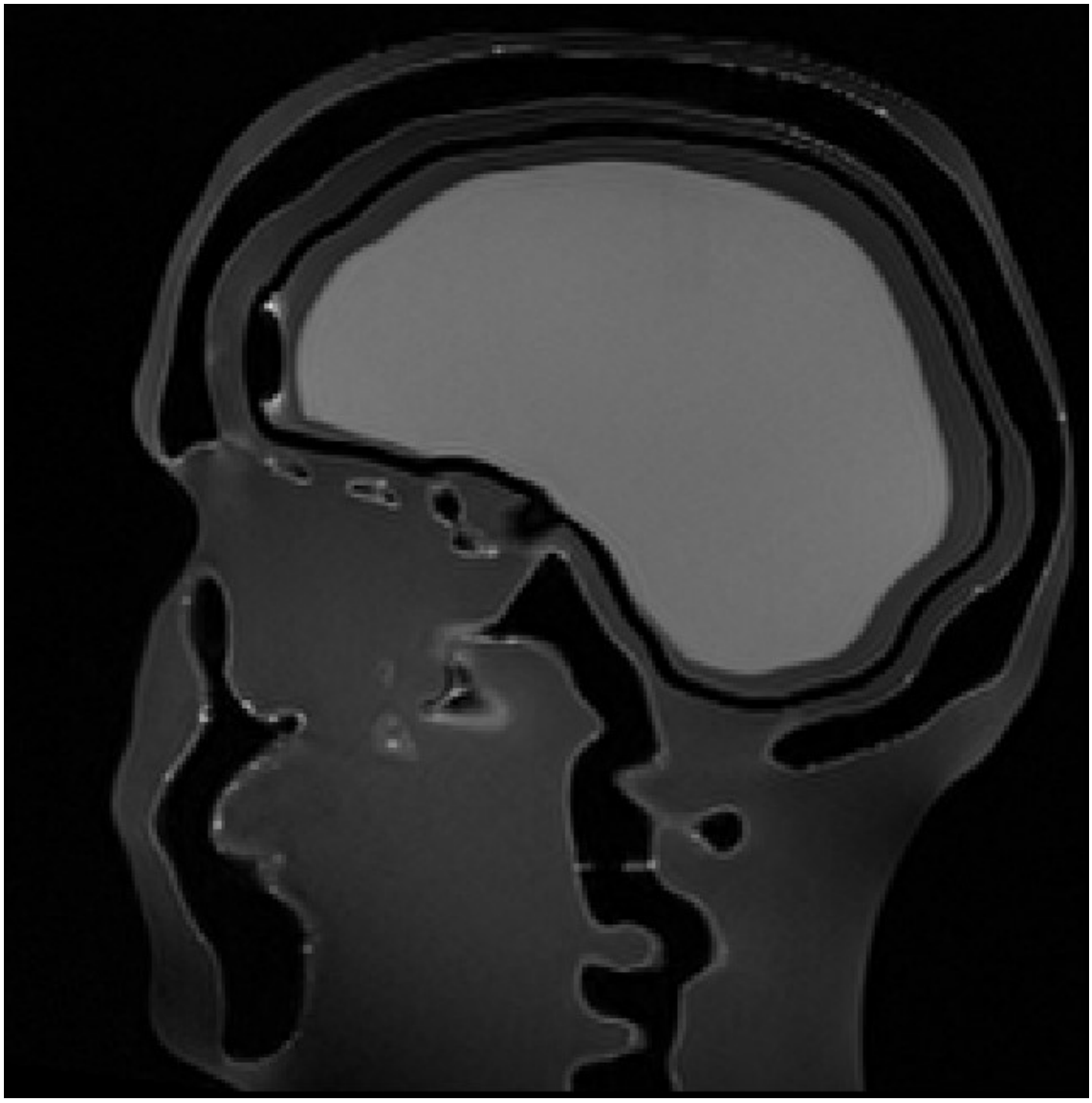

Figure 6-18 Sagittal slice through an MRI FLAIR image of the head section of the phantom. An air bubble anterior to the brain surrogate material can be seen.

T_1 and T_2 mapping were also performed using the mapping sequences described previously (see Section 6.3.1). T_1 and T_2 relaxation times were measured for the brain and soft tissue. It was found however that measurement of the T_2 of the soft tissue surrogate was not possible due to the sequence optimisation for brain imaging, where expected T_2 values are approximately one hundred times faster than for fluid water. The results are shown in Table 6-4.

	Phantom surrogate		Human tissue/ ms	
	T_1/ ms	**T_2/ ms**	**T_1/ ms**	**T_2/ ms**
Brain	1387 ± 6	139 ± 4	700-1000 (white matter) 1400-1700 (grey matter)	~70 (white matter) ~100 (grey matter)
Soft tissue	3350 ± 9	n/a	900-1400 (muscle)	25-45 (muscle)

Table 6-4 T_1 and T_2 relaxation times of the phantom tissue surrogate materials measured on a Siemens Prisma 3 T scanner. Values from the literature (179) are also shown for comparison. No measurement of the T_2 relaxation time of the soft tissue surrogate was acquired.

T_1 relaxation times for the PVA-c brain surrogate material were in good agreement with the literature (179) and were well matched with the initial measurement of 1395 ms. This value lies between the T_1 relaxation times of white and grey matter, but is better matched with grey matter, which makes up the bulk component of the brain. The T_2 relaxation time of the brain surrogate material measured higher than the T_2 times of white and grey matter reported in the literature (179), as well as being higher than the T_2 of the initial measurement of the material in Section 6.3.3.1 of 106 ms. The clinical acceptability of this material will be assessed through the production of an sCT of the phantom in future work.

In this phantom prototype, water was used to fill the remainder of the phantom. However the results in Table 6-4, show that this is not a suitable filling material and will likely result in a poor quality sCT. Therefore future work should focus on filling the phantom with a material which appears more similar to soft tissue when MRI imaging is performed. For example, Niebuhr et al., (56) found that agarose gels containing sodium fluoride and a gadolinium-based contrast agent, were a good surrogate for various soft tissues in MRI and CT imaging. Longevity of the agarose solution may be problematic however, and it may be that a doped liquid or a typical MRI phantom solution (e.g. nickel, sodium or manganese chloride) may be a more suitable surrogate.

6.6.2 CT scanning

CT scanning of the phantom was performed using the clinical head and neck protocol. The phantom set-up is shown in Figure 6-19. A head rest was used to simulate the patient set-up position. An immobilisation mask for the phantom will also be created in the future to further replicate patient scanning. Reference markers will be added to this mask to allow for a reproducible phantom set-up. HU measurements were taken within the brain, soft tissue, bone, lungs and air cavity sections of the phantom by measuring the values within three ROIs in the CT scan for each tissue surrogate. The results can be seen in Table 6-5. Mean HU values of the tissues in typical patient CT scans are also shown in Table 6-5 for comparison. As was previously performed in Section 6.3.1, ten ROIs were drawn within these patient tissues and the mean HU values calculated. A greater number of ROIs were used for the patient measurements due to the larger HU heterogeneity in patient tissue compared to the tissue surrogate materials.

	Mean HU value of phantom surrogate	**Mean HU value of patient measurements**
Air cavities	-972.1 ± 9.3	-945.3 ± 9.5
Bone	457.7 ± 42.6	1107.8 ± 75.9
Brain	20.0 ± 0.4	35.1 ± 1.1 (Grey matter) 11.8 ± 1.0 (White matter)
Lung	-765.4 ± 0.8	-709.7 ± 28.8
Soft tissue	-2.9 ± 1.5	58.9 ± 1.4

Table 6-5 The mean CT HU values of the tissue surrogate materials within the phantom. Mean HU values measured from typical patient CT scans are also shown for comparison. The error values represent the standard error of the mean.

