The optimisation of oropharyngeal radiotherapy to reduce the risk of long-term toxicities

A thesis submitted to The University of Manchester for the degree of Doctor of Medicine in the Faculty of Biology, Medicine and Health

Christina J. Hague

2020

School of Medical Sciences

Contents

WORD COUNT: 59,517 words

List of Figures

List of Tables

Glossary

ABAS   Atlas based auto-segmentation

AI   Artificial intelligence

ART   Adaptive radiotherapy

CBCT   Cone beam computer tomography

CT   Computer tomography

CTV   Clinical target volume

2D   Two-dimensional

3D   Three-dimensional

DIR   Deformable image registration

DSC   Dice similarity coefficient

DTA   Distance to agreement

DW-MRI   Diffusion weighted-magnetic resonance imaging

GTV   Gross tumour volume

Gy   Gray

HNSCC   Head and neck squamous cell cancer

HPV   Human papilloma virus

ICRU   International Commission of radiation units and measurements

IGRT   Image guided radiation therapy

IMPT   Intensity-modulated proton therapy

IMRT   Intensity-modulated radiation therapy

MFO   Multi-field optimisation

MRI   Magnetic resonance imaging

MRIgRT   Magnetic resonance image guided radiotherapy

MRL   MR Linac

NTCP   Normal tissue complication probability

OAR   Organ at risk

OE-MRI   Oxygen enhanced magnetic resonance imaging

OPC   Oropharyngeal cancer

SFO   Single field optimisation

SMG   Submandibular gland

TMJ   Temporo-mandibular joint

TNM   Tumour, nodal, metastasis

PTV   Planning target volume

QOL   Quality of life

VMAT   Volumetric modulated arc therapy

Abstract

The University of Manchester
Christina Hague, Doctor of Medicine
The Optimisation of Oropharyngeal Radiotherapy to reduce the risk of long-term toxicities

Purpose: The prevalence of long-term survivors with head and neck cancer is increasing. A number of strategies are required to optimise organ sparing, to improve long term quality of life. These include: establishing a correct dose threshold for organs at risk (OAR) in avoidance planning; using guidelines and auto-contouring models to standardise volumes for optimal dose delivery; adaptive re-planning to de-escalate normal tissue dose in response to tumour shrinkage and anatomical changes; exploiting the therapeutic advantages of proton beam therapy as an alternative modality and exploring the willingness of patients to travel to receive proton beam therapy in the UK's first proton trial.

Aims: (1) To establish a tolerance dose for the masticatory apparatus for use in avoidance radiotherapy planning to reduce trismus. (2) To use a novel muscles of mastication atlas to standardise volumes and improve consistency and optimise dose delivery. (3) To assess the benefits of a novel CT deep learning auto-contouring model to improve clinician workload and reduce inter-observer variability. (4) To develop and evaluate a novel MR deep learning auto contouring model for adaptive re-planning. (5) To investigate the dosimetric consequences of uncertainties in set up and range with proton beam therapy in post-operative oropharyngeal and oral cavity cancers. (6) To evaluate if patients are willing to travel and stay away from home to receive proton beam therapy in the UK's first proton trial. (7) To determine if oxygen-enhanced MRI (OE-MRI) is feasible in head and neck cancer by assessing ability to detect an oxygen signal and patient tolerability.

Results: (1) There was a significant association between doses >40 Gy to the ipsilateral block, lateral pterygoid and masseter and deterioration in trismus. (2) The atlas significantly reduced interobserver variability for the muscles of mastication and improved contouring consistency by trainees compared with consultants. (3) An optimised CT deep learning auto contouring model (model$_{CT}$) reduced time and inter-observer variability compared with manual contours for OAR delineation. (4) The performance of a novel MR deep learning auto-contouring model (model$_{MRI}$) to define contours was sensitive to differences in image acquisition parameters but compared with CT models, geometric accuracy with manual contours increased. (5) Multi-field optimisation is robust to inter-fraction uncertainties in set-up and range, without compromising OAR mean dose in postoperative oropharyngeal and oral cavity cancer. (6) Patients are willing to travel and stay away from home to receive proton beam therapy. (7) The OE-MRI study has received ethical approval, but is yet to begin study recruitment.

Conclusions: (1) Tolerance dose to the ipsilateral block, lateral pterygoid and masseter of ≤40 Gy for tumours not invading the masticatory apparatus may improve morbidity. (2) A muscles of mastication atlas improved standardisation of volumes and has a role as an educational tool for trainees and radiographers. (3) Model$_{CT}$ has the potential to improve the adaptive radiotherapy workflow by providing quick and efficient contours. (4) Image sequence optimisation on the MR-Linac will improve the use of model$_{MRI}$ in MR image guided radiotherapy to standardise OAR delineation and reduce dose. (5) Development of a robust analysis protocol will identify plans that need additional individualisation and improve consistency of plan reporting and evaluation amongst centres. (6) Involvement of patients early in a trial design to gain feedback on the patient pathway and study endpoints is invaluable to better inform and shape the trial design.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university of other institute of learning.

Copyright Statement

Acknowledgements

I would like to thank my principal supervisor Professor Nick Slevin for his continued guidance, support and expertise. To all the patients whom without I would not have been able to have undergone this research. I am grateful to the support and collaboration from colleagues at the Christie NHS Foundation Trust, Radiotherapy Related Research Group and the University of Pennsylvania.

I was fortunate to have four outstanding co-supervisors. Professor Catharine West provided invaluable guidance and support during my MD and in preparation of papers and my thesis; Professor Marcel van Herk provided medical physics expertise; Dr David Thomson and Dr Andrew McPartlin both helped as clinical supervisors to further develop my knowledge and skills as well as offering academic guidance and support. I would also like to thank my advisor Professor Corrine Faivre-Finn for her advice and encouragement.

I would like to thank the Head and Neck Charitable fund for supporting my work and to the Tomkins Family Bursary of which without I would not have been able to complete my fellowship at the University of Pennsylvania.

I would like to thank my mum, my husband James, family and friends for their support over the last 4 years.

Dedication

To my husband James for his patience and constant support and daughters' Imogen aged 3 and Lauren aged 1.

To my mum for always being there with a listening ear and positive encouragement and to my dad who died before my research began, but whom inspired me to pursue a career in oncology and to always achieve my dreams.

Preface

This thesis is presented in the literature report and journal paper format. The published papers are outlined and will be presented.

This thesis investigates novel strategies to reduce normal tissue toxicities and improve quality of life, in long term survivors with head and neck cancer. Strategies include: (1) standardising volumes where no contouring guidelines are available; (2) normal tissue dose avoidance in adaptive re-planning and (3) exploiting the benefits of different treatment modalities such as proton beam therapy to further widen the therapeutic ratio.

Chapter 1 is a systematic literature review summarising current knowledge of the background, epidemiology and management of oropharyngeal cancer. Areas that still need to be addressed have been identified and a number of strategies are outlined in the seven thesis aims.

Chapter 2 details the methods used to address the seven aims outlined in this thesis. Chapters 3,4,5,6,7 and 8 are in the form of journal articles, of which all except chapters 5 and 6 have been accepted for peer reviewed publication. Chapters 3 and 4 focus on developing a dose threshold and contouring guideline to help standardise volumes for use in avoidance planning, to reduce trismus. Chapters 5 and 6 compare novel CT and MR deep learning auto contouring tools to produce quick and consistent volumes in adaptive re-planning. Chapters 7 and 8 evaluate methods to optimise plan uncertainty in proton beam therapy and explore patients' willingness to travel for proton beam therapy as part of the UK's first proton beam therapy trial. Chapter 9 outlines the first study protocol to investigate the use of imaging predictive biomarkers, to detect hypoxia, with a view to dose escalation in adaptive radiotherapy.

Chapter 10 summarises the results and limitations of the previous chapters and discusses the possibilities for future work, using the outcomes of this thesis.

Author biography

I graduated from medical school at the University of Manchester in 2006. Between 2009 and 2010 I spent a year working alongside oncologists in Australia. Here I developed a keen interest in research and was co-author on the LEAP study in breast cancer. On my return in August 2010, I was awarded the MRCP and was successfully ranked first in National Recruitment for Clinical Oncology training. I began Clinical Oncology speciality training at The Christie in August 2011 and completed the final FRCR in 2015. During training I developed a keen interest in head and neck research and the importance of investigating methods to minimise radiotherapy related long term toxicities and improve patients' quality of life.

To date my postgraduate Master's Degree has led to four peer-reviewed publications with a fifth in draft form. I enjoy the blend of clinical and technical radiotherapy research with the patient journey at the core. I hope to continue clinical research in my future career as a clinical oncology consultant.

1.0 Introduction

1.1 Head and neck cancer

The incidence of head and neck squamous cell cancer (HNSCC) in the UK and worldwide has risen by a third since the early 1990s. In the UK in 2017, 12,200 new cases of HNSCC were diagnosed per annum and there were an estimated 3,989 deaths (1). There has been a 16% increase in the number of cases of HNSCC diagnosed over the last decade particularly in deprived social areas. HNSCC is currently the eighth most common type of cancer in the UK, with 46% to 88% of cases being preventable (2). The peak age at diagnosis is 70-74 years and 69% of cases are in males (1). Worldwide 708,000 HNSCC cases are diagnosed and there are an estimated 358,000 deaths per annum (3).

The main causative factors are smoking and alcohol. Analyses by the large international head and neck cancer epidemiology consortium (INHANCE), based on 14 studies published between 1981-2010, showed smoking and alcohol are the strongest risk factors for developing HNSCC. Data from these studies showed by increasing smoking intensity from ≤ 20 to >20 cigarettes per day, an individual's 20-year absolute risk of HNSCC is predicted to increase by 4.9% (95% CI 6.4%-11.3%). The absolute risk of HNSCC can be calculated using a risk prediction model that combines multiple factors. For example, an individual's 20-year absolute risk of HNSCC is 0.1% if male, aged 45 years, a lifetime non-smoker, alcohol intake <1 drink per day, no family history and educated beyond high school (4). The risk of HNSCC dramatically increases to 9.3% for an older male (aged 60 years), >20 pack year smoking history, alcohol intake >3 drinks per day, with a family history and lower levels of education (4). The strongest associations were observed for cancers of the oral cavity and larynx.

Overall survival from HNSCC is dependent on tumour subtype and is highest for those aged between 15-49 years (2). Men and women diagnosed with oropharyngeal cancer (OPC) have superior 1 year and 5 year survival rates, compared with cancers of the hypopharynx, e.g., 84% versus 60% by 1 year, falling to 60% versus 27% by 5 years (5).

OPC is the second commonest HNSCC subsite in the UK and has a recognised association with human papilloma virus (HPV). Data from the office for National Cancer statistics updated in July 2018 reported 2,295 oropharynx cases in men and 712 in women per annum in the UK (6). Between 60-80% of OPCs are HPV positive. HPV16 and 18 are the commonest subtypes, with >80% due to HPV16. In a large multi-centre UK cross sectional retrospective study, OPC tissue blocks were assessed to determine the HPV status (n=1,602). The prevalence of HPV positivity was highest in cancers of the tonsil (61.8%) and lowest for those originating in the soft palate or uvula (9.1%) (7). HPV positive HNSCCs are considered a distinct entity. They commonly present in patients who are younger (median age 57 v 64 years) and less likely to smoke (8,9). HPV positive HNSCCs carry a much better prognosis with superior treatment response, loco-regional control and 3 year survival rates (82% v 35%) compared with HPV negative (10). The absence of smoking and low nodal burden within the neck are good prognostic factors. Ang et al. described superior 3 year overall survival rates for HPV positive OPC of 82% v 57% for HPV negative, which increased to 93% in non-smokers and with low nodal disease (N0-N2) within the neck (11). Loco-regional control at 3 years in HPV positive OPC is 95% for early stage and 78% for advanced stage compared with 76% and 62% in HPV negative OPC (9). It is therefore important to spare late effects in those with HPV positive OPC.

1.2 Anatomy and staging of oropharyngeal cancers

The anatomy of the head and neck is complex, due to the number and proximity of critical structures. The pharynx has three portions, the nasopharynx, oropharynx and laryngopharynx. The oropharynx lies behind the oral cavity, extending inferiorly to the level of the hyoid bone and is lined by squamous epithelium. It consists of the posterior third of the tongue, vallecula, bilateral tonsils, inferior border of the soft palate and the uvula (Figure 1.0).
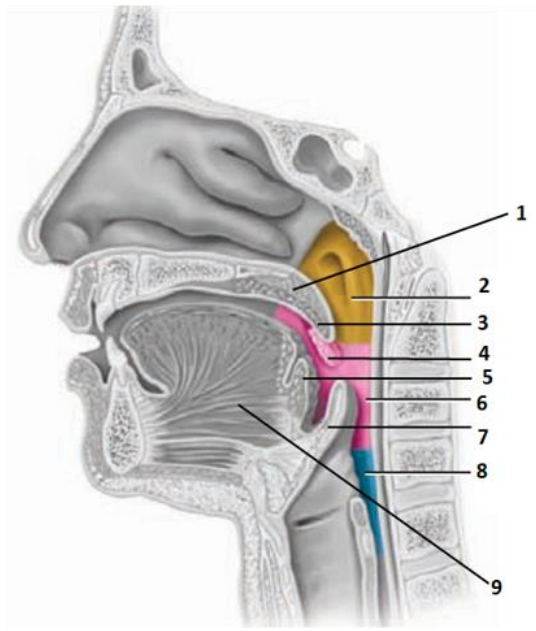
1 Soft palate

2 Nasopharynx

3 Uvula

4 Palatine tonsil

5 Vallecula

6 Oropharynx

7 Epiglottis

8 Laryngopharynx

9 Base of Tongue

Figure 1.0. Sagittal diagram of the pharynx with oropharynx subsites[1]

1.2.1 Function of the oropharynx including muscles of mastication

The muscles and structures of the oropharynx are responsible for swallowing and speech. There are four sets of masticatory muscles: medial, lateral pterygoids, masseter and temporalis as illustrated in Figure 1.2. The masticatory muscles open and close the jaw, by moving the mandible at the level of the temporo-mandibular joint (TMJ), as shown in Figure 1.3 and Table 1.0. Histologically the muscles vary in their muscle fibre make up. The lateral pterygoid, the only muscle to open the jaw consists of type 1 fibres, whereas the jaw closing muscles contain type 1 and 2 fibres are arranged in a complex pattern (12,13). The masseter and medial pterygoid muscles work synergistically with each other, the masseter being the strongest.

---

[1] Image adapted from (302)

1 Temporalis

2 Lateral pterygoid

3 Zygomatic arch

4 Masseter

5 Mandible

6 Medial pterygoid

7 Temporo-mandibular joint (TMJ)
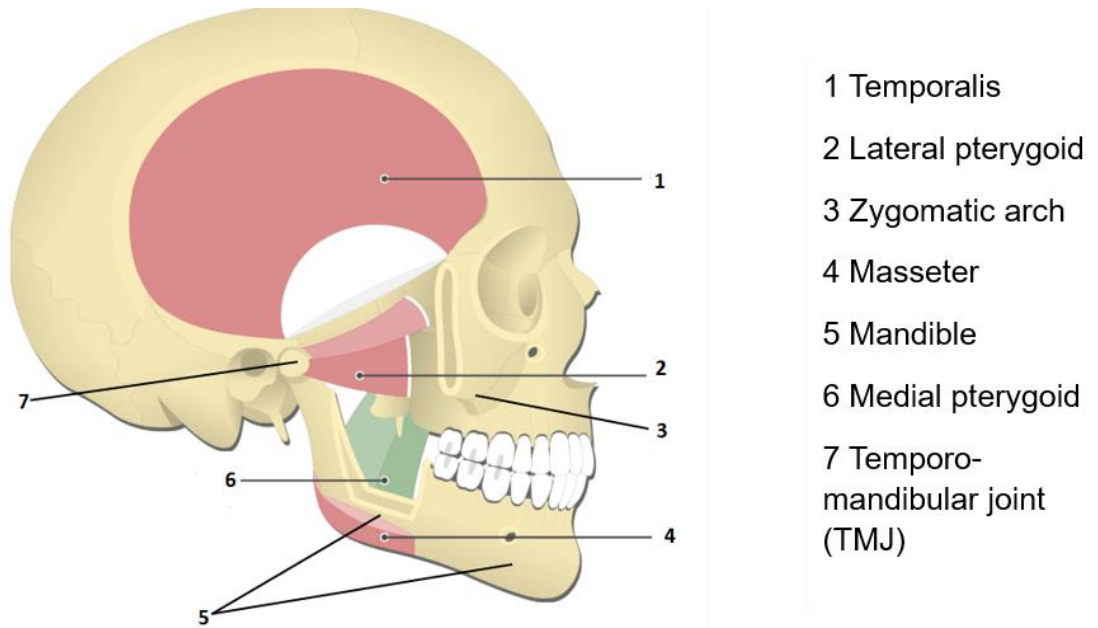
Figure 1.1. Diagram illustrating the muscles of mastication including the TMJ[2]

Lift

Depress

Retract

Protrude
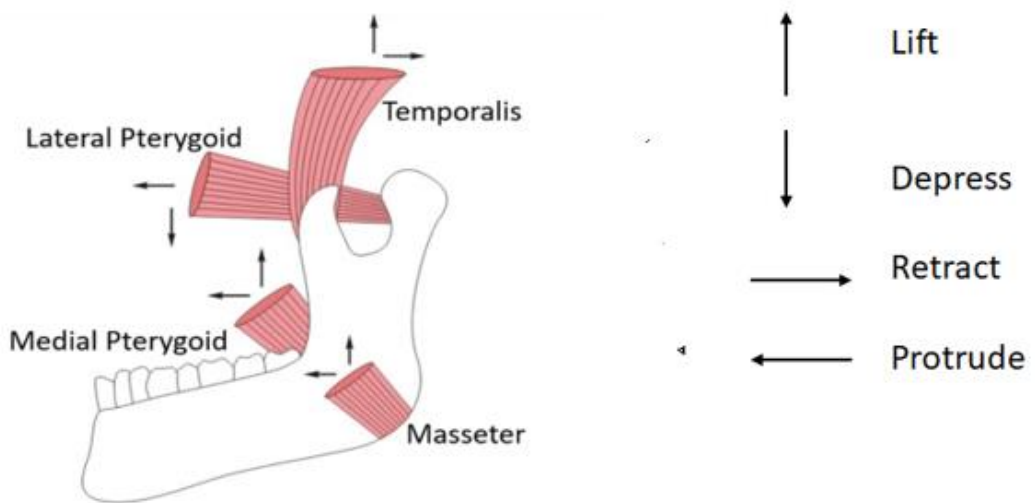
Figure 1.2. Movement of the jaw by action of the masticatory muscles [3]

[2] Image adapted from (303)
[3] Image adapted from (304)

Table 1.0. Comparison of the structure and function of the masticatory muscles

Masticatory muscle	Muscle structure	Muscle fibre	Origin	Insertion	Function
Masseter	Quadrangular 2 heads: Superficial & Deep	Type 1 and 2	Zygomatic arch	Ramus of mandible	Elevates mandible Closes the jaw
Temporalis	Fan-shape	Type 1 and 2	Temporal fossa	Coronoid of mandible	Elevates mandible Closes the jaw
Medial pterygoid	Quadrangular shape 2 heads: Superficial & Deep	Type 1 and 2	Superficial head: maxilla Deep head: lateral pterygoid plate of sphenoid bone	Ramus of mandible	Elevates mandible Closes the jaw
Lateral pterygoid	Triangular shape 2 heads: Superior & Inferior	Horizontal and type 1	Superficial head: greater wing of the sphenoid Deep head: lateral pterygoid plate of sphenoid bone	Neck of mandible	Protracts the mandible Opens the jaw Moves the jaw from side to side

1.2.2 Staging of oropharyngeal cancer

HNSCCs are staged following the TNM staging system by the American Joint Committee on Cancer (AJCC) version 8. Approximately 40% present with early stage (I-II) disease and between 60-70% with locally advanced disease (2,14). Due to the superior prognosis of HPV positive OPC, the new version 8 includes an additional section for HPV positive and negative OPC. In a study of 253 patients, low tumour grade and a tumour arising either in the tonsil or base of tongue have an increased association with HPV positivity (15). HPV positive OPC typically present with cancers of a smaller tumour stage (64%), histologically poorly differentiated with basaloid features and more advanced nodal disease in the neck (69%), compared with HPV negative cancers (11,12). Table 1.1a and b highlight the differences between HPV p16 positive and negative OPC (16,17).

Table 1.1. Tumour and nodal classifications by p16 status in oropharynx cancer

	p16 positive	p16 negative
Tx	Removed from classification	Tumour cannot be assessed
T0	Removed from classification	Removed from classification
Tis	No primary identified	Carcinoma in situ
T1	≤ 2cm in maximum dimension	≤ 2cm in maximum dimension
T2	>2cm and ≤4cm in maximum dimension	>2cm and ≤4cm in maximum dimension
T3	>4cm or extends to lingual epiglottis	>4cm or extends to lingual epiglottis
T4	Moderately advanced disease. Invades larynx, extrinsic muscles of the tongue, medial/lateral pterygoids, mandible or beyond.	T4a Tumour invades the larynx, extrinsic muscle of tongue, medial pterygoid, hard palate, or mandible or beyond.
	T4b removed from classification	T4b Tumour invades lateral pterygoid muscle, pterygoid plates, lateral nasopharynx, or skull base or encases carotid artery.
	p16 positive	p16 negative
Nx	Regional lymph nodes cannot be assessed	
NO	No regional lymph nodes to be assessed	
N1	Metastasis in a single ipsilateral node ≤ 6cm	Metastasis in a single ipsilateral node ≤3cm, ENE negative
N2	Metastasis in contralateral or bilateral lymph nodes, none >6cm	2a metastasis in single ipsilateral node >3cm ≤6cm, ENE negative
		2b metastasis in multiple ipsilateral nodes ≤6cm, ENE negative
		2c metastasis in bilateral or contralateral nodes ≤6cm, ENE negative
N3	Metastasis in any lymph node (s) >6cm	3a metastasis in any node >6cm, ENE negative
		3b metastasis in any node and ENE positive

Abbreviations: ENE= Extra nodal extension

Table 1.2. Overall staging for oropharyngeal cancers

P16 positive							
T Stages	cN0	pN0	cN1	pN1	cN2	pN2	cN3
T0	NA	NA	I	I	II	II	III
T1	I	I	I	I	II	II	III
T2	I	I	I	I	II	II	III
T3	II	II	II	II	II	III	III
T4	III	II	III	II	III	III	III
Any M1 is stage IV c=clinical, p=pathological							

P16 negative				
T Staging	N0	N1	N2a,b,c	N3a,b
T1	I	III	IVa	IVb
T2	II	III	IVa	IVb
T3	III	III	IVa	IVb
T4a	IVa	IVa	IVa	IVb
T4b	IVb	IVb	IVb	IVb
Any M1 is stage IVC				

1.2.3 Treatment of oropharyngeal cancers

The choice of treatment for OPC depends on patient, tumour and treatment factors as summarised in Table 1.3. Treatment options are largely dependent on tumour stage and site. Single modality treatment with curative surgery or primary radiotherapy is the preferred option for patients with early stage (I and II) OPC. Retrospective analyses reported similar 3-year loco-regional and 5-year overall survival rates for surgery versus radiotherapy, but different adverse events (18). A Medline search from 1970 to 2000 of retrospective, non-randomised studies (n=13 to 101) compared surgery ±radiotherapy with radiotherapy ±neck dissection for patients diagnosed with SCC of the base of tongue or tonsil. Primary surgery or radiotherapy demonstrated similar loco-regional and 5-year survival rates for base of tongue SCC (60% v 69% and 49% v 52%) and for SCC tonsil (65% v 69% and 47% v 43%). Severe complications and functional deficit were significantly greater however, in those treated with surgery than radiotherapy (23% v 6%) (19). The treatment of locally advanced (III and IV) OPC is multi-modality. Options include: surgery followed by post-operative radiotherapy (if R1 resection or extra nodal extension) +/- chemotherapy, concurrent chemoradiotherapy or neoadjuvant chemotherapy followed by concurrent chemoradiotherapy. In an updated meta-analysis published in 2009 by Pignon et al. an absolute 5-year survival benefit of 6.5% in patients with stages III/IV disease treated with cisplatin compared with radiotherapy alone was observed (20). A modest 2% survival benefit in those treated with induction chemotherapy was also noted (20). Concurrent chemotherapy with high dose cisplatin is therefore the current non-surgical standard of care for locally advanced OPCs in those aged <70 years with a good performance status (21).

Table 1.3. Table of patient, tumour and treatment related factors which inform treatment choice for oropharyngeal cancer

Patient factors	Tumour factors	Treatment related factors
Age	Site	Functional status
WHO Performance status	Stage	Radiotherapy- target volume, total dose, fractionation
Co-morbidities	Grade or differentiation	Chemotherapy or immunotherapy
Smoking	Human Papilloma Virus	
Alcohol	Resection margins	
	Extra-nodal extension	

Approaches have been sought to de-intensify treatment with HPV positive OPC due to its favourable prognosis. The UK De-ESCALaTE HPV trial investigated the use of the epidermal growth factor inhibitor cetuximab, as an alternative to cisplatin to minimise toxicity, for low risk HPV positive cancers. The trial in 334 patients showed no reduction in treatment related toxicity with cetuximab, but inferior survival outcomes, with an higher number of patients developing distant metastases (22). In the randomised phase 2 ECOG 3311 trial, de-escalation of adjuvant treatment in low to intermediate risk locally advanced OPC following transoral surgery produced favourable 2 year PFS rates >90% when compared with standard adjuvant therapy (23). The phase 3 randomised multi-centre PATHOS trial is evaluating de-intensifying adjuvant treatment following transoral surgery, to reduce long term swallowing damage, in low risk HPV OPC (24).

1.3 Radiotherapy

Radiotherapy involves delivering maximum dose to the tumour to improve local control and reduce metastatic spread, whilst sparing normal tissues to minimise side effects. It can be given either as a primary treatment or following surgery to reduce local recurrence due to adverse pathological features (R1 resection or extra nodal extension). Radiotherapy has evolved from conventional 2D, to more conformal techniques including 3D and intensity modulated radiation therapy (IMRT). IMRT is an advanced form of radiotherapy which can deliver a highly conformal dose to the target volumes and reduce doses to neighbouring OAR by

modifying the beam shape and intensity (25). IMRT is the standard of care for patients undergoing primary and postoperative treatment for HNSCC in the UK, with the exception of T1 larynx and certain palliative regimens (26). Standard international primary radiotherapy is delivered to a total dose of 70 Gy in 2 Gy fractions over 7 weeks. In the UK many centres adopted altered fractionation regimens such as 65 to 66 Gy in 30 fractions over 6 weeks.

1.3.1 The therapeutic ratio

The goal of radiotherapy can be explained by the therapeutic ratio, which is the relationship between the probabilities of tumour control (TCP) and normal tissue complications (NTCP; Figure 1.3) (27) (28). The therapeutic ratio is the horizontal separation between the two sigmoid curves. The optimum treatment is one that can widen the separation with a, e.g. radiosensitiser to shift the TCP curve to the left with no additional toxicity. For example, a retrospective study of 86 OPC patients assessing response with quality-of-life (QOL) questionnaires showed an increase in the probability of dysphagia by 19% for every additional 10 Gy beyond 55 Gy, indicating a causal relationship between dose and OAR toxicity (29). There are several techniques to improve the therapeutic ratio such as using: different fractionation schedules, radiosensitisers, IMRT and image guided radiotherapy (IGRT).
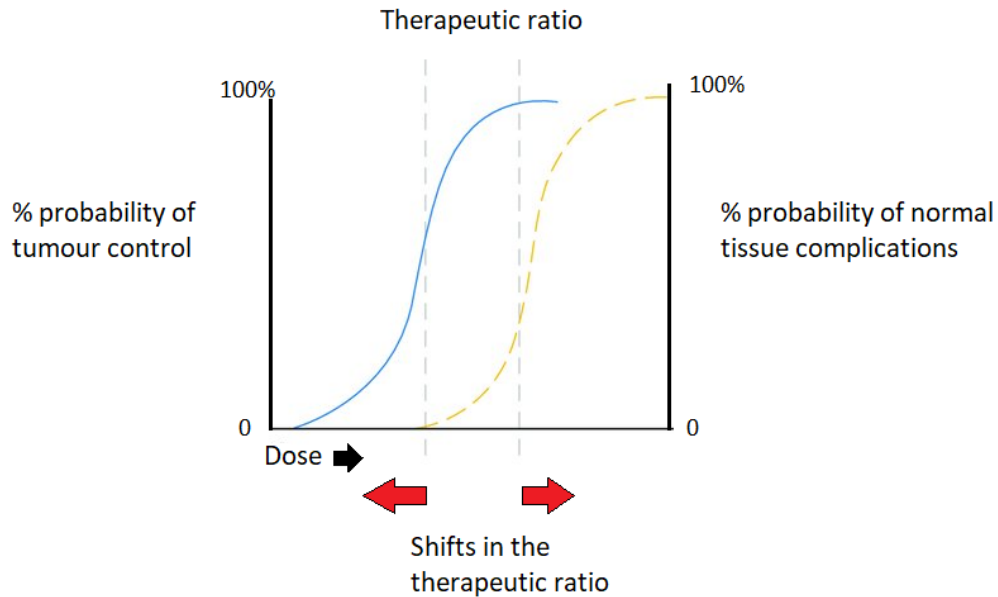
Figure 1.3. Sigmoid dose-response curves to illustrate the therapeutic ratio

1.3.2 Intensity-modulated radiotherapy

IMRT uses multiple beams of varying intensities to deliver a highly conformal dose. The ability to modulate the shape of the beam using multi leaf collimators (step and shoot or dynamic) increased the use of IMRT to accurately treat the complex shapes and anatomy of target volumes in the head and neck. IMRT improves the therapeutic ratio by reducing doses to normal tissues but is reliant upon accurate selection and delineation of structures (25). IMRT can be delivered either with a sequential or simultaneous integrated boost (SIB-IMRT) technique. SIB-IMRT enables multiple clinical target volumes (CTV) to be treated at the same time at different doses with the same number of fractions in a single treatment plan without increasing toxicity. It is the preferred approach as it improves conformity of the target volume, reduces overall treatment time by using a single plan and in some studies has shown to improve parotid gland sparing (30,31). In a study of 5 patients the mean dose to the ipsilateral parotid gland reduced by 21% (52.9 Gy to 41.9 Gy) when using SIB-IMRT versus sequential IMRT (32). Clinical outcomes with SIB-IMRT in locally advanced HNSCC are good. In a retrospective analysis of 108 patients treated with either definitive SIB-IMRT or adjuvant SIB-IMRT,3-year loco-regional control and overall survival rates were 64% v 78% and 52% v 57%,

respectively (33). Despite the improvement in dose delivery with SIB-IMRT, rates of acute G3 dysphagia (44%) and late xerostomia (42%) remain high (33).

An alternative form of IMRT is volumetric modulated arc therapy (VMAT). VMAT uses rotational arcs to deliver a highly conformal dose by varying the gantry speed, beam shape and dose rate. VMAT based plans deliver treatment over a shorter duration of time which can improve workload efficiency and patient compliance (34). Shorter treatment times with VMAT can also reduce the chance of intra-fractional errors due to organ motion. Use of VMAT to spare swallowing structures has been evaluated. In a comparative planning study of 20 patients treated with VMAT, mean dose to the swallowing OARs (upper pharyngeal constrictor muscles and supraglottic larynx) was reduced by 4 Gy and 5 Gy respectively, which induced a 9% reduction in RTOG Grade 2-4 dysphagia (35). Nithya et al. compared VMAT with IMRT step and shoot treatment plans for base of tongue tumours and found similar tumour coverage but greater parotid gland sparing with VMAT (36). The steep dose gradients of IMRT make it very sensitive to changes in patient and tumour position due to weight loss or tumour response, which may alter the planned dose distribution and result in a geographical miss. Adequate target volume definition, safety margins and methods of image-guided radiotherapy (IGRT) to track tumour progress during treatment are prerequisites to the success of IMRT.

1.3.3 Organs at risk and dose constraints

An OAR is defined as any normal healthy tissue located within the radiation field during radiotherapy. Establishing a threshold dose for OARs to be used in avoidance planning helps reduce toxicity and maintain normal function. Studies by Emani and QUANTEC (quantitative analysis of normal tissue effects in the clinic) suggested tolerance doses for some OARs but others such as the masticatory muscles are not included (37,38). To spare OARs, dose constraints need to be set when designing the IMRT plan. To help establish a dose constraint, OAR volumes need to be standardised and well defined. Contouring atlases, such as those from Radiation Therapy Oncology Group (RTOG), European Organization for Research and Treatment of Cancer (EORTC) and others developed as part of clinical trials,

may help improve the definition and standardisation of OAR volumes and reduce inter-observer variability in clinical practice (39). In a systematic review by Brouwer et al. head and neck atlases were shown to reduce inter-observer variability by clinicians following a consensus guideline (40). Current contouring atlases do not include all OARs such as the masticatory muscles. There is also significant variability in practice amongst and between institutions, which will impact on dose delivery to target volumes and OARs and in the development of multi-institutional clinical trials.

1.3.4 Delineation of target volumes and margins

The ICRU (International Commission of Radiation Units and Measurements) reports 50 and 62 are used as the gold standard for target volume definition (41).These reports include the gross tumour volume (GTV), clinical target volume (CTV) and planning target volume (PTV) (41). The GTV is the location and extent of the primary tumour and is defined by clinical examination and adequate imaging. Daisne et al. found the GTV defined on pre-treatment imaging underestimated the mucosal disease extent supporting the need for a CTV expansion (42). The CTV contains the GTV (tumour ±nodal regions) plus a margin to account for microscopic spread. In head and neck radiotherapy planning two CTV dose levels are used, a high dose CTV1 to the head and upper neck and a lower CTV2 to treat elective nodal levels at risk. The size of the CTV margin affects indirectly the volume of the OAR in the high dose radiation field (44). The PTV is a margin applied around the CTV in treatment planning systems. The PTV incorporates geometric uncertainties due to set up error and organ motion to ensure the CTV is adequately treated to maximise the probability of cure. The majority of loco-regional failures are within the high dose treatment volume. For example in a retrospective analysis of 56 HNSCC patients treated with IMRT, 59% failed within the GTV, 19% within the high dose GTV and 7% within the high dose CTV (43).

The method of determining the GTV-CTV expansion margin remains controversial and is an area of much debate. Traditionally an anatomical expansion from GTV to CTV was applied using prior knowledge of expected tumour spread, but the approach was shown to be more heterogeneous and to produce larger volumes compared with a geometric expansion. In a retrospective study by Caudell et al.,

the type of GTV-CTV expansion (anatomical or geometric) was compared with loco-regional failure rates. Poorer loco-regional control was found in those treated with an anatomical expansion of GTV-CTV, however the study was small and expansions varied between primary tumour sites and stages (44). In a recent study across four different centres, a geometric GTV-CTV expansion produced more uniform CTVs and greater consistency of target volumes irradiated between centres (45). International guidelines published by Gregoire et al. recommend a smaller 5+5 mm margin based on histopathology from a surgical series evaluating the extent of microscopic spread. The 5+5 margin is a 5 mm expansion from GTV to high dose CTV-P1 and a further 5 mm from CTV1 to the low dose CTV-P2 (46). These guidelines should improve consistency within and between centres. Results from the PROCAHN study showed using the international guidelines reduced inter-observer variability in particular for contouring the elective CTV2 nodal levels (47). Retrospective studies such as that by Corkum et al. evaluated removing the CTV-P2 margin in order to reduce further manual editing of natural barriers and thus workload. In this small study of 27 patients, 95% of the CTV-P2 received the expected dose of 56 Gy as was already covered within the CTV-P1 (48). Larger prospective studies are needed to validate this finding.

Patients with HNSCC treated with radiotherapy are estimated to lose between 5% to 15% of their initial weight (49,50). This weight loss can lead to incorrect fitting of immobilisation devices and as a consequence under and over dosage of the target and normal tissues. In a pilot study by Barker et al. (n=14), of which 9 patients received treatment for OPC with primary or combined chemo radiotherapy, a mean weight loss of 7.1% (range +5.2% to -13%) by the end of treatment was observed. The loss in weight significantly correlated with a reduction in external skin contours at the second cervical vertebrae due to tissue loss. The parotid glands also moved medially by 3 mm into the high dose radiation field, increasing the risk of xerostomia (49). In a study of 10 OPC patients, a similar mean weight loss of 7.5±3.1 kg by the end of treatment was reported by Bhide et al. as well as a 10% reduction in CTV in the first 2 weeks of treatment (50).

1.3.5 Image guided radiotherapy (IGRT)

IGRT is the daily imaging of patients in the treatment position to check for changes in the target volume, safety margins and patient position and make the necessary

adjustments relative to the initial planning CT. IGRT is a two-step process. The first step involves registration of the planning CT with the cone beam CT (CBCT). Images can then be manually reviewed on the CBCT and the necessary corrections made. The types of uncertainties or 'errors' that can occur during radiotherapy are either systematic or random (51). Systematic errors are uniform over the treatment and can occur between (inter-fraction) or during (intra-fraction) each fraction. Systematic errors are due to changes in target delineation or set up. Random errors are unpredictable, may occur intra-fractionally and are due to organ motion (44). Systematic errors shift the dose, whilst random errors blur the dose distribution (52). Some of the anatomical and dosimetric consequences that may occur during radiotherapy due to uncertainties are summarised in Tables 1.4 and 1.5.

Table 1.4. Anatomical changes during head and neck radiotherapy

Anatomical changes	Patients	IGRT	Ref
Parotid gland volume ↓ 4% per week CTV nodes: ↓10% by week 5	19	Daily CBCT	(53)
Right parotid gland volume ↓ 4.4 cm^3 by end week 5 Left parotid gland volume ↓4.5 cm^3 by end week 5	10	Weekly CBCT	(54)
Parotid gland volume ↓18 % with a 4.2mm medial shift GTV primary volume ↓ 63% GTV nodes volume ↓ 52%	20	Repeat CT	(55)
Ipsilateral parotid gland ↓29.7% Contralateral parotid gland↓ 28.4%	10	Weekly CBCT	(56)
Parotid gland volume ↓ 4.9 % per week, 0.85mm medial shift	15	Weekly CBCT	(56)
Parotid gland volume ↓ 13%	18	Daily CBCT	(57)

Table 1.5. Dosimetric changes during head and neck radiotherapy

Dosimetric changes	Patients	IGRT	Ref
Ipsilateral parotid gland ↑1.2 Gy	19	Daily CBCT	(53)
Right parotid gland ↑ V26 7.5% Left parotid gland ↑ V26 8.8%	10	Weekly CBCT	(54)
Parotid gland ↑ mean dose 20 % (5 Gy) Spinal cord ↑ D2 5 % (1.9 Gy) GTV primary mean dose stayed the same GTV nodes mean ↑ 0.6 Gy	20	Repeat CT	(55)
No change in dose to parotid gland, spinal cord, brainstem, larynx and oral cavity	10	Weekly CBCT	(56)
Parotid gland ↑mean dose by 2.6%	15	Weekly CT	(56)
Parotid gland mean dose increase by 0.9 Gy	18	Daily CBCT	(57)

Abbreviations: GTV=gross tumour volume; Gy=Gray; CTV= clinical target volume; CBCT= cone beam computed tomography

As shown in Tables 1.4 and 1.5, there is disparity in published data between the amount of volume change in OARs and at what time interval during radiotherapy changes occur. This variation can be explained by differences in the location of the tumours treated and the small sample sizes. Most studies showed anatomical changes occurred in the first 2 to 3 weeks of radiotherapy, which may suggest a potential need for adaptive re-planning at this time point. However, larger prospective studies are needed to validate this finding (50).