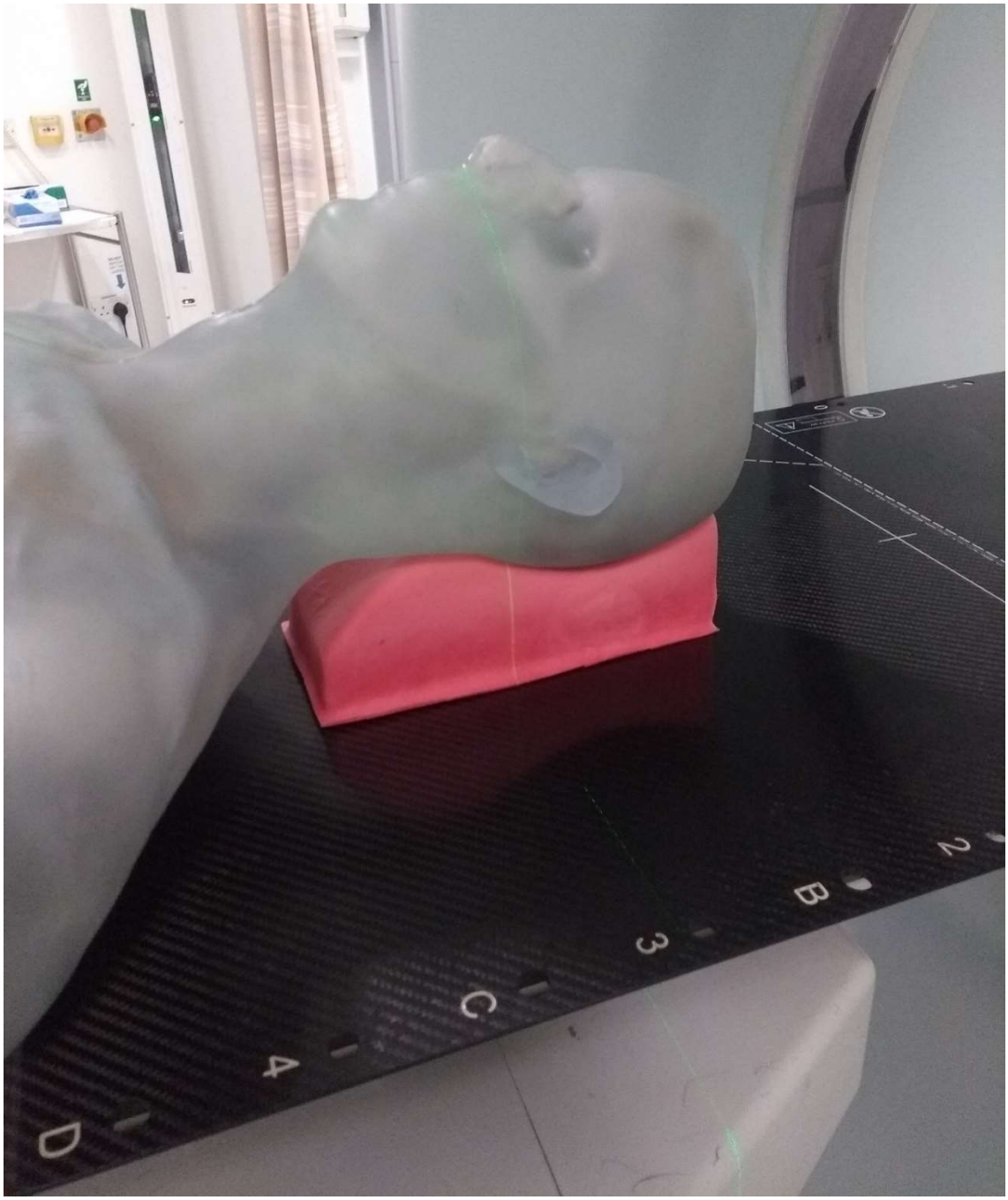

Figure 6-19 The phantom set-up in the CT scanner. The phantom was imaged on a head rest, as is the case for the clinical head and neck protocol.

Generally the HU values of the tissue surrogate materials appear well matched to the patient tissue measurements. However the HU of the ceramic material used as a bone surrogate appears lower than that of actual patient bone. This is an unexpected result as the measurement of the ceramic material was equal to 973.5 ± 10.2 HU in the preliminary measurements (Table 6-3). Upon discussion

with LTO, it was found that a different bone surrogate material had been used. This will be modified in a future phantom version.

The mean HU value of the brain tissue surrogate material (PVA-c) lies in-between the HU values of grey and white matter as expected. Air and lung (a low density polyurethane material added by LTO) also appear well matched with the mean HU values of the surrogate materials and patient tissues being within two SEMs of each other.

The mean HU of the soft tissue surrogate material is lower than that of actual patient soft tissue. This is expected however as water was used to simulate an approximation of soft tissue in this phantom prototype. However it is important to note that patients are also composed of fat as well as soft tissue. HU measurements were therefore also taken with a patient CT scan within regions of fat, and a mean HU of -101.4 ± 5.6 HU was found. Therefore it may be that using water is a good approximation of the remaining average tissue HU within a patient in a CT phantom. As discussed in Section 6.6.1, this is however unsuitable as a soft tissue surrogate in MRI imaging.

CT images through the head and thorax sections of the phantom can be seen in Figure 6-20. The complex structures, particularly the skeleton, appear to be well modelled. An air bubble can be seen in the CT image of the head section of the phantom, which will need to be corrected.

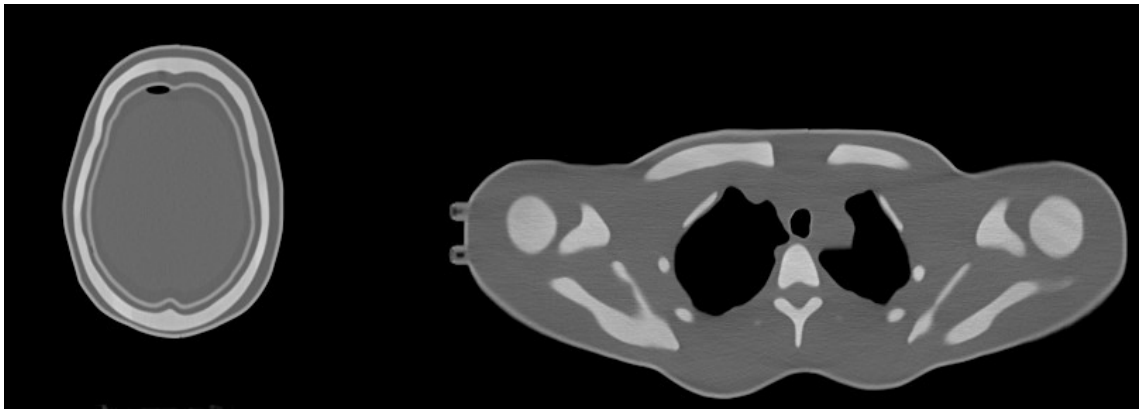

Figure 6-20 Transaxial CT images of the head and thorax sections of the phantom. The filling caps can be seen in the thorax section.

6.7 Future improvements

As has been noted in Section 6.6, future work should focus on improvements of the MRI and CT properties of the soft tissue and bone surrogates respectively. Additionally adding in missing components such as the fat section and the dental filling will improve the usefulness of the phantom.

It was noted on scanning the phantom that small air bubbles were found to exist within the soft tissue and brain sections (see Figure 6-18 and Figure 6-20). A small leak in the seal of the bottom of the phantom was also identified, which was likely contributing to the air bubbles. This was reported to LTO in order to be remedied in future phantom versions.

6.8 Conclusions

A novel head and neck anthropomorphic phantom prototype has been created for the purposes of end-to-end testing of an MRI-only radiotherapy workflow. This phantom has been shown to successfully replicate the anatomical structure of the head and neck on both MRI and CT, with some adjustments needed to the bone surrogate and phantom filling material. The suitability of using the phantom to produce sCTs should now be assessed in order to ensure that the phantom is fit for its intended purpose.

Following the acceptance of the prototype, alterations to the phantom such as the inclusion of a removable dental filling and an adipose section would improve the usefulness of the phantom and should be the focus of future work. It should be ensured that enough cavities for dosimetric chambers are present in the phantom as to make the phantom suitable for head and neck testing.

Chapter 7 Future work and Conclusions

7.1 Future project work

The aim of this project was to investigate the feasibility of MRI-only radiotherapy planning for brain patients and to find solutions to some of the challenges faced when implementing this technique. The work undertaken in this thesis was focussed around two main areas of interest. Firstly, the acceptability of brain sCTs, as well as their associated automatic organ at risk (OAR) contours, for clinical use, and secondly, the development of a quality assurance (QA) workflow related to MRI-only radiotherapy, including the assessment of geometric distortions and the end-to-end testing of the workflow.

7.1.1 Clinical implementation of sCTs

Brain sCTs produced through the multi-atlas and neural network techniques generally performed well across quality metrics and results were in-line with those reported in the literature. However it was shown in Chapter 3 that a small number of dosimetric metrics were clinically unacceptable. It is predicted that performing MRI scanning whilst the patient is in the treatment position, and in particular inside the immobilisation mask, will improve the quality of the training data and consistency of these dosimetry metrics with those recorded when radiotherapy planning is performed on the CT.

It is recommended that the first 15-20 brain patients (in line with recommendations in the literature (90)) scanned in this way, preferably on an MRI simulator, are used in the training dataset and that sCTs are generated for these patients. These should be compared to the CT both in terms of HU difference and a dosimetric comparison of radiotherapy plans, as performed in Chapter 3. This will allow confirmation that these issues, mainly regarding the volume of the planning treatment volume (PTV) which receives 105% and 107% of the prescribed dose, have been resolved. Upon commissioning an MRI-only

radiotherapy technique for clinical use, a CT should continue to be taken along with the MRI at first (182). The radiotherapy should be planned using the CT as well as the sCT and the dosimetry of the two plans compared to confirm that no significant differences between the two exist.

It should also be validated that the sCTs can be used in place of CT scans for on-line positioning verification as part of image-guided radiotherapy. For example, the sCT scans could be used as a reference for cone-beam CT (CBCT) imaging. Here, the geometric shifts needed to match the CBCT to the sCT should be compared to those needed to match the CBCT to the CT. This could also be tested with known positional errors introduced in the CBCT scans.

Before the technique is implemented clinically, the dosimetric impact of using sCTs for patients with PTVs or OARs close to large bone resections should be analysed. It may be the case that for patients in this category, MRI-only radiotherapy using an atlas method is contraindicated.

7.1.2 Automatic OAR contouring in MRI-only radiotherapy planning

The use of an atlas method for the generation of sCTs also means that OAR structures associated with the image can be automatically delineated. The main advantages of using automatic contours is that inter-observer variability is removed, as well as the resulting time saving. In order to implement this automatic contouring however it must be shown that the structures are consistently clinically acceptable, or at least would only require minor manual adjustment by a clinician. If this is not the case, then the aforementioned benefits would not exist.

It was found in Chapter 4 that large structures within the brain, such as the brainstem, spinal cord, optic nerves and globes were generally contoured well by the atlas model, meaning that these are likely candidates for introduction into the clinic. This study was focussed on proof of principle and therefore, as for the sCTs, the clinical acceptability of these structures, as well as comparison to manually drawn contours, should be assessed during commissioning of MRI-only

radiotherapy for the first cohort of patients. After the technique has been clinically implemented, automatic contours should continue to be assessed for each patient by a clinician, and adjusted where necessary.

It was found in this study that smaller structures, particularly the lacrimal and pituitary glands, were more prone to large discrepancies from the manually delineated expert contours. As these structures are small, it may not be time efficient to investigate automatic contouring for these structures further at this point. However if this was desired, it could be tested whether improvements in the placement and delineation of these structures is observed when more attention is paid to recruiting an age and gender diverse population which can then be used for training. Using advanced MRI sequences, such as ultra-short echo time (UTE) imaging, which allows cortical bone to be visualised, and fluid attenuation inversion recovery (FLAIR) sequences, where cerebrospinal fluid is suppressed, could also be included in the MRI imaging session. If information from these is incorporated into the atlas model and automatic contour generation process, it is hypothesised that improvements in the automatic contours of these structures will be observed.

In this study contouring was performed by experienced neuro-oncologists, who had previously discussed the delineation methodology, as outlined in Section 4.2.1. However it is possible that the input of an experienced neuro-radiologist during the training phase may have improved the consistency of the clinician contouring and the quality of the training dataset. For example, it was reported by the clinicians involved in this study that one area of delineation which was difficult to judge was the point at which the chiasm joined the optic nerves. A further training session with an expert radiologist may lead to improvements in the consistency of the contouring of these structures between delineators. This should be investigated in future work. Improvements in the dataset as a result could be signalled by a reduction in inter-observer variability.

Comparison of the auto-contours with intra-observer variability was not performed in this investigation due to limited clinician availability. Therefore

comparisons with inter-observer variability only were performed as an initial assessment of auto-contour acceptability. Bridge et al., (183) argued that once a clinician has acquired adequate experience and training, intra-observer variability is likely to reflect a range of clinically acceptable contours. Therefore comparison of the auto-contours with intra-observer variability, where contours have been delineated by an experienced oncologist who has received training from an experienced neuro-radiologist, would be a useful method of assessing the clinical acceptability of the auto-contours. This will be an area of investigation in future work.

Qualitative assessment of the clinical acceptability of the auto-contours was based on the clinical opinion of an experienced neuro-oncologist. To confirm these findings multiple clinicians should also carry out the same assessment. This work will be performed in a future study.

7.1.3 Quality assurance of MRI geometric distortions

It was shown in Chapter 5 that the temporal stability of MRI geometric distortions on each MRI scanner should be assessed prior to a decision being made on the required frequency of geometric distortion QA for an MRI-only radiotherapy workflow. This should be carried out for the sequences to be used in the clinical workflow. QA tests similar to those used in Chapter 5 could be incorporated in the QA protocol itself.

During this project an error was found in the software provided by Spectronic Medical AB involving incorrect distortion calculation for markers positioned at large distances from the isocentre. If this software is to be used in the future, this limitation should be taken into account, with these markers being excluded from analysis and the field of view (FOV) being limited to a radial distance-to-isocentre of less than 250 mm. This should not be of concern for brain patients, but may cause issues for large pelvic patients if MRI-only radiotherapy is used for this anatomical site. If a larger FOV is to be analysed, either a modified and validated Spectronic Medical AB code should be created, or code must be produced in-house to determine the centroids of the spheres on MRI and CT images.

7.1.4 Development of a head and neck MRI/ CT phantom

For the purposes of aiding QA of an MRI-only radiotherapy workflow, a head and neck phantom, suitable for imaging on MRI and CT, was designed and developed. It is now necessary for this phantom to undergo testing in order to ensure that it is fit for purpose.

To fully assess the suitability of the phantom for use in the end-to-end testing of an MRI-only radiotherapy workflow, sCTs of the phantom must be generated. This can be assessed for the brain using the previously developed models (see Chapter 3). The sCT images of the phantom created from MRI images should also be compared to patient sCTs and the CT of the phantom. This will be the focus of phantom validation in the future, when a PVA-c plug has been produced which can be inserted into the chamber space in the brain compartment. The production of the sCT will also mean that the acceptability of the outer casing of the phantom can be assessed. The phantom has been designed to encompass the head and neck in order to ensure that the phantom is as useful for MRI-only radiotherapy as possible. Therefore in the future it is hoped to use this phantom to create sCTs using a head and neck model.

The longevity of the phantom has not yet been tested. It has been advised by LTO that the PVA-c used to create the brain mould should last for a period of years under normal conditions. The phantom will be exposed to radiation however and it is not known the effect of this on the PVA-c. This will be in tested in the phantom prototype by comparing the HU values and MRI relaxation times to the baseline values after continuously irradiating the phantom with a large radiation dose. The geometric integrity of the PVA-c after this irradiation will also be checked by comparing the baseline and new CT scans.

The remaining components of the phantom should also remain stable over time, however the integrity of the phantom components should be monitored. It is suggested that CT and MRI scans of the phantom be acquired regularly, for example monthly, in order to ensure that the HU, T_1 and T_2 values of the components have not changed and the components still have geometric integrity.

The sCT produced from the MRI scan of the phantom should also be compared to the sCT image acquired using the phantom when it was first produced i.e. a subtraction image, in order to ensure that the two are sufficiently similar. The latter test however is not necessarily an assessment of the integrity of the phantom, as other factors can affect this. It should instead form part of the end-to-end testing of the MRI-only workflow.

An end-to-end MRI-only radiotherapy planning quality assurance protocol should be created. This should include details of the tests to be carried out, as well as the suggested frequency, for example monthly if the sCT generation is shown to be stable. The quality assurance tests should not only assess the consistency of the generated sCTs, but also the dosimetry of the plans produced using the sCT as a reference for treatment planning. The phantom should also be used for imaging with CBCT and the shifts when using the phantom sCT, MRI and CT as a reference determined. This analysis should include introducing known errors into the phantom set-up.

7.2 Conclusions

This aim of this project was to determine the clinical feasibility of MRI-only radiotherapy for brain patients and to develop processes and equipment to aid the clinical implementation of the technique i.e. to develop MRI geometric distortion and end-to-end QA tests. Through the assessment of brain sCTs, it has been shown that it is feasible to apply sCT generation techniques, previously developed for the prostate, to the brain site for patients recruited at a different centre. These sCTs were shown to be of generally high quality, however it is likely that MRI scanning of the patient in the treatment position within the immobilisation mask, will improve these further. This should be carried out before the technique is implemented clinically. It has also been shown that generation of automatic contours of large brain structures, such as the brainstem, spinal cord, optic nerves and globes, is feasible using the same technique. This is likely to be of clinical use if MRI-only radiotherapy planning is implemented.

Careful consideration should be given to QA of new radiotherapy techniques, particularly when a shift in workflow is being implemented, as is the case here. The main areas considered in this study were the assessment of geometric distortions in MRI images to be used for radiotherapy and consideration of end-to-end testing of the MRI-only workflow. It has been shown that prior to implementing the technique, full FOV phantom measurements should be taken on the specific MRI scanner in order to determine the stability of the geometric distortions for clinical MRI sequences. This could be done using the methods outlined in this study. From this the frequency of geometric distortion QA measurements can be determined. For end-to-end testing of the MRI-only pathway an anthropomorphic head and neck phantom, which could be imaged on MRI and CT, was developed. This would allow testing of the workflow for brain, as well as head and neck patients. The feasibility of the production of this dual modality anthropomorphic phantom has been shown and acceptance and commissioning of the phantom should now be carried out. The production of these phantoms for other areas of the body where MRI-only radiotherapy techniques are being developed should be the focus of future research.