Anatomical changes can induce dosimetric changes. In the literature reviewed the dose distribution to the primary tumour remains robust, despite changes in anatomy. In a study of 20 patients by Nisha et al. a reduction in GTV tumour and nodal volume by 63% and 52% occurred midway through treatment with minimal increase in mean dose to the primary volume of 0.6 Gy (55). Castedot et al. found no increase in dose to the primary or nodal CTV when evaluated on weekly CBCT scans in 10 patients treated with chemoradiotherapy (59). In contrast, a study by Bhide et al. showed a 3.2% and 10% reduction in the primary CTV1 and CTV2 levels resulted in reduced mean doses to PTV1 and PTV2 of 2 Gy and 3.9 Gy, respectively (50). In a study by Barker et al. of 14 patients treated with chemoradiotherapy (n=1 IMRT, n=13 3D conventional) tumour and nodal GTV volumes reduced in an asymmetrical manner by 1.8% per day, but changes in dose were not evaluated (49).

Anatomical and dosimetric changes to the parotid glands during radiotherapy have been most widely studied. The parotid glands are highly radiosensitive and an increase in dose due to anatomical changes has been correlated with an increased risk of xerostomia and detrimental effect on QOL. The changes in dose to the parotid gland can be explained by a reduction in volume and medial shift into the high dose radiation field. Nisha et al. found that the parotid gland migrated medially by 4.2 mm by the end of treatment with a decrease in volume by 18% and increase in dose of 5 Gy (55). This was supported in a prospective observational study of 20 patients whereby a 15% decrease in the ipsilateral parotid gland volume increased the mean dose by 2.7 Gy by the end of the second week of radiotherapy. In a study of 15 patients by Castelli et al., the ipsilateral left parotid gland shrank by 4.9% per week over 7 weeks which resulted in an increase in the mean left parotid gland dose by 2.6 Gy (56). Fiorentino et al. showed no difference in the amount of parotid gland volume

reduction based on laterality, right and left parotid gland volumes reduced by 43.5% and 44% respectively (60). However, it is unclear in this study if both glands were treated in the high dose CTV-P1. Brouwer et al. in a systematic review of 51 original studies (n= 10 to 87) found the average (± standard deviation) volume of the parotid glands reduced by 26 ±11% with an average increase in mean dose of 2.2 Gy during radiotherapy. Volume loss was reported at different time points however and cohorts were small and retrospective (61). The exception to these studies was a study of 10 patients with OPC by Ho et al. where no increase in dose was found, despite a reduction in the volume of the parotid gland by 25% and weight loss of >10% (58). Dosimetric consequences due to changes in anatomy have been observed for other OARs such as the spinal cord and submandibular glands (SMG). In a prospective study of 10 patients with locally advanced OPC, a similar reduction in the volume of the SMGs occurred throughout radiotherapy by the end of week 6 (right 30% and left SMG 27%). Structures such as the spinal cord and brainstem however do not appear to change during treatment as shown by Jin et al (54).

1.3.6 Adaptive radiotherapy

Adaptive radiotherapy (ART) is an advanced form of IGRT that uses image guidance to edit a patient's treatment plan in response to changes during treatment. Over the last decade, the quality and type of image guidance approaches have evolved. CBCT scans are currently used as standard of care but have poor image quality. The poor image quality of CBCT scans is due to the quality of image capture equipment and the desire to keep additional radiation exposure low. The poor imaging of structures with CBCT scans makes calculation of the delivered dose inaccurate. Deformable image registration (DIR) is one method to improve the assessment of the cumulative doses received (56). DIR involves the overlay of images from the planning CT to match the CBCT. The method allows more accurate manual or automated contouring of the tumour or OAR volumes and assessment of the cumulative dose received (62). However, delineation of contours using CT imaging may be

challenging, due to the lack of soft tissue contrast and image degradation, secondary to dental or metal artefacts.

ART can occur either offline between treatments, online immediately before and in real time during treatment. In a study by Veiga et al. the concept of "dose of the day" was evaluated. In this proof-of-principle study, weekly offline dose calculations, by deforming the planning CT scan to match daily CTs was shown to be one step in the further development of adaptive radiotherapy (63).
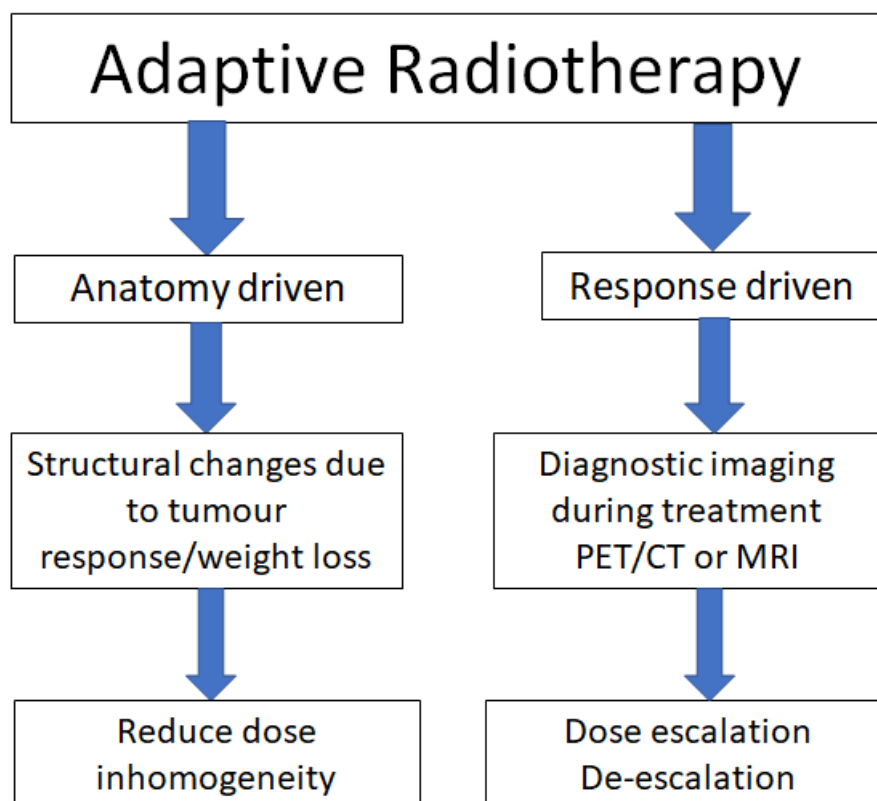
Figure 1.4. Schematic illustrating anatomy and response adapted radiotherapy.

1.3.6.1 Anatomy or response driven adaptive radiotherapy

Adaptive radiotherapy can be classified as anatomy or response adapted. The differences are illustrated in Figure 1.4. Anatomy driven ART involves re-imaging patients at different time points during radiotherapy in response to structural changes such as tumour response and weight loss. The ideal timing of when to re-image has yet to be confirmed, partly due to the heterogeneity between patients and lack of pre-treatment ART models to predict who may

benefit. Studies have attempted to develop predictive tools based on pre-treatment factors to select those who may benefit the most from ART using changes observed in the parotid gland. Broggi et al. in a study of 87 patients from four institutions found initial parotid volume and mean dose were the strongest pre-treatment factors to select those who may benefit from ART to reduce the risk of xerostomia (64). Brouwer et al. in a larger prospective cohort study of 113 patients confirmed these findings using a multivariable linear regression model to analyse ipsilateral and contralateral parotid glands. In this study, data were collected to develop (cohort A, n=113) and validate (cohort B, n=43) a model to select HNSCC patients who may benefit from ART, based on dose deviations to the ipsilateral and contralateral parotid glands and subsequent risk of xerostomia. In cohort A, ΔDmean was defined as the difference between the mean dose to the parotid gland on the post treatment CT at 6 weeks and the mean dose to the parotid gland on the planning CT. In cohort B, ΔDmean was defined as the mean accumulative dose to the parotid gland calculated on weekly CT scans and the mean dose to the parotid gland on the planning CT. A ΔDmean of >3 Gy above a threshold dose of ≥22.2 Gy which equated to 3-10% NTCP risk of xerostomia was used to select those who may benefit from ART. Between 18-20% of the patients in this study had a ΔDmean >3 Gy above the threshold of ≥22.2 Gy and were therefore selected for ART (65). Of the pre-treatment factors included in the model, such as BMI, stage, and tumour location, planned mean dose to the parotid gland on the pre-treatment scan, was the only significant factor to predict for benefit with ART (65).

Response adaptive radiotherapy uses diagnostic imaging such as CT, PET-CT or MRI to adapt the target volumes and doses based on treatment response. In a phase 1 dose escalation study of 21 patients using PET-CT guided ART, persistent FDG avid uptake was seen by week 2 of radiotherapy indicating areas of radio-resistant disease. Treatment adaption based on the second PET-CT scan acquired after fraction 8 identified a 41%, 18% and 14% reductions in the GTV, high dose CTV1 and PTV1 respectively enabling much smaller target areas to be treated and a reduction in dose to neighbouring OARs (66).

The pattern of recurrence in HNSCC remains largely in-field. In a study of 57 patients treated with OPC, a total of 31 local failures were reported of which 29

were in-field, defined as >50% within the GTV (67). The ability to dose escalate with response-ART to improve loco-regional control is currently being investigated in two phase II randomised trials, C-ART-2 (NCT01341535) and ARTFORCE (NCT01504815). Both trials compare PET-CT guided dose escalation with standard chemoradiotherapy with locoregional control (LRC) as the primary endpoint. ART is also being explored as a strategy to de-intensify treatment such as in HPV positive OPC with a view to reducing normal tissue toxicity. A phase II randomised trial, entitled PEARL (NCT03972072), in HPV positive OPC is evaluating the use of PET-CT at the end of week 2 to reduce target volumes secondary to tumour response and decrease dose to OARs.

1.3.6.2 Role of MRI in adaptive radiotherapy

MRI may be superior to CT in guiding the amount of tumour response in ART. In a study of five patients with HPV positive OPC, the average planned GTV dose visualised on MRI reduced at weeks 2, 4 and 6 by 44%, 90% and 100% respectively (68). The mean doses to the parotid gland (3.3 Gy) and swallowing apparatus were reduced and there was a decrease in the probability of developing > grade 2 dysphagia at 6 months by 11% (68). The soft tissue contrast and functional properties of MRI have been shown to better visualise the primary tumour and lymph nodes to help in the assessment during and following treatment. In a systematic review of 63 original papers (24 retrospective and 39 prospective), T1-T2 weighted MRI was superior to CT for detecting cervical nodal levels with a specificity of 0.81 v 0.72 and an ability to visualise nodes with a minimum axial diameter of 10 mm compared with 12 mm for CT (69). This finding was supported by a prospective study of 22 patients with locally advanced disease where diffusion weighted (DW)-MRI had better sensitivity (89% v 47%), specificity (97% v 42%) and accuracy (96% v 82%) for the detection of lymph nodes compared to CT/ T1-T2 weighted MRI (70). DW-MRI yields less false positives compared with CT or PET CT when evaluating persistent disease or new lymph nodes <10 mm (n=26) (71). The ability, however, of MRI to correctly detect the extent of microscopic disease has been questioned. In a study of 8 patients with locally advanced OPC, fiducial markers were placed at the edges of the target volume. MRI scans performed at weeks

3 and 6 demonstrated a greater reduction in the target volume relative to displacement of the markers, implying the need for further treatment to eradicate microscopic disease (72).

The challenges of MR image guided radiotherapy (MRIgRT) historically are due to scanning patients in a separate scanner. This may lead to errors in position and registration. To try and overcome this issue, several hybrid devices have been developed that use different strengths and orientations of a magnetic field combined with a linear accelerator. The hybrid devices enable patients to be imaged pre-, during and post-treatment in real time with the 'beam on' to track moving objects and visualise anatomical or tumour response, without the need to use a separate MR scanner or expose patients to additional ionizing radiation. The first hybrid device to be developed in 2014, was the ViewRay MRIdian, which instead of a linear accelerator, combined a magnet with three CO-60 heads. Since then the ViewRay MRIdian MR-Linac (MRL) was developed and is the commonest MRL worldwide and the first to be CE marked. This uses a 0.35T magnet with a 6MV linear accelerator (73). Similar to this machine, the Elekta MRL which is also being used in clinical practice combines a higher strength 1.5T magnet with an Elekta linear accelerator.

Frequent imaging is important in HPV positive OPC as large anatomical changes due to tumour response are expected which may result in excess dose to OARs. In a study of 120 patients with node positive nasopharyngeal or HPV positive OPC, 38% required a re-plan by fraction 22 (74). The MRL can adapt to tumour response and anatomical change. As previously discussed, HNSCC patients can lose up to 15% of their weight and radiosensitive OARs such as the parotid glands can reduce in volume and displace, increasing the risk of xerostomia. The MRL has the potential to map changes in OARs such as the mastication muscles, salivary glands and swallowing apparatus which are not as well defined on CT so adjustments in dose can be made (75).

MR guided ART may help to dose de-escalate due to tumour response in HPV positive OPC and dose escalate in HPV negative poor responders. The MR-ADAPTOR phase II study is investigating the potential of weekly MRI to guide adaptation of the high dose regions based on tumour response, in low risk HPV positive OPC and compare LRC with standard non-adapted IMRT (76). Plan optimisation on an MRI can be time consuming, uncomfortable and

claustrophobic for patients. Integration of the MRL into the radiotherapy workflow may reduce time on the scanner by online adaptation, but is potentially limited by the increase in workload of manually editing plans and patient compliance (77).

1.3.7 Side-effects of radiotherapy for oropharyngeal cancer

Radiotherapy side-effects of particular importance in OPC are xerostomia, dysphagia and trismus, which are recognised dose limiting toxicities that have a significant negative impact on patients' QOL (74). Radiotherapy induced xerostomia is caused by damage to the salivary acinar cells resulting in reduced saliva production. The reduction in saliva production is greatest from the onset of radiotherapy to 3 months post completion. An estimated 50%- 60% reduction in saliva has been reported in the first week post radiotherapy (79). In a retrospective QOL study of 39 long term HNSCC survivors treated with conventional radiotherapy across 30 years, 64% of patients reported G2 or G3 xerostomia using LENT-SOMA scores and visual analog scales (VAS) (80). Xerostomia will lead to problems with chewing, swallowing and dental caries. Radiotherapy related dysphagia is caused by inflammation and fibrosis to the larynx and pharyngeal constrictor muscles. Patients who develop dysphagia are at an increased risk of aspiration and needing enteral feeding replacement. In a study of 529 patients by Langendijk et al. the prevalence of RTOG ≥G2 dysphagia at 6,12,18 and 24 months following radiotherapy completion was 23.1%, 15.6%, 13.2% and 13.2% respectively. Prognostic factors for developing dysphagia were identified including advanced tumour stage (T3/4), bilateral neck irradiation, weight loss at baseline, primary OPC or nasopharyngeal cancer and use of accelerated and chemoradiotherapy (81). Similar results were shown in the large DAHANCA 6 and 7 randomised studies. The prevalence of acute severe dysphagia (G3 or G4) in HNSCC treated with accelerated versus conventional radiotherapy was 47% versus 38% respectively. Chronic dysphagia above grade 0 was 46%, 32%, 29%, 24%, 23% at years 1,2,3,4 and 5 respectively with no difference between the conventional and accelerated arms (82). Both T stage and tumour site were identified as

potential risk factors for developing dysphagia with the highest prevalence of Grade 3 or 4 dysphagia observed in oral cancer (20% at 5 years) (82).

Organ sparing studies to help reduce long term toxicity for good prognostic cancers have therefore been evaluated (83). In the PARSPORT trial there was an absolute reduction in ≥ grade 2 xerostomia in the parotid sparing IMRT group compared with conventional radiotherapy, both at 12 months (35%) and 2 years (54%) (84). An improvement in salivary flow in the parotid sparing IMRT group was observed at both time points, resulting in improved patient reported QOL. In a study of 186 patients, a reduction in mean doses with swallowing-sparing IMRT was confirmed with an NTCP model, (mean change in dysphagia 4.9%; 22.6% v 27.5%) (85). Using patient reported outcome measures studies showed an improvement in dysphagia but swallowing function did not return to baseline by 12 months (86,87). This finding highlights the ability of some patients to re-adjust to a new 'functioning level' following radiotherapy.

Radiotherapy is also a major cause of trismus. The natural progression of trismus shows a peak at 6 months with some recovery, but incomplete recovery by 1 year. The incidence of trismus following radiotherapy ranges between 5% and 38% (88,89). The large variability in published data is due to lack of a uniform criteria to define trismus, heterogeneity in patients and retrospective analysis (90). In a study of 75 patients treated for OPC, the incidence of trismus pre-treatment was 9%, at 6 months 38% and at 1 year 28% (91). Patients with cancers of the tonsil were more prone to developing trismus and EORTC QLQ H&N 35 scores indicated greater problems in xerostomia, pain and dysphagia in those with trismus up to 1 year post treatment (91).

Trismus is defined as a maximal inter-incisor distance of ≤35 mm (92) and is caused by progressive impaired function of the four sets of masticatory muscles as described in Section 1.3.1. Trismus has a heterogeneous pathology including smoking, tumour stage and location, surgery, chemotherapy and radiotherapy. In a recent study published in 2019, factors associated with the development of trismus included: increasing age, advanced tumour stage/extension (T3/4), tumour location near the TMJ and masticatory muscles (mandible, maxilla, cheek, salivary glands, oropharynx, unknown primary), free soft-tissue flap reconstruction after surgery, re-irradiation and chemotherapy (93).

Radiation induced trismus is caused by fibrosis and atrophy of the masticatory muscles' secondary to ischaemia. In a study of 139 patients treated with radiotherapy alone, the prevalence of trismus 16 months post IMRT was 24% (94). The incidence of trismus is increased in patients receiving more than one treatment modality. In a retrospective study of 259 patients with HNSCC there was a significant lower incidence of trismus in those treated with surgery alone (17%) versus either chemoradiotherapy (29%) or surgery plus post-operative radiotherapy (34%) (95). Reduced QOL scores have been reported in patients receiving tri-modality as opposed to single modality treatment, due to an increased pain, dysphagia and trismus (96).

Trismus can result in impaired speech, oral intake and dental hygiene, as well as psychological difficulties such as low self-esteem, depression and suicidal ideations (97). In a cross-sectional study of 78 patients treated for HNSCC who developed radiotherapy related trismus, all recorded a significantly higher hospital anxiety and depression score (HADS). Health related quality of life scores (HRQOL) based on symptoms such as pain and social functioning were also much lower (98). Bensadown et al. in a systematic review identified five papers which all confirmed the benefit of early "gentle passive motion" after completing radiotherapy (99). An example of gentle passive motion is jaw opening exercises with a therabite. A therabite is an aid to jaw opening exercises that has been evaluated to reduce the development of trismus, when used before, during and after radiotherapy. In a feasibility study of 71 HNSCC patients with a sense of jaw tightening at baseline (defined as trismus), mouth opening measurements were measured at different time points following a trial of proactive jaw exercises using a therabite or wooden spatula. Mouth opening measurements improved by 6 months post radiotherapy, irrespective of the type of exercise intervention used. Proactive jaw exercises are a useful intervention to help reduce the development of trismus in HNSCC, but are reliant upon adequate patient compliance (97).