It has been shown that MRI-only radiotherapy planning techniques developed for prostate patients can be used for the generation of sCTs for brain patients, along with automatic contouring of larger brain structures. This project has developed an end-to-end QA phantom and also provided recommendations for the monitoring of MRI geometric distortions when the scanner is to be used for MRI-only radiotherapy. The next phase of the project will focus on clinical implementation of the technique, for which it is recommended that patients are scanned in the treatment position and a QA framework for the technique is produced.

List of References

1. Dirix P, Haustermans K, Vandecaveye V. The value of magnetic resonance imaging for radiotherapy planning. *Seminars in Radiation Oncology*. 2014;24(3):151-9. doi:10.1016/j.semradonc.2014.02.003.

2. Nyholm T, Jonsson J. Counterpoint: opportunities and challenges of a magnetic resonance imaging only radiotherapy work flow. *Seminars in Radiation Oncology*. 2014;24(3):175-80. doi:10.1016/j.semradonc.2014.02.005.

3. Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. *Radiation Oncology*. 2017;12(1):28. doi:10.1186/s13014-016-0747-y.

4. Liney GP, Moerland MA. Magnetic resonance imaging acquisition techniques for radiotherapy planning. *Semin Radiat Oncol*. 2014;24(3):160-8. doi:10.1016/j.semradonc.2014.02.014.

5. Ishikura S. Quality assurance of radiotherapy in cancer treatment: toward improvement of patient safety and quality of care. *Jpn J Clin Oncol*. 2008;38(11):723-9. doi:10.1093/jjco/hyn112.

6. Kutcher GJ, Coia L, Gillin M, et al. Comprehensive QA for radiation oncology: report of AAPM Radiation Therapy Committee Task Group 40. *Med Phys*. 1994;21(4):581-618. doi:10.1118/1.597316.

7. Fraass B, Doppke K, Hunt M, et al. American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: quality assurance for clinical radiotherapy treatment planning. *Med Phys*. 1998;25(10):1773-829. doi:10.1118/1.598373.

8. Chang DS, Lasley FD, Das IJ, Mendonca MS, Dynlacht JR. Therapeutic Ratio. *Basic Radiotherapy Physics and Biology*. Cham: Springer; 2014.

9. Bentzen SM. 3 - Fractionation Effects in Clinical Practice. In: Hoppe RT, Phillips TL, Roach M, editors. *Leibel and Phillips Textbook of Radiation Oncology (Third Edition)*. Philadelphia: W.B. Saunders; 2010. p. 40-54.

10. Stuben G, Thews O, Pottgen C, Stuschke M, Sack H. Tumour oxygenation during fractionated radiotherapy--comparison with size-matched controls. *Acta Oncol*. 1999;38(2):209-13.

11. Pereira GC, Traughber M, Muzic RF. The Role of Imaging in Radiation Therapy Planning: Past, Present, and Future. *BioMed Research International*. 2014;2014:9. doi:10.1155/2014/231090.

12. Elert G. Photoelectric effect. In: *The Physics Hypertextbook*. 2014. http://physics.info/photoelectric/. Accessed 10/03/2019.

13. Hubbell JH. Electron–positron pair production by photons: A historical overview. *Radiation Physics and Chemistry*. 2006;75(6):614-23. doi:https://doi.org/10.1016/j.radphyschem.2005.10.008.

14. Compton AH. A Quantum Theory of the Scattering of X-rays by Light Elements. *Physical Review*. 1923;21(5):483-502. doi:10.1103/PhysRev.21.483.

15. Goyal S, Kataria T. Image Guidance in Radiation Therapy: Techniques and Applications. *Radiology Research and Practice*. 2014;2014:10. doi:10.1155/2014/705604.

16. McRobbie D, Moore E, Graves M, Prince M. *MRI from picture to proton*. 2nd ed. Cambridge: Cambridge University Press; 2007.

17. Devic S. MRI simulation for radiotherapy treatment planning. *Med Phys*. 2012;39(11):6701-11. doi:10.1118/1.4758068.

18. Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. *Med Phys*. 2017;44(7):e43-e76. doi:10.1002/mp.12256.

19. Lagendijk JJ, Raaymakers BW, Raaijmakers AJ, et al. MRI/linac integration. *Radiother Oncol*. 2008;86(1):25-9. doi:10.1016/j.radonc.2007.10.034.

20. Kupelian P, Sonke J-J. Magnetic Resonance Guided Adaptive Radiotherapy: A Solution to the Future. *Seminars in Radiation Oncology*. 2014;24(3):227-32. doi:10.1016/j.semradonc.2014.02.013.

21. Schmidt M, A. , Payne G, S. . Radiotherapy planning using MRI. *Phys Med Biol*. 2015;60(22):R323. doi:10.1088/0031-9155/60/22/R323.

22. Creak AL, Tree A, Saran F. Radiotherapy planning in high-grade gliomas: a survey of current UK practice. *Clin Oncol (R Coll Radiol)*. 2011;23(3):189-98. doi:10.1016/j.clon.2010.11.003.

23. Hanks G, Chavaudra J, Dobbs J, et al. Report 50. *Journal of the International Commission on Radiation Units and Measurements*. 2016;os26(1). doi:10.1093/jicru/os26.1.Report50.

24. Aoyama H, Shirato H, Nishioka T, et al. Magnetic resonance imaging system for three-dimensional conformal radiotherapy and its impact on gross tumor volume delineation of central nervous system tumors. *International Journal of Radiation Oncology*Biology*Physics*. 2001;50(3):821-7. doi:https://doi.org/10.1016/S0360-3016(01)01598-X.

25. Stejskal EO, Tanner JE. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *The Journal of Chemical Physics*. 1965;42(1):288-92. doi:10.1063/1.1695690.

26. Pantelis E, Papadakis N, Verigos K, et al. Integration of functional MRI and white matter tractography in stereotactic radiosurgery clinical practice. *Int J Radiat Oncol Biol Phys*. 2010;78(1):257-67. doi:10.1016/j.ijrobp.2009.10.064.

27. Demol B, Boydev C, Korhonen J, Reynaert N. Dosimetric characterization of MRI-only treatment planning for brain tumors in atlas-based pseudo-CT images generated from standard T1-weighted MR images. *Med Phys*. 2016;43(12):6557. doi:10.1118/1.4967480.

28. Nyholm T, Nyberg M, Karlsson MG, Karlsson M. Systematisation of spatial uncertainties for comparison between a MR and a CT-based radiotherapy workflow for prostate treatments. *Radiation Oncology*. 2009;4(1):1-9. doi:10.1186/1748-717x-4-54.

29. Andreasen D, Van Leemput K, Hansen RH, Andersen JAL, Edmund JM. Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain. *Med Phys*. 2015;42(4):1596-605. doi:http://dx.doi.org/10.1118/1.4914158.

30. Ulin K, Urie MM, Cherlow JM. Results of a multi-institutional benchmark test for cranial CT/MR image registration. *Int J Radiat Oncol Biol Phys*. 2010;77(5):1584-9. doi:10.1016/j.ijrobp.2009.10.017.

31. Lagendijk JJ, Raaymakers BW, Vulpen M. The magnetic resonance imaging-linac system. *Semin Radiat Oncol*. 2014;24. doi:10.1016/j.semradonc.2014.02.009.

32. Hanson LG. Is quantum mechanics necessary for understanding magnetic resonance? *Concepts in Magnetic Resonance Part A*. 2008;32A(5):329-40. doi:10.1002/cmr.a.20123.

33. Elster A. Questions and Answers in MRI 2019. Available from: http://mriquestions.com/. Date accessed: 08/11/2019.

34. Bloch F. Nuclear Induction. *Physical Review*. 1946;70(7-8):460-74. doi:10.1103/PhysRev.70.460.

35. Westbrook C, Kaut Roth C, Talbot J. *MRI in practice*. 4th ed. Chichester, UK: Wiley-Blackwell; 2011.

36. Liney G. *MRI in Clinical Practice*. 1st ed. London, UK: Spinger-Verlag; 2006.

37. Karlsson M, Karlsson MG, Nyholm T, Amies C, Zackrisson B. Dedicated magnetic resonance imaging in the radiotherapy clinic. *Int J Radiat Oncol Biol Phys*. 2009;74. doi:10.1016/j.ijrobp.2009.01.065.

38. Sprawls P. *Magnetic Resonance Imaging: Principles, Methods, and Techniques*: Medical Physics Publishing; 2000.

39. Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Med Phys*. 1998;25(5):656-61. doi:10.1118/1.598248.