1.4 Strategies to minimise long term toxicities and improve patient outcome in oropharyngeal cancer

Strategies are needed to reduce treatment related morbidity in good prognostic HPV positive OPC, where long term QOL should be optimised as shown in Table 1.6.

Table 1.6. Some strategies to reduce long term toxicities in oropharynx cancer

Strategy	Rationale	Section
To define anatomically what constitutes the OAR and develop a tolerance dose for avoidance planning	To define for the first time a tolerance dose for the muscles of mastication to reduce radiation related trismus	1.4.1
To improve the consistency of clinician contouring for OARs in the treatment of OPC and evaluate the roles of CT and MRI in auto-contouring	To standardise volumes and reduce inter-observer variability with contouring atlases and auto-contouring models to improve the accuracy of avoidance planning and dose delivery and reduce toxicity	1.4.2
To evaluate the potential benefit of protons in the treatment of OPC	To exploit the sharp fall off and lack of exit dose with protons to reduce doses to OARs	1.4.3
To evaluate the role of patient and public involvement to aid development of the UK's first proton trial in HNSCC	To understand patients' views and perceptions of the trial's primary objectives and study logistics to help shape and inform the trial design	1.4.4
To evaluate the benefit of non-invasive oxygen enhanced MR imaging in head and neck treatment planning	To understand the predictive and prognostic implications of tumour hypoxia to aid adaptive planning and improve clinical outcome	1.4.5

1.4.1 Tolerance dose and contouring atlas for the masticatory muscles

The masticatory muscles are not routinely contoured as OARs and therefore no tolerance doses have been established in clinical practice. This is partly due to the variability in the literature, which is summarised in Table 1.7. The medial pterygoid and masseter muscles are most strongly associated with trismus. Gebre Medhin et al. in a retrospective analysis of 139 patients suggested a mean threshold dose of 60 Gy to the ipsilateral masseter. The prospective ARTSCAN study defined V40-V60 Gy as a tolerance dose to the ipsilateral masseter (100). In a small retrospective analysis of 22 patients, a mean dose to

the lateral pterygoid of <42 Gy was suggested. This is much less compared with suggested doses to the adjacent medial pterygoid, where constraints in excess of 55 Gy have been reported (101,102).

Table 1.7 Literature summarising tolerance doses for the masticatory muscles

Study design	N	Masticatory muscle	Tolerance dose	Ref
Retrospective	139	Ipsilateral masseter	Mean 60 Gy	(94)
Retrospective	421	Ipsilateral medial pterygoid	V68 Gy	(103)
Retrospective	40	Ipsilateral medial pterygoid/ masseter	Mean ≥55 Gy	(102)
Prospective	124	Ipsilateral masseter	V40-60 Gy	(100)
Retrospective	55	Masseter	Mean, V20, 40, 60 Gy	(104)
Retrospective	56	Pterygoid muscles	10 Gy ↑ in dose >40 Gy ↑ probability of trismus by 24%	(105)
Retrospective	22	Lateral pterygoid	Mean <42 Gy	(101)

One approach to establish an avoidance dose for the muscles of mastication, which may be more sensitive than considering the muscles as separate entities, would be to investigate the relationship between dose to a combination of muscles defined as a block with the development of trismus (Figure 1.5). The outline of the block is a composition of all the structures. The pink area highlights the interface between the structures composed of fat, nerves and fascia.
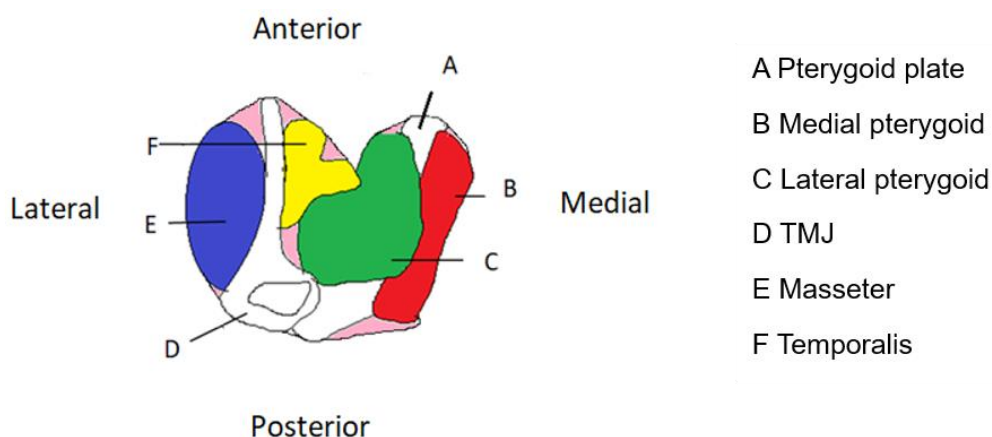
Figure 1.5. Diagram illustrating the concept of the block structure.

Using a block approach has the potential to take account of areas in the interface of muscles, such as fat, nerves and fascia of which little is known about the dose-response relationship. It may also help improve clinician inter-

and intra-observer variability, reduce delineation time and be incorporated into the adaptive radiotherapy algorithm. Development of a novel muscle of mastication atlas would also help standardise contours, improve reproducibility and could be used as an educational tool for trainees and in clinical trials.

1.4.2 Auto-contouring models in head and neck radiotherapy planning

OAR contouring in head and neck radiotherapy planning is labour intensive and subject to observer variability due to the number and complexity of the structures involved. Auto-contouring tools have been shown to reduce contouring time of target volumes and OARs (106). In a systematic review of 10 original studies investigating auto-segmentation in targets and OARs for HNSCC IMRT, time savings of 31% to 59% were observed (107). Stapleford et al. recorded an average time saving of 11.5 minutes per patient using atlas based auto contours for lymph nodes in the neck (108). This was similar to that reported by Teguh et al. where ABAS auto-segmentation of 30 structures (bilateral neck levels I-V, spinal cord, bilateral parotid glands, SMGs, mastication and swallowing muscles) took on average 7 minutes per patient, (n=10). The auto contours still required manual editing which took up to 66 minutes per patient. This was still a time saving of 59% compared with manual contours (109).

There is considerable inter-observer variability in target volume and OAR definition which will affect the doses delivered. Hong et al. in a study across 20 institutions reported substantial variation in contouring of levels I, V and the contralateral neck with a range of CTV volumes, 37–676 cm^3 (110). Loo et al. in a study of 10 patients with OPC treated with parotid sparing IMRT reported significant variation in parotid gland contours amongst 4 oncologists and 3 radiologists when compared with the initial clinical plan. Of the 70 study contours reviewed, 46% were sufficiently different from the initial treatment plan highlighting the interobserver variability amongst clinicians (111). Geets et al. compared interobserver variation of the parotid glands and spinal cord when delineated on CT and MRI. Interobserver variation was similar for the parotid glands but significantly reduced when using MR-fusion for the spinal cord (112). Auto-contouring tools have been largely studied to reduce inter-observer

variability in contouring OARs and lymph node levels. ABAS auto-contours for the parotid glands improved mean DICE and DTA 0.8 and 2.4 mm compared with manual contours (109). Walker et al. in a prospective study using a commercial auto-contouring software, reported mean DSC values of 0.97, 0.90 and 0.89 for the brainstem, spinal cord and parotid glands (n=40) (113).

1.4.2.1 Types of auto-contouring models

Auto contouring models can be classified as atlas or model based. Atlas based models such as ABAS (Elekta) and SPICE (Philips), use expert contours from a "reference library" and apply them to a new patient image, to fit the patient's anatomy. Contours are propagated from the atlas and deformably registered on the patient image data set. Atlas based models are reliant upon the accuracy of the expert contours sometimes termed the 'ground truth'. Contours can be developed from single or multiple data sets to refine the image. The accuracy of the 'ground truth' is determined by a single person's contours and is thus subject to personal perception and uncertainty. Comparing multiple contours of the same structure from different clinicians to produce a 'consensus contour' may help reduce uncertainty and bias. Atlas-based models are unable to adapt to changes between the reference library and patient contours and are strongly dependent on the accuracy of the registration and anatomical similarity between the atlas and patient image. In a study by Ayyalusamy et al. of 10 head and neck patients, better matching of anatomy resulted in better segmentation (114).

Model or machine-based learning uses prior knowledge from expert contours to extract features of the structures, which are then included in the model. Machine learning is a type of artificial intelligence which recalls the characteristic shapes of structures that have previously been manually delineated. Machine learning can be classified as random forest and deep learning and is sensitive to the quality of the data imported (115,116). Deep Learning Contouring (DLC) Expert by Mirada medical is a new advanced form of machine-based learning. Studies validating the use of DLC Expert in clinical practice, have shown the similarity of auto-contours to manual contours. DLC

Expert uses 'neural networks' to learn about features and tasks given to it directly from a set of data, without the need to manually identify specific features in an image. Like the method of neurons receiving and passing signals in the human body, DLC Expert uses the input data it receives to predict a particular output. The information is passed via several "deep" artificial layers and is very sensitive to the quality of the data imported. The neural network is illustrated in Figure 1.6.
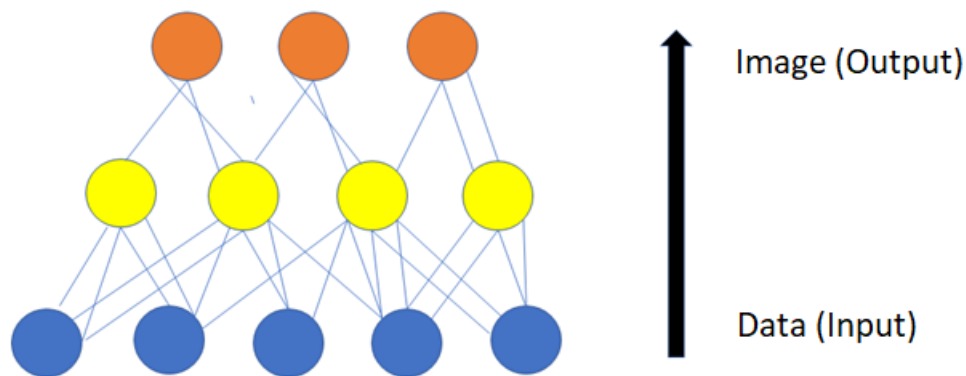
Figure 1.6. Artificial neural network showing information passing between the input to the projected output layer [4]

Deep learning models can be trained on the maximum amount of data available and do not exclude differences in anatomy between image sets. Unlike atlas-based models, deep learning approaches are not affected by artefacts such as metal. This advantage creates a diverse model, with the ability to adapt to changes in anatomy secondary to tumour volume, shape and treatment position. Atlas based models are very labour intensive and unlike deep learning models cannot adapt for changes in anatomy between the atlas and the target image. The main differences between atlas and model-based auto-contouring models are summarised in Table 1.8 (117).

---

[4] Taken from (305)

Table 1.8. Atlas versus machine-based approaches to auto-contouring

	Atlas based	Machine based
Registration of images	Yes	No
Misalignment of atlas on target image due to variable anatomy	Yes	No
Trained on large diverse data sets	No	Yes
Adapts for artefacts i.e. metal	No	Yes
Performs better for smaller structures	No	Yes

A recent publication by van Dijk et al. compared atlas (ABAS) and machine-based (DLC Expert) learning models in 693 head and neck patients over a 9-year period. DLC Expert was preferred over ABAS for all OARs except the glottis. DLC Expert showed superiority for all glandular structures and vessels including the carotid arteries. Difference in doses between the manual contours and the auto contours for all OARs were reduced as was inter-observer variability (117). ABAS contours were more prone to obvious errors and required frequent corrections. Using the Turing test, it was difficult to distinguish between human and DLC Expert contours with similar corrections observed between the two groups, 7% v 9% respectively. DLC Expert performed exceedingly well for the parotid, thyroid gland, cricoid inlet and pharyngeal constrictor muscles (117).

Delineation accuracy is a potential limitation of auto-contouring tools. Studies reported insensitivity of auto-contouring models to recognise tissue boundaries leading to inaccurate delineation and potential under and over dosage of target volumes and OARs. In a study by Thomson et al. SPICE contours were compared with manual contours from five clinicians. SPICE contours were acceptable for the parotid and submandibular glands (median DSC 0.79/0.80) but were inferior to manual contours for the larynx, pharyngeal constrictor muscles and cochleae (118). Voet et al. reported a reduction in the contoured CTV for nodal levels using auto-contours and subsequent under-dosage (119). Walker et al. reported over-dosage of the optic chiasm and subsequent blindness, due to an inability of the auto-contouring model to correctly define the optic chiasm (113). Tumour stage may also impact on delineation accuracy of

auto-contouring tools with a lower DSC reported for T4a compared with T1-T3 HNSCCs (120).

There are currently no published auto-contouring tools in the post-operative setting in HNSCC to reduce variation and time. Inter-observer variation can be increased in post-operative treatment plans due to the absence of a GTV and the effect of anatomical changes following surgery leading to larger treatment volumes (121). In a study in non-small cell lung cancer, however of 418 patients, deep learning assisted contours significantly reduced variation compared with manual contours on CTV datasets, with an increase in DSC (0.75 v 0.72, p<0.001) and reduction in mean DTA (2.97 v 3.07 mm, p<0.001) (122). Auto-contouring tools are an attractive addition to the adaptive radiotherapy workflow to reduce time and standardise volumes.

1.4.2.2 MR versus CT in improving image quality of auto-contouring models

The superior soft tissue contrast of MRI has the potential to increase contouring accuracy and volume definition (123). In patients with nasopharyngeal cancer, MRI compared with CT is able to detect intracranial involvement (41%) (n=258) (124). Co-registration of MRI with CT is the international standard of care for base of skull radiotherapy planning due to optimal visualisation of perineural/ intracranial spread (70, 76). The scatter radiation from dental artefacts that affects the accuracy of volumes contoured in CT planning is reduced with MRI. Furthermore, MRI can visualise the lower neck better than CT which is affected by shoulder artefacts (126).

Several studies have evaluated the size of the GTV contoured on MRI compared with CT. In a recent study of 10 patients with SCC of the tongue, tumour and nodal GTV contours were 80% and 70% larger than the GTV contours on the baseline CT (127). The larger volumes contoured with MRI may reduce the likelihood of geographical miss due to organ motion but have the potential to overdose OARs. In contrast, other studies reported smaller GTVs contoured on MRI due to better visualisation compared with CT (128,129). In a small study of 10 patients with OPC, inter-observer variability was also reduced with MRI compared with CT or PET CT (130). Current international guidelines

recommend the use of MRI for contouring tumours of the oral cavity, oropharynx, nasopharynx and OARs including spinal cord, brainstem, optic apparatus and parotid glands (40).

MRI auto-contouring software tools have been evaluated in studies focusing on GTV delineation. In a study of 10 patients with OPC and laryngeal cancer, MRI auto contouring software tools demonstrated a time saving of 7 minutes per patient to contour the GTV using auto versus manual contours (mean 45 versus 88 seconds per axial slice) (131). The mean distance similarity coefficient (DSC) between the gold standard and the auto contours was 0.79 with a volume concordance of 86.5% v 74.1%. This finding was further supported in a retrospective study of 1,021 patients with nasopharyngeal cancer where an MR AI auto contouring tool for the GTV reduced contouring time by 39.4% (132). In a study of 14 head and neck patients by Wardman et al., atlas-based auto-segmentation tools using T1-weighted MRI were compared with standard CT for OARs and lymph node levels. T1-weighted MRI automatic segmentation was significantly better for orbits, parotid glands, brainstem, and lymph nodes, whilst the spinal cord was better delineated, using CT auto-segmentation (133). The optimum sequence and treatment position required to develop an MRI based auto-contouring tool is yet to be agreed, however, and no validated MRI auto-contouring tools are available for clinical practice.

1.4.3 Role of proton beam therapy in the treatment of oropharyngeal cancer

Proton beam therapy is a maturing treatment in the management of HNSCCs that exploits the unique physical properties and depth dose characteristics of proton beams. Protons deposit their maximum dose in a defined area called the Bragg Peak. The Bragg peak is situated at the distal end of the beam. Unlike with photons, there is minimal to no exit dose. Normal tissues beyond the treated tumour are therefore spared. Protons are an exciting potential treatment for OPC to help reduce dose to OARs and minimise long term toxicities, in those with good prognostic HPV positive cancers.

IMRT   IMPT
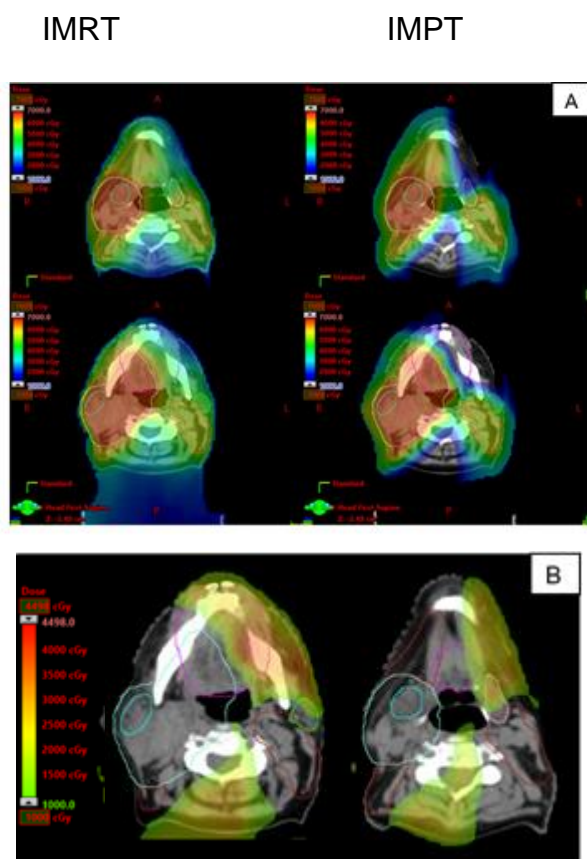
Figure 1.7. shows the sparing effect of protons by comparing the dose distributions of an IMRT (left) versus an IMPT (right) plan in the same patient with oropharyngeal cancer.

The difference in dose between the IMRT and IMPT plans is shown in figure B. Only differences ≥10Gy are shown.

Proton beam therapy is limited in clinical practice due to cost, the number of centres worldwide and a lack of high-quality published data defining the benefit of protons over IMRT. Patient selection is important to determine who will benefit most from proton beam therapy (71). Most studies comparing the toxicities of Intensity Modulated Proton Therapy (IMPT) and IMRT are based on small retrospective analysis as shown in Table 1.9.

Table 1.9. Studies comparing IMPT versus IMRT in the treatment of HNSCC

Design	Size	Toxicity	Survival	Ref
Retrospective Oropharyngeal 2010-2014	IMPT n=50 IMRT n=10	Median length gastrostomy insertion IMPT v IMRT 2.8 v 4.8 months Weight loss >12 months IMPT v IMRT 8% v 25%	IMPT OS*94.3% IMRT OS*89.3% *3 years	(134)
Prospective Oropharyngeal 2010-2014	IMPT n=50	G3 mucositis 58% G3 dermatitis 46% Acute G3 dysphagia 24% Late G3 dysphagia 12% Gastrostomy insertion in 22%	LRC 92%* OS 94.5%* PFS 88.6%* *2 years	(135)
Retrospective Salivary gland or cutaneous SCC 2011-2014	Passive Scattering n= 18 IMRT n=23	Passive Scattering v IMRT ≥G2 acute dysgeusia 5.6 v 65.2% ≥ G2 acute mucositis 16.7 v 52.2% ≥ G2 acute nausea 11.1 v 56.5%		(136)
Retrospective Nasopharynx or paranasal sinus cancer 2012-2014	IMPT n=14 IMRT n=26	IMPT v IMRT Rate of gastrostomy dependence IMPT: on completion and 3 months n=0 gastrostomy dependent (OR 0.03; 0.11 respectively)		(137)

Abbreviations: IMPT= Intensity modulated proton therapy, IMRT= Intensity modulated radiotherapy, LRC= locoregional control, OS= overall survival; PFS= progression free survival; G3= grade 3

Proton beam therapy can spare the anterior part of the tongue and major salivary glands leading to reduce rates of mucositis and xerostomia (138). IMPT should be advantageous in the treatment of oropharyngeal cancers due to improving the ability to shape doses to simultaneously treat large primary tumours and cover the neck, whilst maintaining normal tissue tolerance. Gunn et al. performed a study in a cohort of patients of whom 98% had stage III/IV disease, 98% were p16 positive and half were non-smokers. Although a small study (n=50), grade 3 toxicities were low and feeding tube placements were less than with comparative IMRT studies (24% v 47%) (135). In a study by Sio et al., patient reported outcomes were compared between 35 patients treated with IMPT and 46 treated with IMRT. There were no significant differences when comparing the acute and chronic side effects in the two groups. In the IMPT group acute toxicities were reduced in the initial three months following completion of treatment (139). IMPT is a favourable option in the treatment of unilateral early tonsil and salivary gland cancers due to the increase in OAR sparing (140). The prevalence of swallowing related side-effects 12 months following treatment with IMPT and IMRT was compared by Jakobi et al., as summarised in Table 1.10 (141). The incidence of grade 2 xerostomia was halved using IMPT (23 % v 46%). Dysphagia, permanent gastrostomy insertion and osteoradionecrosis of the jaw were also reduced with IMPT (Table 1.10). This finding was further supported by a study of 61 patients where grade 3 radiation oesophagitis was reported in 4% treated with IMPT versus 12% with IMRT. Survival outcomes were similar amongst the two groups (142).

Table 1.10. Comparison of grade >2 toxicities 12 months post IMPT versus IMRT[5]

	IMRT	IMPT
	Grade > 2 toxicities 12 months	
Xerostomia	46%	23%
Dysphagia	23%	18%
Permanent gastrostomy by 1 year	9%	2%
Osteoradionecrosis of the jaw	8%	2%

Bagley et al. reported QOL scores related to xerostomia in 69 patients treated with IMPT. In this study, xerostomia related QOL scores were reported to be worse six

weeks following completion of radiotherapy with a significant improvement noted at ten weeks and within the first year. This may be due to improved parotid gland sparing with IMPT compared with other radiation techniques (143).

1.4.3.1 Dosimetric and physical properties of protons

Figure 1.8 illustrates the finite beam range of protons known as the Bragg peak. The difference in the beam path and dose distribution compared to photons is highlighted. The reduction in exit dose can reduce the integral dose to the OARs. Such properties may enable dose escalation to the target without exceeding dose to adjacent normal healthy tissues. Protons are considered to be slightly more effective in killing cancer cells than photons. In clinical practice a relative biological effectiveness (RBE) of 1.1 irrespective of tumour type is assumed for protons (144). In clinical practice there is uncertainty as to the value of the RBE at the end of the proton beam range which, if higher, could increase toxicities such as temporal lobe necrosis (145,146).
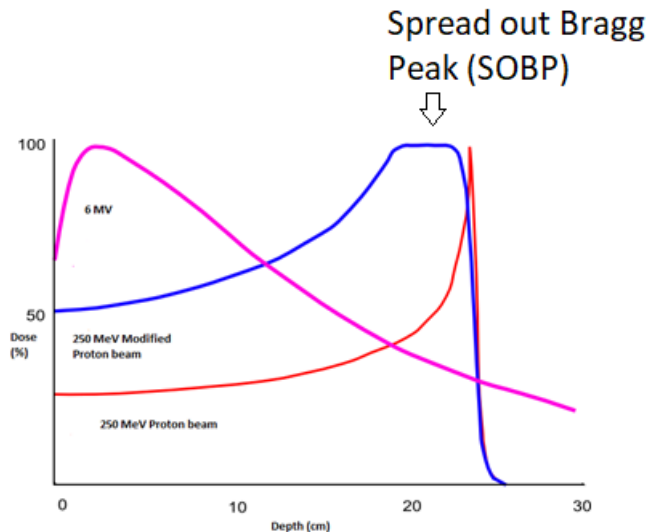
Figure 1.8. Dose-depth distribution of a 6MV photon (lilac line), a modified 250 MeV proton beam (blue line) and a mono-energetic 250 MeV proton beam. The Bragg peak can be seen at the distal end of the mono-energetic proton beam.

1.4.3.2 Mode of delivery of protons

Proton beam therapy is delivered by passive scattering or pencil beam scanning. Passive scattering places a scattering material into the proton path to produce a large field. Compensators and collimators can be used to shape the scattered field both laterally and distally. These may be adjusted to ensure the target is covered despite range and set up uncertainties (discussed in Section 1.4.3.3). Protons have a low entrance dose compared to the Bragg peak. As tumours vary in size and position the proton beam can be modified ensuring a homogeneous dose is delivered to cover the target area. To ensure coverage in depth a spread-out Bragg peak (SOBP) is created, delivering protons at different energies and consequently different ranges. The SOBP can also increase the entrance dose especially for superficial tumours resulting in an increase in skin toxicity. Despite the clinical application of a constant RBE of 1.1, the RBE is also thought to vary with the position of the SOBP. In a study by Hojo et al. the RBE was found to be higher at the distal end of the Bragg peak compared to the proximal part (147). There are still large uncertainties as to the exact RBE value that should be used, dependent on the range and position of the Bragg peak.

Pencil beam scanning uses magnetic fields to deflect and steer the proton beam. The beam is delivered as an array of spots. The size of the spot can be measured in σ. The size of the spot will affect the width of the proton beam penumbra and therefore the amount of tissue treated. The proximal conformation of pencil beam scanning results in a more conformal dose to the target and allows greater skin sparing, compared with passive scattering. The position and depth of the beam can be controlled by steering the direction of the spot and changing its energy. Plans may be optimised using single (SFO) or multi-field optimisation (MFO). In SFO, each spot is individually optimised and representative of the overall plan. In MFO all spots are simultaneously optimised, which can result in a more conformal dose distribution. MFO however is more sensitive to uncertainties due to the larger number of in-field dose gradients compared with SFO. MFO plans provide better organ sparing than SFO but are more sensitive to the effect of changes in set up and anatomy. The position of the spot pattern will determine where the dose is delivered. Incorrect spot position risks under- or over-dosing the tumour and OARs

respectively. The size and position of the spot is determined by the proton range, which is sensitive to inter and intra-fractional motion. Inter and intra-fractional motion during treatment can lead to misplacement of spots and inaccurate treatment delivery.

1.4.3.3 Uncertainties with proton beam therapy

Proton beam treatment planning is not without uncertainty. As with IMRT the ideal dose distribution is affected by uncertainties caused at the start of the planning process, such as target volume delineation. Additional challenges of proton beam delivery are range/stopping power uncertainty and set up and organ motion. Immobilisation, image guidance and safety margins can help mitigate set up errors, but daily variations are inevitable. Incorrect positioning of proton Bragg peaks and the effect on the overall dose distribution can have detrimental clinical outcomes. The proton beam range is affected by several variables (Figure 1.9).
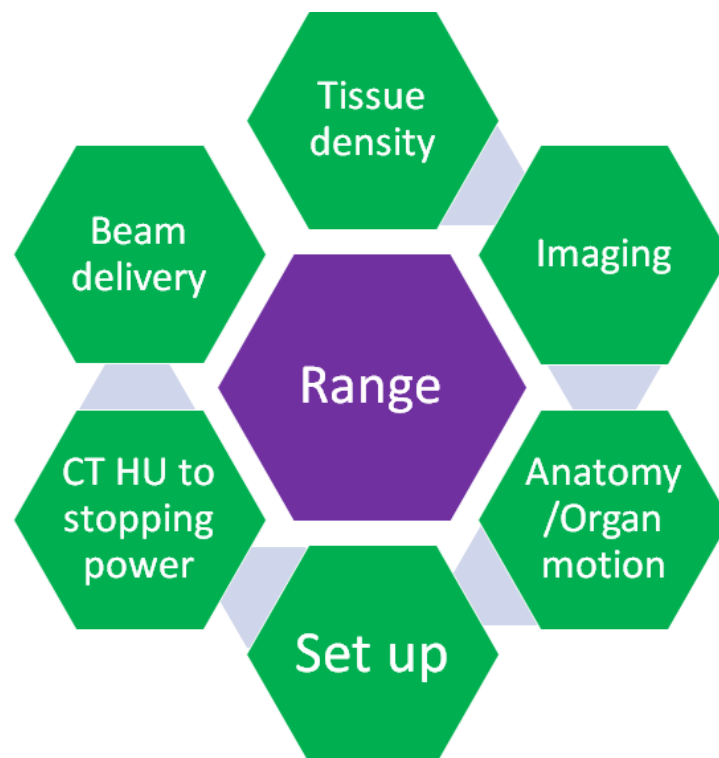
Figure 1.9. Variables affecting the uncertainty of the range of the proton beam.

Addressing the uncertainties summarised in Figure 1.9 will help avoid misplacement of the Bragg peak and reduce safety margins enabling greater normal tissue sparing. Range uncertainties can be up to several millimetres and are dependent on the location of the tumour being treated. The range uncertainty margin is therefore estimated to lie between 2.5% to 5% of the beam length with an additional 1.5-3 mm margin (148). Range uncertainties lead to the systematic and random errors discussed in Section 1.3.4.

In conventional radiotherapy a geometric PTV margin or PRV margin for OARs is applied to compensate for geometric uncertainties during the planning process. The CTV to PTV margin recipe as described by van Herk et al. ensures coverage of the CTV to within 95% of the prescription dose (51). The margin does not account for anatomical changes. The physical properties of protons mean they are more sensitive to uncertainties than photons. A PTV margin as described by van Herk is inadequate for use in protons as does not take into account differences in range, motion and tissue densities (149). The PTV margin assumes a "static dose cloud", by which the dose distribution moves rigidly within the patient. Anatomical uncertainties can alter the beam path. For tumours close to serial OARs such as the spinal cord, the risk of cold and hot spots within the target due to alteration of the beam path, has the potential for serious clinical consequences. Moyers et al in the treatment of lung cancers suggested including motion and setup uncertainties into the treatment plan design and abandoning the concept of a PTV margin for proton beam therapy (150). Robust plan optimisation is an alternative method to the PTV safety margin. Robust plan optimisation attempts to include uncertainties in range and set up that may arise during a patients treatment, into the optimisation algorithm, prior to treatment delivery (151,152).

1.4.3.4 Robust plan optimisation

Robustness of an IMPT plan is defined by how well a plan can account for uncertainties in set up and range/stopping power. Robust plan optimisation incorporates uncertainties due to patient setup and proton range directly into the treatment plan design and plans to the CTV under uncertainty instead of the PTV (153). Various methods of robust plan optimisation have been described. The two

commonest methods used in clinical practice are the worst case scenario method, and stochastic or probabilistic programming (154). Both methods attempt to minimise placement of the distal field edge of the Bragg peak directly in front of the OARs, to reduce in-field dose gradients (155). Avoidance of dental or surgical artefacts can help minimise large tissue heterogeneities and should be considered as part of the proton planning process. An example of an MFO oropharyngeal plan is shown in Figure 1.10.
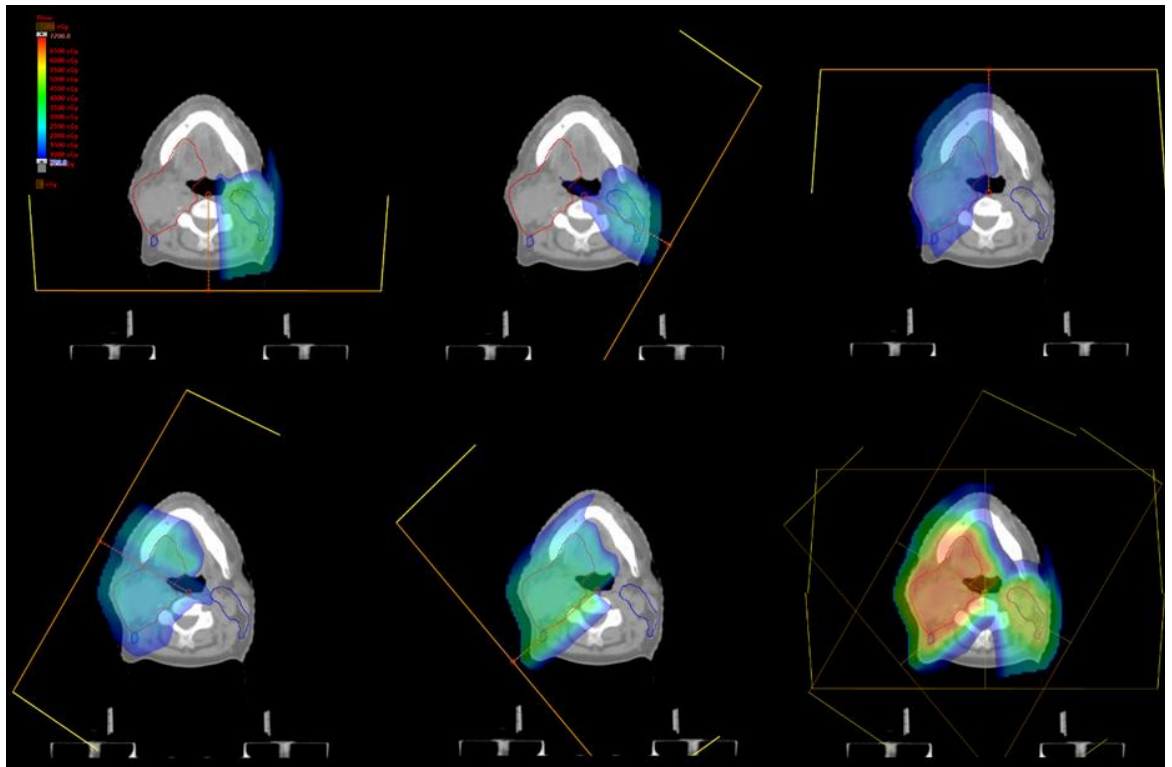
Figure 1.10. Multi-field optimisation plan for a patient with oropharyngeal cancer.

Dose patterns are shown for five individual fields and the combined dose distribution (bottom right) for IMPT.

In probabilistic planning optimisation is based on the expectation value of several scenarios (118). In the worst-case scenario approach, optimisation is based on the worst case of the scenarios. In a study of 14 patients the worst case scenario method under uncertainty, improved target coverage D95 GTV 94.6 % v 91.9%, and reduced dose to the spinal cord, brain stem, parotid glands and oral cavity (156). Studies have addressed whether to optimise to the PTV or CTV where the GTV remains intact. In a study of 14 cases by Liu et al., CTV based MFO led to superior target coverage and improved dose homogeneity to OARs (156). The use

of robust MFO has also shown promise in reducing dose to the ipsilateral parotid gland in the treatment of unilateral HNSCCs compared with SFO (157).

HPV positive OPC can present with large tumour volumes which respond early to treatment. The challenges of definitively treating large oropharynx tumours with protons is predicting the exact position of the Bragg peak due to anatomical and positional changes during treatment. Alternative methods to de-intensify treatment for favourable HPV oropharyngeal have been sought. Transoral robotic surgery (TORS) for example, has shown to improve functional and survival outcomes in early and as part of tri-modality treatment in advanced stages of OPC (158). Post-operative/adjuvant radiotherapy may be required post TORS due to positive surgical margins or extra-capsular extension. The absence of a GTV in the adjuvant setting may reduce the impact of gross anatomical changes potentially making the dose distribution more predictable. This relies, however, on good image quality and the potential need to co-register images, in order to ensure the virtual GTV is covered. Also whilst the total radiation dose is lower in the adjuvant setting, long-term functional impairment is known to be higher in those treated with dual or tri-modality treatment (159). Proton beam therapy may improve OAR sparing in the post-operative setting and reduce functional impairment.

SFO and MFO are sensitive to tissue changes throughout treatment. The benefit of MFO in the post-operative setting was a reduction in toxicities in patients with non-small cell lung cancer. The perceived benefit of MFO in OPC patients requiring adjuvant treatment is largely based on small retrospective analyses. Prospective studies are needed to evaluate the optimal MFO process and quantify the clinical benefit of adjuvant proton beam therapy in those at risk of recurrence following primary surgery.

1.4.4 Patient and public involvement in the design of the UK's first proton trial in the treatment of oropharyngeal cancer

The potential clinical advantages of proton beam therapy are clear, but due to cost and resources, access in the UK is limited. Since 2009, UK patients have travelled overseas for proton beam therapy. Due to patient inconvenience and treatment cost, the Department of Health and NHS centres have established proton beam

centres here in the UK. The UK model accounted for geographical location and demand and established two UK centres. The first proton centre in Manchester opened in December 2018 to serve Northern regions. The second proton centre in University College London Hospital serving the South is due to open in 2021. The current indications for proton beam therapy include paediatric cancers, soft tissue and bone sarcomas, some head and neck and base of skull tumours. Proton beam therapy in the treatment of head and neck squamous cell cancers will be available as part of a clinical trial. These trials will involve eligible patients travelling and staying away from home for treatment. As proton beam therapy is a new treatment for OPC, there is a need to understand the views of patients on the logistical challenges and issues of being randomised to proton beam therapy. Patient and public involvement (PPI) research is increasingly recognised to be important in helping in the design of clinical trials (160).

1.4.5 Oxygen enhanced MRI (OE-MRI)

1.4.5.1 Implications of tumour hypoxia

Tumour hypoxia is an adverse prognostic feature (74). High levels of hypoxia in a tumour increase the risk of treatment resistance, metastatic progression and death (161–164). Numerous studies showed HNSCC patients with high versus low levels of tumour hypoxia have poorer local control and survival rates. These studies used a variety of approaches to assess tumour hypoxia that included direct measurements of pO_2 (165), exogenous probes such as pimonidazole (166), endogenous probes such as CAIX (167) and HIF-1α (168,169), gene signatures (170–172) and imaging (173).

There is a high level of evidence that giving hypoxia targeted treatments with radiotherapy improves not only local control but also survival (174). The evidence is strongest in HNSCC whereby hypoxia modification with radiotherapy improved loco-regional control (odds ratio 0.71, 95% CI 0.63–0.80; p<0.001) and overall survival for 60% of patients (164). A number of approaches are available to target hypoxia that are of interest in HNSCC, which include:

(1) hypoxia modifiers to directly target the hypoxic cells such as nimorazole (27,77);

(2) increasing oxygen delivery for example using carbogen and nicotinamide (175);

(3) imaging to detect tumour hypoxia prior to or during radiotherapy with a view to dose painting/adaptive radiotherapy (66,176–178);

(4) high LET radiation to target radio-resistant hypoxic cells (179,180);

(5) combination of low LET radiation with hyperthermia to target radio-resistant hypoxia cells (181).