40. Torfeh T, Hammoud R, McGarry M, Al-Hammadi N, Perkins G. Development and validation of a novel large field of view phantom and a software module for the quality assurance of geometric distortion in magnetic resonance imaging. *Magnetic Resonance Imaging*. 2015;33(7):939-49. doi:https://doi.org/10.1016/j.mri.2015.04.003.

41. Schad L, Lott S, Schmitt F, Sturm V, Lorenz WJ. Correction of spatial distortion in MR imaging: a prerequisite for accurate stereotaxy. *J Comput Assist Tomogr*. 1987;11(3):499-505.

42. Schad LR, Ehricke H-H, Wowra B, et al. Correction of spatial distortion in magnetic resonance angiography for radiosurgical treatment planning of cerebral arteriovenous malformations. *Magnetic Resonance Imaging*. 1992;10(4):609-21. doi:https://doi.org/10.1016/0730-725X(92)90012-O.

43. Kawanaka A, Takagi M. Estimation of static magnetic field and gradient fields from NMR image. *Journal of Physics E: Scientific Instruments*. 1986;19(10):871.

44. Wang D, Strugnell W, Cowin G, Doddrell DM, Slaughter R. Geometric distortion in clinical MRI systems: part I: evaluation using a 3D phantom. *Magnetic Resonance Imaging*. 2004;22(9):1211-21. doi:https://doi.org/10.1016/j.mri.2004.08.012.

45. Wang D, Doddrell DM, Cowin G. A novel phantom and method for comprehensive 3-dimensional measurement and correction of geometric distortion in magnetic resonance imaging. *Magnetic Resonance Imaging*. 2004;22(4):529-42. doi:https://doi.org/10.1016/j.mri.2004.01.008.

46. Weygand J, Fuller CD, Ibbott GS, et al. Spatial precision in magnetic resonance imaging–guided radiation therapy: the role of geometric distortion. *International Journal of Radiation Oncology*Biology*Physics*. 2016;95(4):1304-16. doi:https://doi.org/10.1016/j.ijrobp.2016.02.059.

47. Janke A, Zhao H, Cowin GJ, Galloway GJ, Doddrell DM. Use of spherical harmonic deconvolution methods to compensate for nonlinear gradient effects on MRI images. *Magnetic Resonance in Medicine*. 2004;52(1):115-22. doi:10.1002/mrm.20122.

48. Bhagwandien R, Moerland MA, Bakker CJG, Beersma R, Lagendijk JJW. Numerical analysis of the magnetic field for arbitrary magnetic susceptibility distributions in 3D. *Magnetic Resonance Imaging*. 1994;12(1):101-7. doi:https://doi.org/10.1016/0730-725X(94)92357-4.

49. Nieder C, Grosu AL, Gaspar LE. Stereotactic radiosurgery (SRS) for brain metastases: a systematic review. *Radiation Oncology*. 2014;9(1):155. doi:10.1186/1748-717x-9-155.

50. Shuryak I, Carlson DJ, Brown JM, Brenner DJ. High-dose and fractionation effects in stereotactic radiation therapy: Analysis of tumor control data from 2965 patients. *Radiotherapy and Oncology*. 2015;115(3):327-34. doi:https://doi.org/10.1016/j.radonc.2015.05.013.

51. Pappas EP, Alshanqity M, Moutsatsos A, et al. MRI-related geometric distortions in stereotactic radiotherapy treatment planning: evaluation and dosimetric impact. *Technol Cancer Res Treat*. 2017;16(6):1120-9. doi:10.1177/1533034617735454.

52. Wilson D, Wright, G., Lymer, K., McRobbie, D. Geometric parameters: spatial resolution and distortion. In: McRobbie D, Semple S, editors. *IPEM Report 112: Quality control and artefacts in Magnetic Resonance Imaging*: Institute of Physics and Engineering in Medicine; 2017.

53. Chen Z, Ma CM, Paskalev K, et al. Investigation of MR image distortion for radiotherapy treatment planning of prostate cancer. *Phys Med Biol*. 2006;51(6):1393-403. doi:10.1088/0031-9155/51/6/002.

54. Walker A, Liney G, Metcalfe P, Holloway L. MRI distortion: considerations for MRI based radiotherapy treatment planning. *Australas Phys Eng Sci Med*. 2014;37(1):103-13. doi:10.1007/s13246-014-0252-2.

55. Klein EE, Hanley J, Bayouth J, et al. Task Group 142 report: quality assurance of medical accelerators. *Med Phys*. 2009;36(9):4197-212. doi:10.1118/1.3190392.

56. Niebuhr NI, Johnen W, Güldaglar T, et al. Technical Note: Radiological properties of tissue surrogates used in a multimodality deformable pelvic phantom for MR-guided radiotherapy. *Med Phys*. 2016;43(2):908-16. doi:10.1118/1.4939874.

57. Menikou G, Dadakova T, Pavlina M, Bock M, Damianou C. MRI compatible head phantom for ultrasound surgery. *Ultrasonics*. 2015;57:144-52. doi:http://dx.doi.org/10.1016/j.ultras.2014.11.004.

58. Chu KC, Rutt BK. Polyvinyl alcohol cryogel: an ideal phantom material for MR studies of arterial flow and elasticity. *Magn Reson Med*. 1997;37(2):314-9.

59. Surry KJ, Austin HJ, Fenster A, Peters TM. Poly(vinyl alcohol) cryogel phantoms for use in ultrasound and MR imaging. *Phys Med Biol*. 2004;49(24):5529-46.

60. Wardman K, Prestwich RJ, Gooding MJ, Speight RJ. The feasibility of atlas-based automatic segmentation of MRI for H&N radiotherapy planning. *J Appl Clin Med Phys*. 2016;17(4):146-54. doi:10.1120/jacmp.v17i4.6051.

61. Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies. *Radiotherapy and Oncology*. 2016;121(2):169-79. doi:https://doi.org/10.1016/j.radonc.2016.09.009.

62. Johnstone E, Wyatt JJ, Henry AM, et al. Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging-only radiation therapy. *International Journal of Radiation Oncology • Biology • Physics*. 2018;100(1):199-217. doi:10.1016/j.ijrobp.2017.08.043.

63. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097.

64. Schad LR, Blüml S, Hawighorst H, Wenz F, Lorenz WJ. Radiosurgical treatment planning of brain metastases based on a fast, three-dimensional MR imaging technique. *Magnetic Resonance Imaging*. 1994;12(5):811-9. doi:http://dx.doi.org/10.1016/0730-725X(94)92206-3.

65. Ramsey CR, Oliver AL. Magnetic resonance imaging based digitally reconstructed radiographs, virtual simulation, and three-dimensional treatment planning for brain neoplasms. *Med Phys*. 1998;25(10):1928-34.

66. Prabhakar R, Julka PK, Ganesh T, et al. Feasibility of using MRI alone for 3D radiation treatment planning in brain tumors. *Jpn J Clin Oncol*. 2007;37. doi:10.1093/jjco/hym050.

67. Wang C, Chao M, Lee L, Xing L. MRI-based treatment planning with electron density information mapped from CT images: a preliminary study. *Technol Cancer Res Treat*. 2008;7(5):341-8.

68. Weber DC, Wang H, Albrecht S, et al. Open low-field magnetic resonance imaging for target definition, dose calculations and set-up verification during three-dimensional CRT for glioblastoma multiforme. *Clinical Oncology (Royal College of Radiologists)*. 2008;20(2):157-67.

69. Chen L, Price RA, Wang L, et al. MRI-based treatment planning for radiotherapy: dosimetric verification for prostate IMRT. *Int J Radiat Oncol Biol Phys*. 2004;60. doi:10.1016/j.ijrobp.2004.05.068.

70. Chen L, R. A. Price J, Nguyen TB, et al. Dosimetric evaluation of MRI-based treatment planning for prostate cancer. *Phys Med Biol*. 2004;49(22):5157.

71. Eilertsen K, Nilsen Tor Arne Vestad L, Geier O, Skretting A. A simulation of MRI based dose calculations on the basis of radiotherapy planning CT images. *Acta Oncologica*. 2008;47(7):1294-302. doi:10.1080/02841860802256426.

72. Karotki A, Mah K, Meijer G, Meltsner M. Comparison of bulk electron density and voxel-based electron density treatment planning. *Journal of Applied*

Clinical Medical Physics. 2011;12(4):3522. doi:http://dx.doi.org/10.1120/jacmp.v12i4.3522.

73. Korsholm ME, Waring LW, Edmund JM. A criterion for the reliable use of MRI-only radiotherapy. *Radiation Oncology*. 2014;9:16. doi:http://dx.doi.org/10.1186/1748-717X-9-16.

74. Chin AL, Lin A, Anamalayil S, Teo BK. Feasibility and limitations of bulk density assignment in MRI for head and neck IMRT treatment planning. *J Appl Clin Med Phys*. 2014;15(5):4851. doi:10.1120/jacmp.v15i5.4851.

75. Doemer A, Chetty IJ, Glide-Hurst C, et al. Evaluating organ delineation, dose calculation and daily localization in an open-MRI simulation workflow for prostate cancer patients. *Radiation Oncology*. 2015;10(1):1-9. doi:10.1186/s13014-014-0309-0.

76. Jonsson JH, Karlsson MG, Karlsson M, Nyholm T. Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions. *Radiation Oncology*. 2010;5:62. doi:http://dx.doi.org/10.1186/1748-717X-5-62.

77. Kristensen BH, Laursen FJ, Logager V, Geertsen PF, Krarup-Hansen A. Dosimetric and geometric evaluation of an open low-field magnetic resonance simulator for radiotherapy treatment planning of brain tumours. *Radiotherapy & Oncology*. 2008;87(1):100-9. doi:http://dx.doi.org/10.1016/j.radonc.2008.01.014.

78. Lambert J, Greer PB, Menk F, et al. MRI-guided prostate radiation therapy planning: Investigation of dosimetric accuracy of MRI-based dose planning. *Radiotherapy & Oncology*. 2011;98(3):330-4. doi:http://dx.doi.org/10.1016/j.radonc.2011.01.012.

79. Lee YK, Bollet M, Charles-Edwards G, et al. Radiotherapy treatment planning of prostate cancer using magnetic resonance imaging alone. *Radiotherapy & Oncology*. 2003;66(2):203-16.

80. Stanescu T, Hans-Sonke J, Pervez N, Stavrev P, Fallone BG. A study on the magnetic resonance imaging (MRI)-based radiation treatment planning of

intracranial lesions. *Phys Med Biol*. 2008;53(13):3579-93. doi:10.1088/0031-9155/53/13/013.

81. Stanescu T, Hans-Sonke J, Stavrev P, Fallone G. 3T MR-based treatment planning for radiotherapy of brain lesions. *Radiology and Oncology*. 2006;40(2):125-32.

82. Ramsey CR, Arwood D, Scaperoth D, Oliver AL. Clinical application of digitally-reconstructed radiographs generated from magnetic resonance imaging for intracranial lesions. *International Journal of Radiation Oncology*Biology*Physics*. 1999;45(3):797-802. doi:http://dx.doi.org/10.1016/S0360-3016(99)00173-X.

83. Chen L, Nguyen TB, Jones E, et al. Magnetic resonance-based treatment planning for prostate intensity-modulated radiotherapy: creation of digitally reconstructed radiographs. *Int J Radiat Oncol Biol Phys*. 2007;68(3):903-11.

84. Hoogcarspel SJ, Van der Velden JM, Lagendijk JJ, van Vulpen M, Raaymakers BW. The feasibility of utilizing pseudo CT-data for online MRI based treatment plan adaptation for a stereotactic radiotherapy treatment of spinal bone metastases. *Phys Med Biol*. 2014;59(23):7383-91. doi:10.1088/0031-9155/59/23/7383.

85. Dowling JA, Sun J, Pichler P, et al. Automatic Substitute Computed Tomography Generation and Contouring for Magnetic Resonance Imaging (MRI)-Alone External Beam Radiation Therapy From Standard MRI Sequences. *Int J Radiat Oncol Biol Phys*. 2015;93(5):1144-53. doi:10.1016/j.ijrobp.2015.08.045.

86. Sjolund J, Forsberg D, Andersson M, Knutsson H. Generating patient specific pseudo-CT of the head from MR using atlas-based regression. *Phys Med Biol*. 2015;60(2):825-39. doi:10.1088/0031-9155/60/2/825.

87. Greer PB, Dowling JA, Lambert JA, et al. A magnetic resonance imaging-based workflow for planning radiation therapy for prostate cancer. *Medical Journal of Australia*. 2011;194(4):S24-7.

88. Dowling JA, Lambert J, Parker J, et al. An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy. *Int J Radiat Oncol Biol Phys*. 2012;83(1):e5-11. doi:http://dx.doi.org/10.1016/j.ijrobp.2011.11.056.

89. Uh J, Merchant TE, Li Y, Li X, Hua C. MRI-based treatment planning with pseudo CT generated through atlas registration. *Med Phys*. 2014;41(5):051711. doi:http://dx.doi.org/10.1118/1.4873315.

90. Siversson C, Nordström F, Nilsson T, et al. Technical Note: MRI only prostate radiotherapy planning using the statistical decomposition algorithm. *Med Phys*. 2015;42(10):6090-7. doi:http://dx.doi.org/10.1118/1.4931417.

91. Andreasen D, Van Leemput K, Edmund JM. A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis. *Med Phys*. 2016;43(8):4742-52. doi:10.1118/1.4958676.

92. Zhou W, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*. 2004;13(4):600-12. doi:10.1109/TIP.2003.819861.

93. Jonsson JH, Johansson A, Soderstrom K, Asklund T, Nyholm T. Treatment planning of intracranial targets on MRI derived substitute CT data. *Radiotherapy & Oncology*. 2013;108(1):118-22. doi:http://dx.doi.org/10.1016/j.radonc.2013.04.028.

94. Johansson A, Karlsson M, Yu J, Asklund T, Nyholm T. Voxel-wise uncertainty in CT substitute derived from MRI. *Med Phys*. 2012;39(6):3283-90. doi:http://dx.doi.org/10.1118/1.4711807.

95. Kapanen M, Tenhunen M. T1/T2 -weighted MRI provides clinically relevant pseudo-CT density data for the pelvic bones in MRI-only based radiotherapy treatment planning. *Acta Oncologica*. 2013;52(3):612-8. doi:http://dx.doi.org/10.3109/0284186X.2012.692883.

96. Korhonen J, Kapanen M, Keyrilainen J, Seppala T, Tenhunen M. A dual model HU conversion from MRI intensity values within and outside of bone

segment for MRI-based radiotherapy treatment planning of prostate cancer. *Med Phys*. 2014;41(1):011704. doi:http://dx.doi.org/10.1118/1.4842575.

97. Korhonen J, Kapanen M, Keyrilainen J, et al. Absorbed doses behind bones with MR image-based dose calculations for radiotherapy treatment planning. *Med Phys*. 2013;40(1):011701. doi:http://dx.doi.org/10.1118/1.4769407.

98. Korhonen J, Kapanen M, Keyrilainen J, et al. Influence of MRI-based bone outline definition errors on external radiotherapy dose calculation accuracy in heterogeneous pseudo-CT images of prostate cancer patients. *Acta Oncol*. 2014;53(8):1100-6. doi:10.3109/0284186x.2014.929737.

99. Yu H, Caldwell C, Balogh J, Mah K. Toward magnetic resonance-only simulation: segmentation of bone in MR for radiation therapy verification of the head. *Int J Radiat Oncol Biol Phys*. 2014;89(3):649-57. doi:10.1016/j.ijrobp.2014.03.028.

100. Kim J, Glide-Hurst C, Doemer A, et al. Implementation of a novel algorithm for generating synthetic CT images from magnetic resonance imaging data sets for prostate cancer radiation therapy. *Int J Radiat Oncol Biol Phys*. 2015;91(1):39-47. doi:http://dx.doi.org/10.1016/j.ijrobp.2014.09.015.

101. Kim J, Garbarino K, Schultz L, et al. Dosimetric evaluation of synthetic CT relative to bulk density assignment-based magnetic resonance-only approaches for prostate radiotherapy. *Radiat Oncol*. 2015;10:239. doi:10.1186/s13014-015-0549-7.

102. Korhonen J, Kapanen M, Sonke JJ, et al. Feasibility of MRI-based reference images for image-guided radiotherapy of the pelvis with either cone-beam computed tomography or planar localization images. *Acta Oncol*. 2015;54(6):889-95. doi:10.3109/0284186x.2014.958197.

103. Koivula L, Wee L, Korhonen J. Feasibility of MRI-only treatment planning for proton therapy in brain and prostate cancers: Dose calculation accuracy in substitute CT images. *Med Phys*. 2016;43(8):4634. doi:10.1118/1.4958677.

104. Robson MD, Gatehouse PD, Bydder M, Bydder GM. Magnetic resonance: an introduction to ultrashort TE (UTE) imaging. *J Comput Assist Tomogr*. 2003;27(6):825-46.

105. Larson PEZ, Han M, Krug R, et al. Ultrashort echo time and zero echo time MRI at 7T. *Magma (New York, N.Y.)*. 2016;29(3):359-70. doi:10.1007/s10334-015-0509-0.

106. Lee H, Zhao X, Song HK, et al. Rapid dual-RF, dual-echo, 3D ultrashort echo time craniofacial imaging: A feasibility study. *Magn Reson Med*. 2019;81(5):3007-16. doi:10.1002/mrm.27625.