Randomised trials have shown an improvement in LRC and overall survival (OS) using the hypoxic radiosensitiser nimorazole. In the DAHANCA-5 trial, patients with supraglottic or pharyngeal cancers were double-blind randomised to receive nimorazole or placebo alongside conventional radiotherapy (n=414). Of those randomised to nimorazole, improvements were seen in LRC (49% v 33%, p=0.002) and 10 year OS (26% v 16%, p=0.32) with 67% of patients reporting no major side effects, only temporary nausea and vomiting (182,183). In the DAHANCA-18 phase II study evaluating the addition of nimorazole to concurrent chemoradiotherapy with cisplatin in locally advanced HNSCC, an improvement in 5-year LRC of 80% (CI 74%-85%) was observed (184). This was similar to 5-year LRC of 70% in the DAHANCA 6 and 7 study. An improvement in 5-year OS to 72% (CI 66%-78%) was also shown, being much higher than that reported in the updated meta-analysis by Pignon et al. of 34% (20). The higher OS in the DAHANCA-18 study can be explained by 75% of the patients having HPV positive OPC. However superior 5- year OS rates were also noted for cancers of the hypopharynx treated with additional nimorazole compared with chemoradiotherapy alone (50% v 30%). The improvement in LRC with hypoxia radiosensitisers is also supported by data from the TROG-0202 whereby the hypoxia-targeting bio reductive agent tirapazamine given alongside chemoradiotherapy produced similar 2 year LRC of 75% (185). The phase III ARCON (accelerated radiotherapy plus carbogen inhalation and nicotinamide) trial randomised 345 patients with T2-T4 laryngeal cancer to receive accelerated radiotherapy (AR) alone or with carbogen and nicotinamide in. The trial reported improved 5-year regional control rates in the ARCON arm (93% v 86%, p=0.04) (175). The improvement in regional control

in those treated with ARCON was thought to be due to those tumours having a significantly higher hypoxic fraction on pimonidazole staining compared with 55% treated with AR (p=0.01). The benefit of adding nimorazole to standard radiotherapy in patients with locally advanced HNSCC unsuitable for either synchronous cisplatin or cetuximab is currently under investigation. In this phase III randomised, placebo double blind UK study (NIMRAD) improvements in LRC without additional toxicity with nimorazole will be evaluated (39).

Hypoxia is an important target and prognostic factor in HNSCC. As discussed, hypoxia can be measured by many methods, e.g., pO_2 histography, exogenous markers (pimonidazole), endogenous markers (HIF-1a, HIF-2a, CA-1X, GLUT-1 or osteopontin) and gene signatures (186). There is good evidence that patients with the most hypoxic tumours identified using some of the methods benefit most from having hypoxia-targeted treatments with radiotherapy (172,175,187). Non-invasive hypoxia imaging techniques are also an attractive method to evaluate the presence and extent of hypoxia in HNSCC and are of current research interest in the UK.

1.4.5.2 MRI for assessing tumours

This thesis chapter is on the use of MRI for assessing tumour hypoxia. The tools include dynamic contrast enhanced (DCE), T2 blood oxygen level dependent imaging (T2BOLD) and oxygen-enhanced (OE) MRI. DCE MRI indirectly measures oxygen delivery and necrosis by providing an estimation of the flow through and the permeability of blood vessels after an injection of a gadolinium-based contrast agent (188). BOLD imaging, images the quantitative difference in the paramagnetic properties of deoxyhaemoglobin in red blood cells defined as R2, or magnetic resonance transverse relaxation rate. Both DCE-MRI and T2 BOLD sequences are limited by spatial resolution, artefact susceptibility and lack of direct hypoxia measurements. OE-MRI is an appealing addition to DCE-MRI and T2 BOLD sequences. It measures the changes in intrinsic tissue relaxation properties which can then be visualised on MRI.

Hypoxia imaging techniques are of interest in radiotherapy for their potential to aid treatment choice (e.g., use of hypoxia targeting treatment), assess response to

hypoxia targeting treatments and guide planning. Imaging approaches have been widely studied as tools to facilitate dose painting or GTV boost to hypoxic regions within a tumour. Hypoxia imaging has the potential to identify small hypoxia-induced radioresistant volumes requiring higher doses. By treating a smaller volume, dose to neighbouring OARs may be reduced. Multiple imaging approaches are of interest including not only MR based but also PET (173). This thesis focuses on OE-MRI.

1.4.5.3 Role of OE-MRI in assessing tumour hypoxia

OE-MRI is a novel and emerging non-invasive imaging technique, which quantifies tissue oxygenation using the paramagnetic properties of oxygen as a contrast medium (189,190). The method involves administering an oxygen challenge while a patient is in an MRI scanner. The oxygen challenge can be delivered either by nasal cannula or by facial mask. The change in concentration of molecular oxygen dissolved in the blood or tissue caused by the oxygen challenge, can be detected by an alteration in the T1 relaxation or longitudinal rate termed R1, as illustrated in Figure 1.11 (191,192).
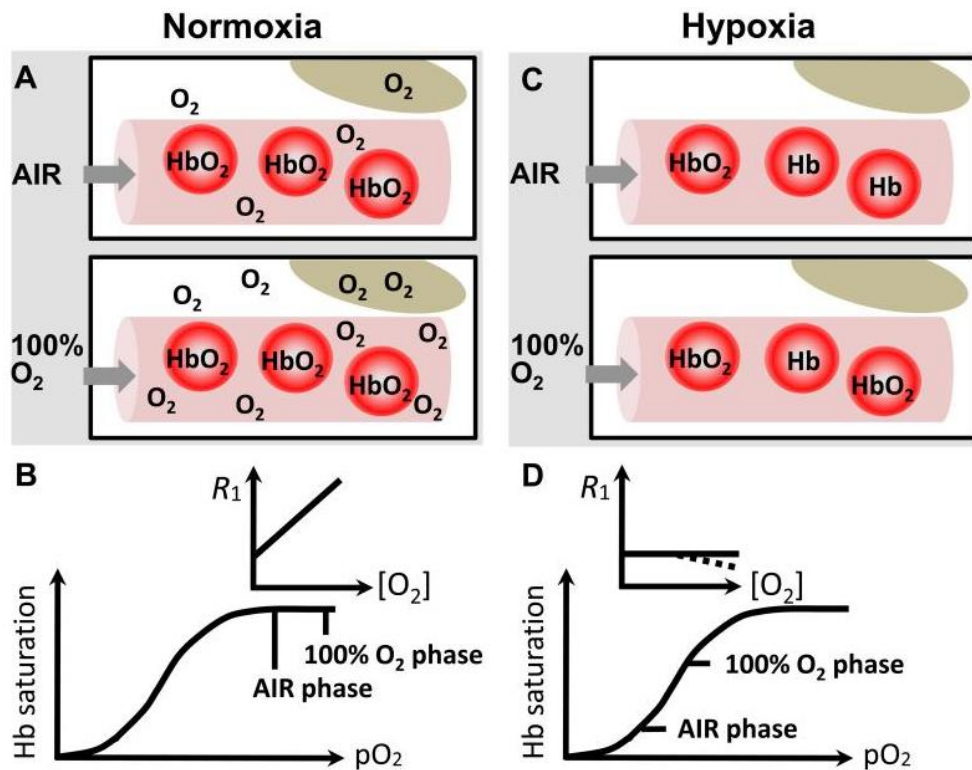
Figure 1.11. Diagram illustrating the methodology of OE-MRI for measuring tumour hypoxia[6].

The difference in saturation of haemoglobin (Hb) molecules following oxygen delivery in normoxic (A) and hypoxic tissues (B). The corresponding change in R1 values is shown on the adjacent graphs. Blood vessels (pink), erythrocytes (red), interstitial space (white), tumour cells (grey ellipse).

In air, the haemoglobin (Hb) molecules in normoxic tissues are saturated with oxygen forming oxyhaemoglobin (HbO_2) which leaves plenty of free oxygen (O_2) in the plasma. As oxygen is inhaled, the amount of oxygen dissolved in the plasma within the blood vessel (pink tube) increases, which causes an exponential increase in R_1. The change in R1 as shown on the y axis in Figure 1.11 causes a signal change, which is measured on the MRI. The change in signal is an estimate of the amount of tissue perfusion and a direct measurement of hypoxia (188). In contrast to normal oxygen saturation, if tissues are hypoxic the Hb molecules are not fully saturated with O_2. A higher concentration of deoxy-Hb is therefore found

---

[6] Image taken from (192)

in the plasma of the blood vessels (pink tubes), than in the interstitial space (white). When oxygen is inhaled, the levels of HbO_2 increase but the plasma O_2 remains the same. The R1 value stays the same as shown by the straight horizontal line or may decrease, which results in no signal detected on MRI. The fraction of tumour tissue refractory to an oxygen challenge reflects the level of hypoxia.

With favourable results of its utility in pre-clinical and clinical studies, there is now a need to establish a role for OE-MRI in clinical practice. Recently published data on the use of OE-MRI to detect variations in hypoxia between tumours from xenografts and patients with non-small cell lung cancer showed OE-MRI was a clinically feasible and reproducible technique (193,194).

OE-MRI like other imaging techniques to image hypoxia has a number of limitations that need addressing:(1) it is unknown whether it identifies true hypoxic regions; (2) small areas of hypoxia cannot be detected due to poor spatial resolution and (3) there can be fluctuations in the location of hypoxic regions. Tumour hypoxia is a dynamic process. Studies showed the percentage of hypoxia in a tumour varies between and within cancer types. For example, the percentage of tumour hypoxia using invasive pO_2 histography was shown to be higher in HNSCC versus pancreatic cancer (1.3%-1.9% v 0.3%-0.4%) (195,196). Therapies can also induce unpredictable changes in oxygenation which is difficult to detect (197). Imaging techniques such as OE-MRI may have the potential to be an image-based tool to assess tumour hypoxia in HNSCC before, during and after radiotherapy. To date, no one has evaluated the safety and feasibility of OE-MRI in head and neck radiotherapy planning.

1.5 Thesis aim and objectives

Radiotherapy to the head and neck region can lead to significant long-term toxicities such as xerostomia, dysphagia and trismus (15). Such side-effects impact negatively on patient's QOL leading to psychological distress, including low self-esteem and depression (198,199). There is a need to optimise treatments to

reduce toxicity for patients particularly those with HPV associated cancers due to their favourable prognosis and increasing prevalence in younger patients (1). IMRT has improved dose conformality but due to the low dose bath effect and anatomical changes during treatment normal tissue toxicities are inevitable. Different strategies to reduce doses to OARs have been evaluated of which all have application in routine clinical practice. To maximise the therapeutic ratio and improve patient outcomes, there is a need to:

1. improve the definition of OARs and dose constraints;
2. develop a contouring atlas to standardise volumes and improve the accuracy of the dose delivered;
3. evaluate MRI auto-contouring models to improve visualisation of complex anatomy in the head and neck, reduce clinician time and inter-observer variability without increasing exposure to ionizing radiation;
4. investigate multi-field robust plan optimisation in the post-operative setting to minimise uncertainties in range and set up and reduce the risk of increasing normal tissue toxicity;
5. evaluate OE-MRI as an image-based tool for assessing changes in the tumour during radiotherapy.

QUANTEC and EMANI published dose constraints for OARs for structures within the head and neck but omitted the masticatory muscles (8). Trismus caused by injury to the masticatory muscles is a significant cause of treatment related morbidity. The muscles of mastication OARs are not well defined and have no established dose thresholds for avoidance planning. The lack of data is partly due to disparity within the literature and small retrospective analyses. Furthermore, the lack of standardised guidelines and contouring atlases for the mastication muscles creates significant inter-observer variability amongst clinicians.

Auto-contouring tools can help standardise volumes, reduce observer variability and clinician workload. The accuracy of auto-contouring models however is limited and dependent on the data and image quality used for model development. Current auto-contouring models are CT based and lack soft tissue discrimination. MRI compared to CT has superior soft tissue visualisation, an ability to distinguish

boundaries and identify smaller structures such as lymph nodes without additional exposure to ionising radiation. MRI is a favourable imaging choice for head and neck auto-contouring but is yet to be evaluated.

Proton beam therapy is a promising alternative to IMRT to spare normal tissues whilst maintaining tumour control. IMPT is limited however by uncertainties in set up, motion, anatomical changes and beam range (81). MFO-IMPT attempts to alleviate such uncertainties. Potential inter-fractional uncertainties due to the beam's range and set up have been studied in definitive IMPT but are yet to be prospectively studied in the post-operative setting.

Use of protons in treating patients with head and neck cancer requires evidence from randomised trials. Patient and public involvement early in the development of the trials can help shape and better inform the study design. As proton beam therapy is only available as limited locations in the UK, there is a particular need to understand patients' views on their willingness to travel and stay at a proton centre.

Tumour hypoxia is a negative prognostic indicator associated with treatment resistance, poor loco-regional control and survival (59). Strategies to adjust treatment based on tumour hypoxia include radiosensitizers such as nimorazole or dose painting based on the identification of radio-resistant hypoxic regions using non-invasive imaging techniques. While some imaging approaches (e.g. PET) have been relatively widely studied, the role of OE-MRI in radiotherapy planning and treatment is not well established. OE-MRI is an emerging technique which has shown promise in lung cancer and pre-clinical studies to identify areas of hypoxia, but is yet to be studied in head and neck cancer (194).

The aim of this thesis is to investigate new techniques to optimise radiotherapy for patients with OPC to reduce treatment related toxicities and improve patient outcome. The specific objectives are to:

1. Compare dose to the individual muscles of mastication versus a combination of the muscles as a novel block with changes in trismus.
2. Develop a novel contouring atlas for the muscles of mastication and test its usability in an inter-observer variability study.

3. Develop an optimised novel CT deep learning auto-contouring model to standardise OAR volumes and reduce clinician workload.

4. Develop a novel MR deep learning auto-contouring model for OAR delineation and evaluate its use in adaptive re-planning compared with CT auto-segmentation.

5. Investigate multi-field robust proton plan optimisation in the treatment of post-operative HPV positive oral cavity and oropharyngeal cancers.

6. Evaluate patient and public opinions to aid the development of a phase III trial of proton beam therapy versus IMRT for multi-toxicity reduction in low risk HPV positive oropharyngeal cancer.

7. Evaluate the potential benefit of OE-MRI in head and neck radiotherapy treatment planning.

2.0 Methods

2.1 Prospective evaluation of relationships between radiotherapy dose to masticatory apparatus and trismus

Chapter 3 is a feasibility study investigating the relationship between dose to the muscles of mastication as individual entities or as a block structure with the risk of developing radiation induced trismus. The study was a prospective analysis of 22 patients with stage III/IV oral cavity or oropharyngeal cancer who had been recruited in Manchester into the three-centre, feasibility phase 3 randomised Trismus trial (NCT01733797) (97). The Trismus trial compared the benefit of proactive jaw opening exercises using a therabite versus a wooden spatula to improve trismus. Seventy-one patients were randomised to receive one of the interventions and asked to follow a protocol of exercises defined as the 5-5-30 regimen beginning 2-3 weeks prior to radiotherapy. The 5-5-30 regimen involved 5 sessions per day of 5 openings/closing per session with a 30 second stretch for each opening (200). Trismus is often defined as a subjective sense of jaw tightening, but here we measured the actual mouth opening with a ruler on a continuous scale.

Two of the 22 patients recruited in Manchester were excluded due to receiving primary surgery. The trial quantified trismus as maximal inter-incisor distance (MID) with the average of two measurements taken prospectively at baseline, 3 and 6 months from the start of radiotherapy. The relative % change in MID was calculated from the difference of the average of the two MID measurements at each time point. Measurements at 3 months were excluded due to poor patient compliance secondary to acute mucositis. Patient, tumour and treatment factors were available for: age, gender, smoking, alcohol, tumour location, tumour and nodal stage, exercise intervention and frequency of exercises, addition of chemotherapy, GTV and total dose to the PTV.

To calculate dose for the ipsilateral and contralateral mastication muscles (medial and lateral pterygoids, masseter, temporalis as well as the temporomandibular joint) each structure was contoured by the same clinician and peer reviewed in Pinnacle version 9.6, Philips Radiation Oncology Systems, Andover, MA. A

muscles of mastication atlas was also developed by a multi-disciplinary team comprising a consultant radiologist, consultant maxillo-facial surgeon, consultant clinical oncologist and clinical oncology research fellow. The atlas was used to aid contouring and standardise volumes. To define the block, an outline of the ipsilateral medial and lateral pterygoids, masseter and the temporalis below the orbital floor was created. The cranial component of the temporalis was excluded as from the literature it was clear that the action of the temporalis on jaw movement at the level of the TMJ, occurs below the orbital floor. The mean dose, equivalent uniform dose (EUD) and V35-60 Gy were calculated for the contours from the block and each individual muscle on the ipsilateral and contralateral side and correlated with the % change in MID at 6 months from baseline. The EUD is a method of reporting radiotherapy dose distributions taking account of non-linearity of tissue dose-response whilst not attempting to make predictions of absolute outcome (201). Data were extracted via a script to calculate dose volume histograms using in-house software.

Analysis of the results revealed a clinical outlier. It was defined as an outlier as despite a mean dose to the ipsilateral muscles of mastication >60 Gy, % change in MID improved by 6 months. The outlier was the only study participant who received a complete response by 1 year for a large (GTV 106 cm^3) T4b HPV positive OPC and who presented with trismus due to significant tumour infiltration into the medial pterygoid. The relative % improvement in trismus can be explained by tumour response rather than reduction in OAR mean dose. By removing the outlier some results became significant (p<0.05) therefore in order to avoid making large assumptions further analysis was performed with and without the outlier.

A Shapiro-Wilkinson test for normality showed a non-parametric distribution of data. Correlations between the different muscles and dose parameters were calculated using Spearman's rank and linear regression models. A p-value of <0.05 was considered statistically significant. The analysis was performed using GraphPad Prism version 6.0 and SPSS version 20.

2.2 Use of a novel atlas for muscles of mastication to reduce inter observer variability in head and neck radiotherapy planning

Chapter 4 used the muscles of mastication atlas developed in Chapter 3 and tested its ability to improve interobserver variability in head and neck radiotherapy planning. The atlas was produced by a multi-disciplinary team as described in Section 2.1. To develop the atlas the paired muscles of mastication: medial and lateral pterygoids, masseter, temporalis and temporo-mandibular joint were contoured on CT axial slices using Pinnacle version 9.6, Philips Radiation Oncology Systems, Andover, MA treatment planning system. The contours were peer reviewed by a consultant radiologist and clinical oncologist to check for consistency. Contours were extracted as DICOM files and converted into an app using in-house software by Dr Andrew Green. The app is available via a web-link using google chrome (https://bit.ly/trismusatlas). The atlas app was written in HTML5, JavaScript and webgl. A non-web version was also developed, so that it can run on a Windows or Mac desktop.

To test the clinical usefulness of the atlas, seven head and neck clinicians (five consultants and two trainees) delineated the paired muscles of mastication on randomly selected CT scans from five patients without the atlas. After a minimum gap of four weeks (to avoid memory effects) in line with previous studies, each clinician was given the atlas and asked to re-contour the same structures on the same five CT scans. Prior to asking the clinicians to contour each structure a reference or "ground truth" contour for each patient was created by the multi-disciplinary team that produced the atlas, which contained none of the above observers. In-house software was used to compare clinician-drawn structures with the reference.