107. Rank CM, Hunemohr N, Nagel AM, et al. MRI-based simulation of treatment plans for ion radiotherapy in the brain region. *Radiotherapy & Oncology*. 2013;109(3):414-8. doi:http://dx.doi.org/10.1016/j.radonc.2013.10.034.

108. Rank CM, Tremmel C, Hunemohr N, et al. MRI-based treatment plan simulation and adaptation for ion radiotherapy using a classification-based approach. *Radiation Oncology*. 2013;8:51. doi:http://dx.doi.org/10.1186/1748-717X-8-51.

109. Johansson A, Karlsson M, Nyholm T. CT substitute derived from MRI sequences with ultrashort echo time. *Med Phys*. 2011;38(5):2708-14. doi:http://dx.doi.org/10.1118/1.3578928.

110. Johansson A, Garpebring A, Karlsson M, Asklund T, Nyholm T. Improved quality of computed tomography substitute derived from magnetic resonance (MR) data by incorporation of spatial information--potential application for MR-only radiotherapy and attenuation correction in positron emission tomography. *Acta Oncologica*. 2013;52(7):1369-73. doi:http://dx.doi.org/10.3109/0284186X.2013.819119.

111. Johansson A, Garpebring A, Asklund T, Nyholm T. CT substitutes derived from MR images reconstructed with parallel imaging. *Med Phys*. 2014;41(8):082302. doi:http://dx.doi.org/10.1118/1.4886766.

112. Hsu SH, Cao Y, Huang K, Feng M, Balter JM. Investigation of a method for generating synthetic CT models from MRI scans of the head and neck for radiotherapy. *Phys Med Biol*. 2013;58. doi:10.1088/0031-9155/58/23/8419.

113. Edmund JM, Kjer HM, Van Leemput K, et al. A voxel-based investigation for MRI-only radiotherapy of the brain using ultra short echo times. *Phys Med Biol*. 2014;59(23):7501-19. doi:10.1088/0031-9155/59/23/7501.

114. Jonsson JH, Akhtari MM, Karlsson MG, et al. Accuracy of inverse treatment planning on substitute CT images derived from MR data for brain lesions. *Radiation Oncology*. 2015;10(1):1-7. doi:10.1186/s13014-014-0308-1.

115. Hsu SH, Cao Y, Lawrence TS, et al. Quantitative characterizations of ultrashort echo (UTE) images for supporting air-bone separation in the head. *Phys Med Biol*. 2015;60(7):2869-80. doi:10.1088/0031-9155/60/7/2869.

116. Paradis E, Cao Y, Lawrence TS, et al. Assessing the Dosimetric Accuracy of Magnetic Resonance-Generated Synthetic CT Images for Focal Brain VMAT Radiation Therapy. *Int J Radiat Oncol Biol Phys*. 2015;93(5):1154-61. doi:10.1016/j.ijrobp.2015.08.049.

117. Zheng W, Kim JP, Kadbi M, et al. Magnetic Resonance-Based Automatic Air Segmentation for Generation of Synthetic Computed Tomography Scans in the Head Region. *Int J Radiat Oncol Biol Phys*. 2015;93(3):497-506. doi:10.1016/j.ijrobp.2015.07.001.

118. Gutierrez S, Descamps B, Vanhove C. MRI-Only Based Radiotherapy Treatment Planning for the Rat Brain on a Small Animal Radiation Research Platform (SARRP). *PLoS One*. 2015;10(12):e0143821. doi:10.1371/journal.pone.0143821.

119. Price RG, Kim JP, Zheng W, Chetty IJ, Glide-Hurst C. Image Guided Radiation Therapy Using Synthetic Computed Tomography Images in Brain Cancer. *Int J Radiat Oncol Biol Phys*. 2016;95(4):1281-9. doi:10.1016/j.ijrobp.2016.03.002.

120. Yang Y, Cao M, Kaprealian T, et al. Accuracy of UTE-MRI-based patient setup for brain cancer radiation therapy. *Med Phys*. 2016;43(1):262-7. doi:10.1118/1.4938266.

121. Grodzki DM, Jakob PM, Heismann B. Ultrashort echo time imaging using pointwise encoding time reduction with radial acquisition (PETRA). *Magn Reson Med*. 2012;67(2):510-8. doi:10.1002/mrm.23017.

122. Gudur MS, Hara W, Le QT, et al. A unifying probabilistic Bayesian approach to derive electron density from MRI for radiation therapy treatment planning. *Phys Med Biol*. 2014;59(21):6595-606. doi:10.1088/0031-9155/59/21/6595.

123. Keereman V, Fierens Y, Broux T, et al. MRI-based attenuation correction for PET/MRI using ultrashort echo time sequences. *J Nucl Med*. 2010;51(5):812-8. doi:10.2967/jnumed.109.065425.

124. Perkins G. (Personal correspondance). Philips Electronics UK Ltd; 2017.

125. Greer PB. (Personal correspondance). Calvary Mater Newcastle, Newcastle, Australia;2017.

126. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys*. 2017;44(4):1408-19. doi:10.1002/mp.12155.

127. Largent A, Barateau A, Nunes JC, et al. Pseudo-CT Generation for MRI-Only Radiation Therapy Treatment Planning: Comparison Among Patch-Based, Atlas-Based, and Bulk Density Methods. *Int J Radiat Oncol Biol Phys*. 2019;103(2):479-90. doi:10.1016/j.ijrobp.2018.10.002.

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

129. Dinkla AM, Wolterink JM, Maspero M, et al. MR-Only Brain Radiation Therapy: Dosimetric Evaluation of Synthetic CTs Generated by a Dilated Convolutional Neural Network. *Int J Radiat Oncol Biol Phys*. 2018;102(4):801-12. doi:10.1016/j.ijrobp.2018.05.058.

130. Maspero M, Savenije MHF, Dinkla AM, et al. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. *Phys Med Biol*. 2018;63(18):185001. doi:10.1088/1361-6560/aada6d.

131. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging*. 2012;30(9):1323-41. doi:10.1016/j.mri.2012.05.001.

132. Eaton J, Bateman D, Hauberg S, Wehbring R. GNU Octave version 4.0.0 manual: a high-level interactive language for numerical computations 2015. Available from: http://www.gnu.org/software/octave/doc/interpreter.

133. Pinter C, Lasso A, Wang A, Jaffray D, Fichtinger G. SlicerRT: radiation therapy research toolkit for 3D Slicer. *Med Phys*. 2012;39(10):6332-8. doi:10.1118/1.4754659.

134. Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *Neuroimage*. 2009;46(3):726-38. doi:https://doi.org/10.1016/j.neuroimage.2009.02.018.

135. Siversson C, Akhondi-Asl A, Bixby S, Kim YJ, Warfield SK. Three-dimensional hip cartilage quality assessment of morphology and dGEMRIC by planar maps and automated segmentation. *Osteoarthritis Cartilage*. 2014;22(10):1511-5. doi:10.1016/j.joca.2014.08.012.

136. Report 83. *Journal of the International Commission on Radiation Units and Measurements*. 2010;10(1). doi:10.1093/jicru/10.1.Report83.

137. Kamalian S, Lev MH, Gupta R. Chapter 1 - Computed tomography imaging and angiography – principles. In: Masdeu JC, González RG, editors. *Handbook of Clinical Neurology*. 135: Elsevier; 2016. p. 3-20.

138. Lustberg T, van Soest J, Gooding M, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. *Radiother Oncol*. 2018;126(2):312-7. doi:10.1016/j.radonc.2017.11.012.

139. Sharp G, Fritscher KD, Pekar V, et al. Vision 20/20: Perspectives on automated image segmentation for radiotherapy. *Med Phys*. 2016;41(5):050902. doi:10.1118/1.4871620.

140. Bauer S, Wiest R, Nolte LP, Reyes M. A survey of MRI-based medical image analysis for brain tumor studies. *Phys Med Biol*. 2013;58(13):R97-129. doi:10.1088/0031-9155/58/13/r97.

141. Eekers DBP, in 't Ven L, Roelofs E, et al. The EPTN consensus-based atlas for CT- and MR-based contouring in neuro-oncology. *Radiotherapy and Oncology*. 2018;128(1):37-43. doi:https://doi.org/10.1016/j.radonc.2017.12.013.

142. Scoccianti S, Detti B, Gadda D, et al. Organs at risk in the brain and their dose-constraints in adults and in children: a radiation oncologist's guide for delineation in everyday practice. *Radiother Oncol*. 2015;114(2):230-8. doi:10.1016/j.radonc.2015.01.016.

143. Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology*. 1945;26(3):297-302. doi:10.2307/1932409.

144. Chuter R, Prestwich R, Bird D, et al. The use of deformable image registration to integrate diagnostic MRI into the radiotherapy planning pathway for head and neck cancer. *Radiotherapy and Oncology*. 2017;122(2):229-35. doi:10.1016/j.radonc.2016.07.016.