To measure the effect of the atlas on interobserver variability, standardised metrics were used. Dice similarity coefficient (DSC), mean distance to agreement (DTA) and the centre of mass difference (COM) for each manually drawn contour were compared to the reference for each volume delineated without and with the atlas. DSC measures the amount of overlap volume. A DSC value of 0 means no overlap and a value of 1 is perfect alignment between two volumes. DSC is commonly used in the literature, but values can be largely affected by the volume of the structure. DTA is a surface-based metric which quantifies the distance

between two surfaces. The distribution of the shortest distances from points on each of the two surfaces to the other surface are collated as a histogram. From the histogram, mean, median, 95th percentile, standard deviation (SD), maximum and minimum values were calculated. In this study the mean and SD across all patients were compared without and with the atlas to assess inter-observer variability. Comparison was also performed split by training grade. Paired t-tests compared the mean DSC, mean DTA and distance to COM without and with the atlas. Standard deviation maps illustrated the variability in contours at different locations of the organ. Statistical significance was defined as p <0.05.

2.3 Clinical validation of a novel MR based deep learning auto-contouring models for organs at risk in head and neck radiotherapy treatment planning

Chapters 5 and 6 focus on further strategies using auto-contouring tools to improve the delineation of OARs in head and neck radiotherapy planning. In the first section of Chapter 5, the use of optimised CT based AI auto contouring models to reduce time and interobserver variability are explored. A novel MR AI auto contouring tool was developed and compared with CT methods.

2.3.1 CT based AI auto contouring model

To create an optimised CT based AI auto contouring model, two head and neck CT based models, A and B trained on local data and one generic CT model trained on external data were evaluated by two independent observers. Between 80-90 image data sets were included in each of the models. Models A and B were developed for other projects and used retrospectively; they were used separately for this project as each had different OARs delineated. The two independent observers scored the OAR auto contours from each of the three models on ten randomly selected CT scans using Mirada medical software. The OARs included; bilateral parotid glands, submandibular glands, eyes, optic apparatus (optic chiasm and optic nerves), larynx (sub glottis, glottis and supra-glottis) oral cavity, mandible, pharyngeal constrictor muscles (superior, middle and inferior), brainstem, spinal cord and crico-pharyngeal oesophagus. A qualitative 'goodness

of fit' descriptive score was used where observers scored the contours from 1-7. A score of 5 was given if less than 50% manual edits were required to meet clinical standards (202). The average scores for each OAR were calculated and compared between the two observers. The mean ± SD for DSC and DTA for each OAR were calculated to compare the contours generated by the three models against manual contours. A combined deep learning contouring expert (DLCExpert) model termed, 'model$_{CT}$' was developed using the OARs with the best goodness of fit scores and the best DTA and DSC from each of the three models. Where goodness of fit scores, DSC and DTA scores conflicted for an OAR, three physicists reviewed the scores and a consensus obtained.

Model$_{CT}$ was evaluated for its ability to reduce contouring time and interobserver variability. Ten different randomly selected CT plans of patients with T1N0M0 SCC of the larynx were used to test the model. Using pinnacle version 16, two consultant clinical oncologists and two trainees were asked to record the time taken to manually contour each OAR on each of the ten CT plans. The OARs contoured included: oral cavity, larynx, mandible, eyes, brainstem, spinal cord, bilateral parotid and submandibular glands. After a minimum gap of four weeks, the same clinicians were asked to record the time taken to manually edit model$_{CT}$ contours for each OAR. The mean, absolute and % difference in time between manual contouring and editing model$_{CT}$ were calculated and compared for each clinician and OAR.

Five CT larynx plans were next used to test the ability of the model$_{CT}$ edited contours to reduce interobserver variability. Each clinician's manual and model$_{CT}$ edited contours were compared with the remaining clinicians. The mean ± SD DTA and DSC values were compared between the manual contours and model$_{CT}$ edited contours on each plan. Paired t-tests were used to compare the DSC and DTA values for the manual and model$_{CT}$ edited contours. Statistical significance was defined as <0.05.

2.3.2 MR AI auto contouring model

A novel deep learning MR auto contouring mode termed model$_{MRI}$ for OARs was developed and compared with model$_{CT}$ edited contours described in Section 2.3.1.

75

To develop the MR auto contouring model, 100 T2 weighted (T2W) 2D MR diagnostic scans with the same sequence were selected and imported as anonymised DICOM files into the RayStation planning system version 6.99 via the theragnostic system. Bilateral parotid and submandibular glands were contoured by the same clinician on each MRI scan and peer reviewed by a consultant radiologist prior to being exported to DLCExpert software for model development. Using Workflow Box (Mirada Medical, Oxford, UK), the model was used to generate the automated contours on three different types of MR scans for validation.

Ten T2W diagnostic 2D MRI scans, ten T2W Dixon 2D MR radiotherapy scans and 8 3D MR-Linac scans (MRL) (Elekta, Stockholm, Sweden) scans were imported anonymised into Raystation version 6.99. A clinician then manually contoured bilateral parotid and submandibular glands on each scan for comparison. Finally, contours for model$_{MRI}$ were compared with model$_{CT}$ edited contours for 10 radiotherapy scans.

2.4 Inter-fraction Robustness of Intensity-Modulated Proton Therapy in the Post-operative Treatment of Oropharyngeal and Oral Cavity Squamous Cell Carcinomas

Chapter 7 describes a retrospective study which I undertook at the Roberts Proton Therapy Centre, University of Pennsylvania, USA. The aim of this study was to investigate dosimetric consequences of set up variation and anatomical change in patients receiving multi-field optimised intensity modulated proton therapy in post-operative oral cavity and OPCs. It comprised six patients with OPC or oral cavity cancers requiring post-operative proton beam therapy to the primary site and elective neck. Patients were treated between July 2017 and April 2018 and planned with multi-field robust optimisation in Eclipse (Eclipse v 13.7, Varian Medical systems). All patients were immobilised in a five-point thermoplastic shell with patient set up checked using daily orthogonal kV imaging. Weekly CBCT images were available for review.

Each patient was treated with two or three CTV dose levels. CTV1 was defined as the surgical bed with CTV2 and CTV3 defining "at-risk sites". CTV1 received 60-63

Gy(RBE), CTV2 received 54-60 Gy(RBE) and CTV3 received 54 Gy(RBE) all treated with a dose of between 1.8 to 2.1 Gy(RBE) per fraction. Each treatment plan was based on a three-field beam arrangement consisting of 2 posterior obliques and 1 anterior field with range shifters. The posterior oblique and anterior fields overlap in the superior-inferior direction over a 2 cm region. The total prescribed dose in the original (nominal) treatment plan was defined such that 95% of each target volume received a minimum of 100% of the prescribed dose to this CTV level. Clinically acceptable plans require 95% of the target volume in the worst-case scenario to receive at least 95% of the prescribed dose and at most 0.03cc of the target volumes to receive ≤110% of the prescribed dose.

Plans were optimised using multi field optimisation. CTVs were robustly optimised using a setup uncertainty of 3 mm and a range uncertainty of 3.5%. Set up uncertainties were simulated along three orthogonal directions and combined with range uncertainties in each position to produce twelve scenarios used in the robust optimisation, (x±3mm), (y±3mm), (z±3mm) and CT Hounsfield Unit (HU) scaling of ±3.5%. To evaluate the delivery, virtual CTs (vCTs) were generated by deforming the planning CT onto the CBCT using Morphon's algorithm. The deformable registration provided a high-quality vCT on which the planned dose was evaluated and an estimate of the delivered dose made. Each target CTV on the vCT was compared with the nominal plan and manually edited to modify the superior and inferior extension. The vCT plan robustness evaluation was performed using a residual set up uncertainty of 1.5 mm and range uncertainty of 3.5%. The 1.5 mm setup robustness evaluation considers the uncertainty in the coincidence between the imaging and radiation isocentres, intrafraction motion, as well as variations in user dependent choice of region of interest for evaluating image registration between the vCT and the planning CT.

All doses and dose volume histograms (DVHs) were calculated using Proton Convolution Superposition (PCS) v13.7 within the Eclipse™ treatment planning system (Varian Medical Systems, Palo Alto, CA). Plan metrics were extracted from the calculated DVHs using a MATLAB (Mathworks, Natick ,MA) script. Maximum and minimum values under uncertainty for $D_{95\%}$ for each CTV dose level and $D_{0.03cc}$ for the high dose CTV were extracted and compared with the original treatment nominal value. Mean dose in the vCT-calculated nominal case to

ipsilateral and contralateral parotid glands, oral cavity, pharyngeal constrictor muscles, larynx and maximum dose to the spinal cord were calculated. The relative percentage change in each patient's weight was recorded weekly during treatment and correlated with CTV and OAR coverage.

2.5 Patient and public involvement in the design of a multi-centre phase 3 randomised trial comparing IMPT and IMRT in oropharyngeal cancer

Chapter 8 is an editorial summarising the outcomes from patient focus groups which were organised to understand patients views on the primary objectives of the UK's first proton trial TORPEDO in HNSCC in order to help shape the trial design. TORPEDO is a phase 3 multi-centre randomised trial which aims to assess the benefit of proton beam therapy in terms of patient reported toxicities and outcome in HPV positive OPC.

Three focus groups were set up in Manchester, Sheffield and Leeds. A focus group team was set up to review the literature and brainstorm ideas on how to achieve the outcomes set out in the focus group. This team included members from the public and patient involvement department based at the NIHR Manchester Biomedical Research Centre (BRC) and two clinicians involved in the TORPEDO trial. Patients were selected based on similar eligibility criteria to that of the TORPEDO trial. All patients had oropharyngeal primaries and were either T1-4, N0-3 HPV positive and <10-year smoking history or T1-4, N0-2b and >10-year smoking history. All patients must have completed radiotherapy over 1 year ago. Fifteen invites were sent from each centre (with an expected accrual rate of 10). Included within the invites were: a patient information sheet, directions to the venue, a reply slip and a stamped and addressed return envelope. Methods of replying included phone, email or post. Each patient who replied confirming attendance at the focus group was contacted up to one week before the focus group meeting to check attendance and to answer any questions.

The venue was chosen at each site based on its accessibility for the relevant patients. The venue needed to be open, non-intimidating and have access to round tables to improve discussion. Food and drink were provided and travel expenses up to £50 were offered as well as an honorarium of £50. Each focus

group was led by three people with each assigned an individual role at the start. The focus group lasted two hours and was split into 4 tasks, with ground rules established at the start. The 4 main areas to understand patients views on were: 1) proton beam therapy, 2) radiotherapy side effects, 3a) randomisation, travel, accommodation, 3b) patient treatment pathway and 4) trial endpoint. Information was presented in the form of power point presentations and flipcharts. Small group work and informal discussions were included. Information was recorded on pre-prepared laminates, patient questionnaires, audio-recordings and via email or telephone contact with selected patients following the event.

To understand patients' thoughts on the trial endpoint we explored which toxicities they experienced at 6 weeks and by 1 year (task 2) and to state which ones they felt to be the most important. Following this their views on the University of Washington questionnaire which has been validated in clinical trials as a tool to collect outcome data at different time points were evaluated. The University of Washington questionnaire has 15 single domains made up of 6 physical, 6 socio-economical and 3 global and is one of the tools that will be used in the TORPEDO trial (203). To understand if the 6 physical measures that would make up the composite score for the trial primary outcome measure would be relevant, we compared the side effects patients reported as the most important to those measured by the University of Washington questionnaire. Written feedback on the usability of the questionnaire was collected. Following the focus groups feedback evaluation forms were given to each patient and a small sample of patients volunteered to be involved in future focus groups.

2.6 Oxygen enhanced MRI measurement in head and neck cancer: validation and efficacy of response

Chapter 9 describes a protocol for a prospective pilot study examining the potential clinical value of oxygen-enhanced magnetic resonance (OE-MRI) biomarkers in head and neck radiotherapy treatment planning. The primary objectives of the study are to evaluate the safety, feasibility and tolerability of OE-MRI and to establish an optimum sequence to be used in radiotherapy planning. This will be done using two cohorts A and B. The eligibility criteria for the study will include participants aged ≥18, ECOG 0-2, Creatinine Clearance ≥30mL/min, ability

to lie comfortably on their back for up to 1 hour and tolerate a thermoplastic shell. Those with a contraindication to MRI scanning will be excluded.

Patients treated with radiotherapy +/-chemotherapy over 6 weeks for head and neck cancer will be recruited into the study. The imaging protocol which will be used in this study is still being optimised but will include dynamic T1, DCE with gadolinium contrast and DWI acquisitions, besides the OE acquisition. The patient participants will receive up to 4 OE-MRI scans; two at baseline separated by 24 hours to 7 days, a third at week 2-3 and a fourth at week 4-5 from the start of radiotherapy. By doing serial scans at different time points the extent and changes in tumour hypoxia that occur during radiotherapy will be evaluated. A translational study will also be performed in HNSCC patients. The translational study will determine the levels of hypoxia in each of the patient's tumour tissue block and correlate them with the OE-MRI scan. Each Formalin fixed paraffin embedded (FFPE) tissue blocks will be requested from the diagnostic biopsy from each patient and evaluated using a multi-gene hypoxia signature.

As this is a novel intervention, participants will be given a patient reported questionnaire after each OE-MRI to collect qualitative data. Patients baseline demographics and imaging data will be collected. Follow-up data will be collected from the date of entry into the study until study closure and will include information on treatment response, toxicities (graded as per CTCAE criteria), and dates of relapse, progression or death.

2.7 Statistical methods used

In this thesis, I have performed all statistical calculations except for the use of ADMIRE to extract metrics of DSC, DTA and COM which was done by the advanced radiotherapy physics group. All statistical tests were performed using GraphPad Prism version 6.0, IBM SPSS version 23 and Microsoft excel 2010.

3.0 Prospective evaluation of relationships between radiotherapy dose to masticatory apparatus and trismus

C. Hague[1], K. Garcez[1], L.W. Lee[1], A. McPartlin[1], A. McWilliam[2,3], D. Ryder[4], A.J. Sykes[1], D. Thomson[1], M. Van Herk[2,3], C.M. West[5], N. Slevin[1]

[1]Department of Head and Neck Clinical Oncology, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, M20 4BX, UK
[2] Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Heath, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
[3] Department of Radiotherapy Related Research, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, M20 4BX, UK
[4] Statistics Unit, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, M20 4BX, UK
[5] Translational Radiobiology Group, Division of Cancer Sciences, Manchester Academic Health Science Centre, University of Manchester, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, M20 4BX, UK

Authors' contribution:

This paper reports the results of a feasibility study in 20 patients with head and neck cancer to define a dose threshold and avoidance structure for the muscles of mastication, to reduce the development of trismus. For this study I contoured structures used in the analysis, analysed data and wrote the manuscript. I received support from the physics department (Dr William Beasley and Dr Alan McWilliam) to extract data from the radiotherapy plans to perform dose calculations. Mr David Ryder provided statistical support.

3.1 Abstract

Introduction. This feasibility study aimed to identify relationships between radiation doses to the masticatory apparatus as a combined block or as individual subunits with changes in trismus following radiotherapy.

Methods. Twenty patients from a single centre were recruited prospectively as part of a randomised trial comparing proactive exercises in the management of trismus. Patients with stage III/IV oral cavity or oropharyngeal squamous cell cancers received Intensity Modulated Radiotherapy (IMRT) with concurrent systemic therapy. All patients had trismus prior to radiotherapy. Maximal inter-incisor distance (MID) was measured pre and 6 months from the start of radiotherapy. Bilateral muscles of mastication: medial and lateral pterygoids (MP, LP), masseters (M), temporalis (T) temporomandibular joint (TMJ) were contoured on CT axial images. The block comprised all muscles excluding the TMJ below the orbital floor. Mean dose, equivalent uniform dose (EUD) and V35-60 Gy were calculated and compared with change in MID.

Results. In six patients the MID deteriorated at 6 months from the start of radiotherapy compared with fourteen whose MID improved. No significant association was observed between age, gender, smoking, alcohol status, exercise compliance, cisplatin, tumour site, stage, V35-60 Gy or EUD with change in MID. A clinical outlier was identified and excluded. Without the outlier (n=19) a significant association was seen between mean dose and change in MID at 6 months for the ipsilateral block (p= 0.01), LP (p=0.04) and M (p<0.01). All patients where trismus deteriorated at 6 months received mean doses >40 Gy to the block.

Conclusion. Higher mean radiation doses to the ipsilateral block, LP and M were significantly associated with deterioration in trismus. Limiting dose to these structures to ≤40 Gy for tumours not invading the masticatory muscles may improve treatment related sequelae. The ipsilateral block, LP and M could be studied further as possible alternative avoidance structures in radiotherapy treatment planning.

3.2 Introduction

Trismus or 'locked jaw' is defined as an 'inability to fully open the mouth'. It is a common treatment related effect in patients with head and neck cancer resulting in significant morbidity. The prevalence of trismus varies considerably. This variability is due to patient and physician under-reporting, study differences in clinicopathologic and treatment factors, and the lack of a universally agreed definition (92,204,205).

Trismus is caused by impaired function of the muscles of mastication secondary to benign or malignant processes and their associated treatment. The muscles include: medial (MP) and lateral pterygoids (LP), masseter (M), temporalis (T) as well as the temporo-mandibular joint (TMJ). The paired muscles of mastication assist with chewing through their attachment onto the mandible. The MP, M and T close the jaw whilst the LP opens the jaw. Injury to these muscles will result in reduced function and range of mandibular motion leading to reduced nutrition, impaired oral hygiene and difficulty speaking (206). The most widely used definition for trismus is a maximal inter-incisor distance (MID) of ≤ 35 mm (90).

There are a number of patient, tumour and treatment related factors which contribute to trismus (92,207). The radiation-induced pathogenesis involves fibrosis and atrophy of the mastication muscles secondary to ischaemia (208). The effect of radiation is not immediate but one that progresses over months to years following treatment. The severity of radiation induced trismus appears to be related to the total dose received and the volume of tissues within the radiation field (209). Significant dose-response relationships have been found but the literature varies considerably on what dose constraints should be used. There is also no consensus as to which of the mastication muscles should be defined as an organ at risk with a view to avoidance planning (207,210,211). Rao and Van der Molen reported that doses to the pterygoids and masseter muscles were the most robust predictors for the development of trismus whilst other studies showed doses to the TMJ and the pterygoids were important (102–104). A limitation of published studies is the use of retrospectively collected data and mouth opening measurements based on dichotomized data (100,212).

Given these limitations a feasibility study was carried out using prospectively collected data to identify and compare different dose parameters to a combination of these muscles defined as a block and as individual mastication muscles. Mouth opening measurements on a continuous scale were used to increase statistical power.

3.3 Methods

3.3.1 Patients

The study population comprised patients recruited in Manchester into the multi-centre phase 3 randomised controlled Trismus trial (213).The trial compared proactive exercises using therabite (platon medical) versus standard wooden spatulas in patients with stage III/IV oral cavity/ oropharyngeal cancers to improve trismus. Patients randomised to either intervention were asked to follow a protocol of exercises beginning 2-3 weeks prior to starting radiotherapy. These included 5 sessions per day of 5 openings/closing per session with a 30 second stretch for each opening (97). Patients had stage III/IV squamous cell carcinomas of the oral cavity/oropharynx and were treated with concurrent radiotherapy and systemic therapy to the primary tumour from February 2013 to January 2015. Of the 22 patients enrolled in Manchester, two patients were excluded due to undergoing primary surgery with mandibulectomy. The remaining 20 underwent ipsilateral, bilateral neck dissections or hemi-glossectomies as shown in Table 3.0.

Patients received 60-66 Gy in 30 fractions (2.0-2.2 Gy per fraction) over 6 weeks. A rotational based Intensity Modulated Radiotherapy treatment (IMRT) plan was calculated using the pinnacle treatment planning system (Pinnacle version 9.6, Philips Radiation Oncology Systems, Andover, MA) with target delineation performed on axial CT images. No dose constraints were applied to the muscles of mastication. All patients had concurrent chemotherapy with cisplatin or cetuximab if cisplatin was contraindicated.