145. Jena R, Kirkby NF, Burton KE, et al. A novel algorithm for the morphometric assessment of radiotherapy treatment planning volumes. *Br J Radiol*. 2010;83(985):44-51. doi:10.1259/bjr/27674581.

146. Hughes GK, Miszkiel KA. Imaging of the Lacrimal Gland. *Seminars in Ultrasound, CT and MRI*. 2006;27(6):476-91. doi:https://doi.org/10.1053/j.sult.2006.09.002.

147. Evanson J. Radiology of the Pituitary. South Dartmouth (MA)2016. Available from: https://www.ncbi.nlm.nih.gov/books/NBK279161/.

148. Isambert A, Dhermain F, Bidault F, et al. Evaluation of an atlas-based automatic segmentation software for the delineation of brain organs at risk in a

radiation therapy clinical context. *Radiotherapy and Oncology*. 2008;87(1):93. doi:https://doi.org/10.1016/j.radonc.2007.11.030.

149. Brouwer CL, Steenbakkers RJHM, Bourhis J, et al. CT-based delineation of organs at risk in the head and neck region: DAHANCA, EORTC, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines. *Radiotherapy and Oncology*. 2015;117(1):83-90. doi:10.1016/j.radonc.2015.07.041.

150. Price RG, Knight RA, Hwang K-P, et al. Optimization of a novel large field of view distortion phantom for MR-only treatment planning. *Journal of Applied Clinical Medical Physics*. 2017;18(4):51-61. doi:10.1002/acm2.12090.

151. Walker A, Liney G, Holloway L, et al. Continuous table acquisition MRI for radiotherapy treatment planning: distortion assessment with a new extended 3D volumetric phantom. *Med Phys*. 2015;42(4):1982-91. doi:10.1118/1.4915920.

152. Torfeh T, Hammoud R, Perkins G, et al. Characterization of 3D geometric distortion of magnetic resonance imaging scanners commissioned for radiation therapy planning. *Magnetic Resonance Imaging*. 2016;34(5):645-53. doi:https://doi.org/10.1016/j.mri.2016.01.001.

153. Moerland MA. *Magnetic resonance imaging in radiotherapy treatment planning*: University of Utrecht; 1996.

154. Ahmed M, Schmidt M, Sohaib A, et al. The value of magnetic resonance imaging in target volume delineation of base of tongue tumours – A study using flexible surface coils. *Radiotherapy and Oncology*. 2010;94(2):161-7. doi:https://doi.org/10.1016/j.radonc.2009.12.021.

155. Mizowaki T, Nagata Y, Okajima K, et al. Reproducibility of geometric distortion in magnetic resonance imaging based on phantom studies. *Radiother Oncol*. 2000;57(2):237-42.

156. Boas M. Binomial distribution. *Mathematical Methods in the Physical Sciences*. 2nd ed: John Wiley and sons; 1983.

157. NIST/SEMATECH. Measures of skewness and kurtosis. In: *NIST/ SEMATECH e-Handbook of statistical methods*. National Institute of Standards and Technology, United States Department of Commerce; 2013. http://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm. Accessed 09/03/2018.

158. Paulson ES, Erickson B, Schultz C, Li XA. Comprehensive MRI simulation methodology using a dedicated MRI scanner in radiation oncology for external beam radiation treatment planning. *Med Phys*. 2015;42(1):28-39. doi:10.1118/1.4896096.

159. Mah D, Steckner M, Palacio E, et al. Characteristics and quality assurance of a dedicated open 0.23 T MRI for radiation therapy simulation. *Med Phys*. 2002;29(11):2541-7. doi:10.1118/1.1513991.

160. Price RG, Kadbi M, Kim J, et al. Technical note: characterization and correction of gradient nonlinearity induced distortion on a 1.0 T open bore MR-SIM. *Med Phys*. 2015;42(10):5955-60. doi:10.1118/1.4930245.

161. Baldwin LN, Wachowicz K, Thomas SD, Rivest R, Fallone BG. Characterization, prediction, and correction of geometric distortion in 3 T MR images. *Med Phys*. 2007;34(2):388-99. doi:10.1118/1.2402331.

162. Kapanen M, Collan J, Beule A, et al. Commissioning of MRI-only based treatment planning procedure for external beam radiotherapy of prostate. *Magn Reson Med*. 2013;70(1):127-35. doi:10.1002/mrm.24459.

163. Damyanovich AZ, Rieker M, Zhang B, Bissonnette JP, Jaffray DA. Design and implementation of a 3D-MR/CT geometric image distortion phantom/analysis system for stereotactic radiosurgery. *Phys Med Biol*. 2018;63(7):075010. doi:10.1088/1361-6560/aab33e.

164. Chen HH, Boykin RD, Clarke GD, Gao JH, Roby JW, 3rd. Routine testing of magnetic field homogeneity on clinical MRI systems. *Med Phys*. 2006;33(11):4299-306. doi:10.1118/1.2359229.

165. Schenck JF. The role of magnetic susceptibility in magnetic resonance imaging: MRI magnetic compatibility of the first and second kinds. *Med Phys*. 1996;23(6):815-50. doi:10.1118/1.597854.

166. Haidekker MA. Magnetic Resonance Imaging. *Medical Imaging Technology*. New York: Springer Science & Business Media; 2013.

167. Rasband WS. ImageJ. U. S. National Institutes of Health, Bethesda, Maryland, USA1997-2018.

168. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*. 2012;9:671. doi:10.1038/nmeth.2089.

169. Leymarie F, Levine MD. Fast raster scan distance propagation on the discrete rectangular lattice. *CVGIP: Image Understanding*. 1992;55(1):84-94. doi:https://doi.org/10.1016/1049-9660(92)90008-Q.

170. Cancer Research UK. Brain, other CNS and intracranial tumours incidence statistics 2019. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/brain-other-cns-and-intracranial-tumours/incidence#heading-One. Date accessed: 09/02/2019.

171. Gallas RR, Hunemohr N, Runz A, et al. An anthropomorphic multimodality (CT/MRI) head phantom prototype for end-to-end tests in ion radiotherapy. *Z Med Phys*. 2015;25(4):391-9. doi:10.1016/j.zemedi.2015.05.003.

172. Fortune S, Jansen MA, Anderson T, et al. Development and characterization of rodent cardiac phantoms: comparison with in vivo cardiac imaging. *Magnetic Resonance Imaging*. 2012;30(8):1186-91. doi:https://doi.org/10.1016/j.mri.2012.04.008.

173. Dental treatments: NHS Choices; updated 25/11/2015. Available from: http://www.nhs.uk/Livewell/dentalhealth/Pages/Dentaltreatments.aspx. Date accessed: 14/08/2017.

174. Mail N, Albarakati Y, Ahmad Khan M, et al. The impacts of dental filling materials on RapidArc treatment planning and dose delivery: Challenges and solution. *Med Phys*. 2013;40(8):081714. doi:10.1118/1.4816307.

175. Allahverdi M, Nisbet A, Thwaites DI. An evaluation of epoxy resin phantom materials for megavoltage photon dosimetry. *Physics in Medicine & Biology*. 1999;44(5):1125.

176. Messroghli DR, Radjenovic A, Kozerke S, et al. Modified Look-Locker inversion recovery (MOLLI) for high-resolution T1 mapping of the heart. *Magnetic Resonance in Medicine*. 2004;52(1):141-6. doi:10.1002/mrm.20110.

177. Formula Saint-Gobain. Crystacal R 2019. Available from: https://www.saintgobainformula.com/product/crystacal-r. Date accessed: 09/02/2019.

178. Formula Saint-Gobain. Dental Plaster 2019. Available from: https://www.saintgobainformula.com/product/dental-plaster. Date accessed: 09/02/2019.

179. Bojorquez JZ, Bricq S, Acquitter C, et al. What are normal relaxation times of tissues at 3 T? *Magnetic Resonance Imaging*. 2017;35:69-80. doi:https://doi.org/10.1016/j.mri.2016.08.021.

180. Kim S, Lee GH, Lee S, et al. Body fat measurement in computed tomography image. *Biomed Sci Instrum*. 1999;35:303-8.

181. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. 2006;31(3):1116-28. doi:10.1016/j.neuroimage.2006.01.015.

182. Persson E, Gustafsson C, Nordstrom F, et al. MR-OPERA: A Multicenter/Multivendor Validation of Magnetic Resonance Imaging-Only Prostate Treatment Planning Using Synthetic Computed Tomography Images. *Int J Radiat Oncol Biol Phys*. 2017;99(3):692-700. doi:10.1016/j.ijrobp.2017.06.006.

183. Bridge P, Fielding A, Rowntree P, Pullar A. Intraobserver Variability: Should We Worry? *Journal of Medical Imaging and Radiation Sciences*. 2016;47(3):217-20. doi:10.1016/j.jmir.2016.06.004.