3.3.2 Evaluation of trismus and dose

Patients enrolled in the trial had trismus defined as a sense of jaw tightening self-reported by each patient prior to radiotherapy. For this study, MID measurements were taken prospectively at baseline and 6 months from the start of radiotherapy. The averages of two MID measurements were taken at baseline and at 6 months. The change in MID for each patient was calculated.

The block was defined as the MP, LP, M and T muscles below the orbital floor to exclude the cranial component of the T. Mean dose, equivalent uniform dose (EUD) and V35-60 Gy were calculated for the contoured volumes of the block and each individual muscle on the ipsilateral and contralateral side. The EUD is a method of reporting radiotherapy dose distributions taking account of non-linearity of tissue dose-response whilst not attempting to make predictions of absolute outcome (201).

An in-house contouring atlas was designed for the masticatory apparatus to aid contouring. Absorbed doses were re-calculated for the block and masticatory apparatus using dose volume histograms. Data were extracted via a script to calculate dose volume histograms using in-house software.

3.3.3 Statistics

Analysis was performed using graph pad prism version 6, SPSS version 23 and Microsoft excel version 2010. Data analysis compared patient, tumour and treatment related factors with percent change in MID at 6 months from baseline. Non-parametric tests of Fishers exact, chi-squared and chi-squared test for trend were used. Correlations between the different muscles and dose parameters were calculated using Spearman's rank and linear regression models. A p value of <0.05 was considered statistically significant.

3.4 Results

The 20 patients reviewed all had established trismus at baseline. In six the change in MID deteriorated from a median baseline of 29 mm (17-34) pre-radiotherapy to

18 mm (6-29) 6 months from the first radiation treatment. In 14 patients the MID improved from a median baseline of 16 mm (8-34) to 31 mm (9-39) at 6 months. At the time of the analysis no patients had recurred with residual disease. Table 3.0 shows the distribution of patients in the deterioration and improvement groups in relation to a number of parameters. Patients with a deterioration in MID had more involved lymph nodes (p=0.04). However, the distribution of exercise frequency did not differ between the deterioration and improvement group (p=1.00), and there were no other statistically significant differences in relation to patient, tumour and treatment related factors. Although not statistically significant, mean doses received by the block and individual ipsilateral muscles were higher in the deterioration group with the exception of the temporalis muscle. A clinical outlier was detected within the improvement group. Further analyses were performed with and without the outlier.

Table 3.0. Associations between patient, tumour and treatment related factors with changes in MID

Variable		Number (%)	
		MID deterioration n=6	MID improvement n=14
Age (years)	Median (range)	60 (48-64)	61 (43-70)
Gender	Male	6 (100%)	11 (79%)
	Female	0	3 (21%)
Smoking status	Current	0	0
	Ex-smoker	4 (67%)	9 (64%)
	Never smoked	2 (33%)	5 (36%)
	Pack years, median(range)	17(14-38	27 (10-91)
Alcohol	Current/Previous heavy	3 (50%)	3 (21%)
	Never heavy	3 (50%)	11 (79%)
Type of exercise device	Therabite	4 (67%)	5 (36%)
	Wooden spatula	2 (33%)	9 (64%)
Frequency of daily exercises	≤2 daily	4 (67%)	8 (57%)
	3-5 daily	2 (33%)	4 (29%)
	Missing	n/a	2 (14%)
Tumour sub-Site	Oral cavity	1 (17%)	5 (36%)
	Oropharynx	5 (83%)	9 (64%)
Tumour stage	Stage I	0	0
	Stage II	0	0
	Stage III	0	3 (21%)
	Stage IVA	5 (83%)	10 (71%)
	Stage IVB	1 (17%)	1(7%)
T stage	T1	1 (17%)	4 (29 %)
	T2	4 (67%)	4 (29 %)
	T3	1 (17%)	2 (14 %)
	T4a	0	3 (21%)
	T4b	0	1 (7%)
Regional lymph nodes	N0	0	2 (14%)
	N1	0	2 (14%)
	N2a	0	1 (7%)
	N2b	4 (67%)	8 (57%)
	N2c	1 (17%)	1 (7%)
	N3	1 (17%)	0
Surgical intervention prior to radiotherapy	Ipsilateral neck dissection	6 (100%)	10 (71%)
	Bilateral neck dissection	0	1 (7%)
	Hemi-glossectomy	0	2 (14%)
	None	0	1 (7%)
Concurrent cisplatin chemotherapy	Yes	4 (67%)	9 (64%)
	No	2 (33%)	5 (36%)
Total radiotherapy dose (over 6 weeks)	Median (range)	63.0 Gy (60-66)	64.5 Gy (60-66)
Mean doses (median, range)	Ipsilateral block	43 Gy (41-49)	40 Gy (18-64)
	Ipsilateral MP	62 Gy (57-65)	57 Gy (26-66)
	Ipsilateral LP	41 Gy (40-58)	39 Gy (3-66)
	Ipsilateral M	42 Gy (36-50)	35 Gy(23-61)
	Ipsilateral T	19 Gy (16-36)	26 Gy (3-65)
	Ipsilateral TMJ	23 Gy (11-44)	20 Gy (4-63)

Negative correlations were observed between mean doses to the block and individual ipsilateral muscles with changes in MID. The correlation improved once the outlier was excluded across the masticatory apparatus (Table 3.1). The size of this effect increased with the outlier excluded and reached statistical significance for the ipsilateral block (β -6.26, p <0.01), MP (β – 6.61, p <0.01), LP (β- 4.15, p <0.01) and M (β -4.58, p=0.02) muscles. Figure 3.0 illustrates a change of 4-7% in mouth opening per gray of dose to the relevant structures.

Table 3.1. Correlations between the mean radiation dose to the muscles of mastication with change in MID

	With outlier (n=20)		Without outlier (n=19)	
	r^s value*	p value	r^s value	p value
Ipsilateral Block	-0.42	0.07	-0.58	0.01
Ipsilateral MP	-0.27	0.25	-0.42	0.09
Ipsilateral LP	-0.32	0.16	-0.41	0.04
Ipsilateral M	-0.44	0.05	-0.61	<0.01
Ipsilateral Temporalis	-0.23	0.65	-0.52	0.39

Abbreviations: Calculated using Spearman rank correlation. MP= medial pterygoid; LP= lateral pterygoid; M= masseter; MID= maximal inter-incisor distance
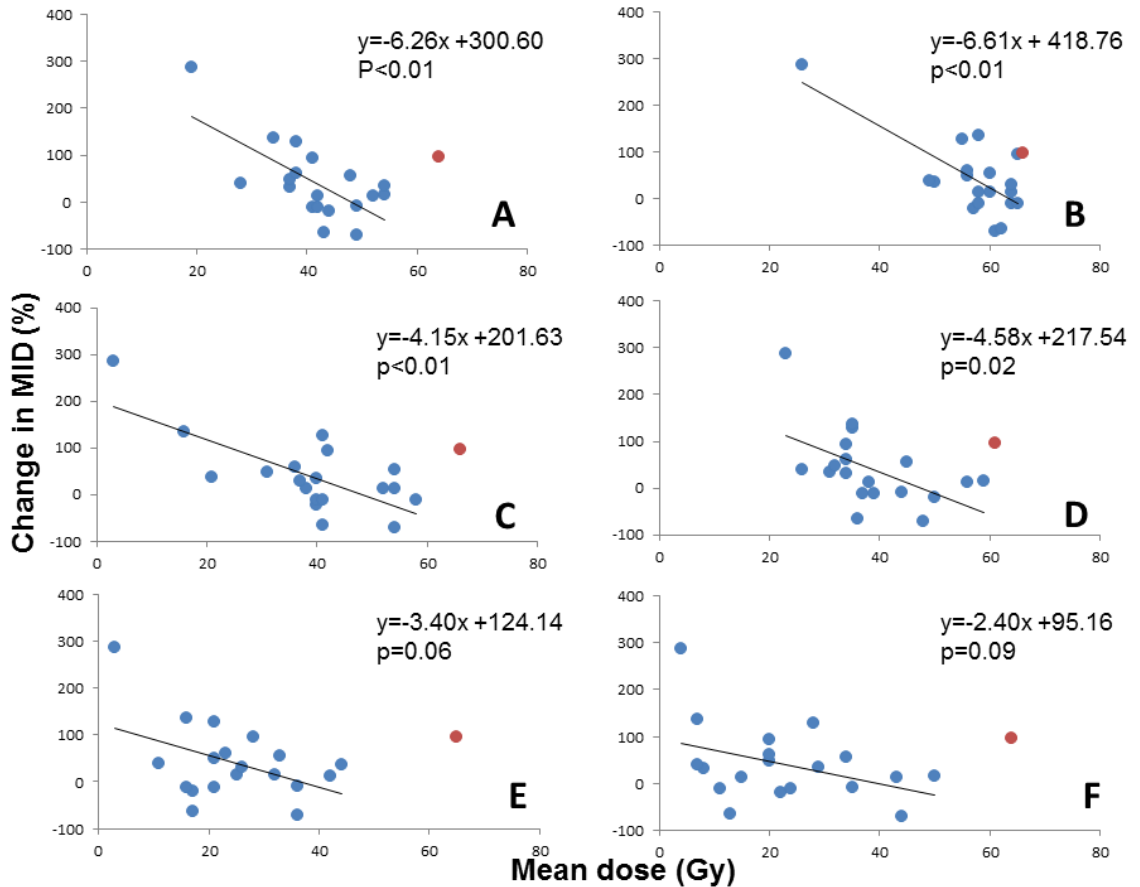
Figure 3.0. Scatter plots showing the relationships between the percentage change in MID at 6 months following radiotherapy and mean doses to the muscles of mastication.

The clinical outlier is shown by the red point. A (ipsilateral block), B (ipsilateral medial pterygoid), C (ipsilateral lateral pterygoid), D (ipsilateral masseter), E (ipsilateral temporalis) and F (ipsilateral TMJ). Data are for 19 patients.

Abbreviations: MID= maximal inter-incisor distance

3.5 Discussion

This feasibility study using prospective data showed an association between mean dose to the ipsilateral block and individual masticatory muscles with the development of trismus. All patients had established trismus prior to radiotherapy measured on a continuous scale. The change in MID 6 months from the start of radiotherapy is a similar metric to that used by Joyce van der Geer et al who observed a peak in the prevalence of trismus at 6 months (204).

In our study there were no statistically significant associations observed between changes in trismus and age, gender, tumour stage or site, concurrent cisplatin or

frequency of proactive exercises during and after treatment. Even though not statistically significant (p=1.00), a relationship was observed between the use of proactive exercises during and after treatment, in those whom the MID improved the frequency of daily exercises performed was higher. Mean doses to the ipsilateral LP (r^s -0.41, p 0.04) and M (r^s -0.61, p <0.01) were significantly associated with change in MID. There was also a correlation between mean dose to the ipsilateral MP but this did not reach statistical significance (r^s -0.42, p 0.09). The relationship between trismus and mean dose to the ipsilateral MP is similar to published data by Kent et al. In a study of 40 patients Kent identified a 45% prevalence in trismus with mean doses to the pterygoids of ≥55 Gy (102). In our study mean dose to the ipsilateral M was most correlated with trismus. This is similar to other retrospective studies. In a study of 139 patients a significant relationship between mean doses to the ipsilateral masseter of 60 Gy was found (94) contrasting with a study of 55 patients where doses of V20-V40 Gy to the masseter were associated with the development of trismus (104) .

In one patient referred to as a clinical outlier, the MID improved at 6 months despite mean doses in excess of 60 Gy received by all masticatory muscles. This outlier was the only HPV positive, stage T4b tumour in the cohort. The large primary base of tongue tumour had invaded into the pterygoid muscles resulting in trismus at diagnosis. The contoured GTV was much larger at 106 cm^3 than the rest of the cohort (median 15 cm^3, range 3-106 cm^3). This patient had very few risk factors for developing late effects being a non-smoker, absence of concurrent platinum-based chemotherapy and no surgery. Post treatment scans confirmed a complete response with no current evidence of recurrence 12 months following treatment. The improvement in trismus can be explained by tumour response and improved pterygoid muscle function.

Although there are known differences in muscle architecture and function between the pterygoid muscles an association was found between mean dose to the MP and LP with trismus. This can be explained by the close proximity of the pterygoid muscles and synergistic function to open and close the jaw (12). The mean dose received by the ipsilateral LP was lower than the MP in those where by trismus improved (39 Gy versus 57 Gy). This is in keeping with Hsieh et al who suggested a mean dose constraint of <42 Gy to the LP to reduce the risk of trismus (101).

The mean doses to the ipsilateral MP, LP, M and TMJ excluding the clinical outlier were greater in those where trismus deteriorated at 6 months from the start of radiotherapy compared to those where trismus improved. There was no significant correlation observed between V35-V60 Gy with changes in trismus across all paired muscles and the block. Our observation contrasts with another study in 124 patients where doses of V40-V60 Gy to the ipsilateral masseter were associated with statistically significant changes in trismus (100). The difference in the dose-response relationships between the two studies may be attributed to the small sample size and short follow up period of 6 months in our study compared with a median follow up of 66 months in the study by Lindbolm et al. In our study there was a weaker and non-significant association seen between the EUD and change in MID compared with the mean dose. The EUD was calculated to evaluate potential non-linear dose effects within a muscle, but the results are less strongly correlated with changes in trismus than the mean dose.

To our knowledge only the effect of dose on the masticatory muscles as individual subunits has been explored previously. An avoidance dose for the masticatory apparatus has not been agreed due to disparity within the literature. Our study is the first to describe and evaluate the concept of a block as an alternative OAR. The block includes the MP, LP, M and T below the floor of the orbit to exclude the cranial component of the T. These muscles were chosen due to their similarities in function and anatomical locations. There was a modest significant correlation (r^s -0.58, p=0.01) between mean dose to the ipsilateral block and change in MID. Patients with a deterioration in trismus 6 months from the start of radiotherapy received mean doses >40 Gy to the ipsilateral block, MP, LP and M. Significant correlations were seen between mean dose and changes in MID with the ipsilateral block, LP and M. Limiting mean dose ≤40 Gy to the ipsilateral block, LP and M for tumours not invading masticatory muscles in those with established trismus prior to radiotherapy and whom underwent proactive exercises could be considered. Mean dose ≤40 Gy is considerably lower than that reported for each individual muscle both in our study and in the published literature. The lower mean dose constraint for the ipsilateral block can partly be explained by the larger volume of muscle within the block compared to the muscles as separate entities. The concept of grouping muscles together as a block would help remove uncertainties as to which masticatory muscles should be avoided. It would also

improve our understanding of the clinical significance of radiotherapy to the regions in the muscle interface such as fat, fascia or nerves of which little is known. Image based data mining has illustrated this point through highlighting an area adjacent to the masseter that has a dose response relationship (214). Avoidance structures such as the block may help reduce the severity of radiation induced trismus but require validation in larger studies.

3.5.1 Limitations to the study

Although this is the first paper to our knowledge that uses prospective pre and post radiotherapy MID measurements on a continuous scale to evaluate dose effects on organs at risk, the study is small. All patients had subjective jaw tightening prior to the start of radiotherapy and underwent proactive exercises during and after treatment. The results are hypothesis generating and require validation in a larger study. Our findings need to be verified in a larger sample and matched control cohort with longer follow up. A larger study would enable development of a normal tissue complications probability (NTCP) model and recommendations for mean dose constraints to use in radiotherapy planning. However, our work does show that further studies using MID methodology would be useful.

3.6 Conclusion

In this prospective study, higher mean radiation doses to the ipsilateral block, LP and M were significantly associated with trismus. The findings suggest that limiting mean dose to the ipsilateral block, LP and M to ≤40 Gy for tumours not invading the masticatory muscles may reduce treatment related morbidity. The ipsilateral block, LP and M should be studied further as alternative OARs and possible avoidance structures in radiotherapy planning in future studies. This suggestion requires validation in a larger study with longer follow up.

3.7 Acknowledgements

4.0 Use of a novel atlas for muscles of mastication to reduce inter observer variability in head and neck radiotherapy planning

C. Hague[1], W. Beasley[2], L. Dixon[1], S. Gaito[1], K. Garcez[1], A. Green[2,3], L.W. Lee[1], M. Maranzano[4], A. McPartlin[1], H. Mistry[2], D. Mullan[5], A.J. Sykes[1], D. Thomson[1], M. van Herk[2,3], C.M. West[6], N. Slevin[1]

[1] Department of Head and Neck Clinical Oncology, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, UK
[2] Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Heath, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
[3] Department of Radiotherapy Related Research, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, M20 4BX, UK
[4] Department of Oral-Maxillo-Facial and Plastic Reconstructive Surgery, Central Manchester University Hospitals, Manchester, UK
[5] Department of Radiology, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, UK
[6] Translational Radiobiology Group, Division of Cancer Sciences, Manchester Academic Health Science Centre, University of Manchester, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, UK

Authors' contribution:

This paper reports a novel muscles of mastication atlas and evaluates its ability to reduce time and improve consistency of contouring. For this study I developed the muscles of mastication atlas with the help from Dr Mullan, Consultant Radiologist, Mr Maranzano, Consultant Maxillo-facial surgeon and Dr David Thomson, Consultant Clinical Oncologist. I performed the data collection, analysis and wrote the manuscript. I received guidance from Mr Hitesh Mistry, Statistician. Mr Andrew Green using in-house software developed the atlas into a free app which is available via google chrome with the link: (https://bit.ly/trismusatlas).

4.1 Abstract

Introduction. Trismus is caused by injury to the masticatory muscles resulting from cancer or its treatment. Contouring these muscles allows avoidance planning to potentially reduce radiation related trismus. Clinician contouring, however, is undermined by interobserver variability. This study aimed to evaluate a novel in-house contouring atlas for the masticatory muscles and assess its ability to reduce interobserver variability.

Methods. An atlas was developed by a multi-disciplinary team. The muscles of mastication in the atlas were: medial and lateral pterygoids (MP, LP), masseter (M), temporalis (T) as well as the temporomandibular joint (TMJ). Seven head and neck clinicians delineated five paired muscles of mastication on CT scans from 5 patients without the atlas. After a minimum gap of four weeks, clinicians were provided with the atlas and asked to re-contour the structures. Using contours generated by the same multi-disciplinary team on the same 5 CT scans as the reference, the dice similarity coefficient (DSC), mean distance-to-agreement (DTA) and the centre of mass (COM) difference were calculated and compared for each volume with and without the atlas. Comparison was also performed split by training grade. Mean and standard deviation (SD) values for all the patients were measured to assess interobserver variation.

Results. Using the three contouring consistency models the atlas reduced interobserver variability for all muscles. Mean DTA significantly improved for the MP ($p=0.01$), M ($p<0.01$), T ($p<0.01$) and TMJ ($p<0.01$). There was an improvement in mean DTA using the atlas for the trainees across all masticatory muscles, with the largest improvement and reduction in variability observed for the T (4.26 ± 7.12 v 1.21 ± 0.38 mm, $p=0.06$) and TMJ (2.06 ± 0.71 v 0.78 ± 0.26 mm, $p<0.01$). Distance between the COM and interobserver variability reduced in all directions for the MP and T.

Conclusion. A new atlas for contouring masticatory muscles during radiotherapy planning for head and neck cancer has been developed. The atlas reduces interobserver variability and could be considered as an educational tool for trainees. Reductions in interobserver variability and improved identification of structures may improve the accuracy of radiotherapy avoidance planning with the aim of reducing radiation-related toxicity.

***Key words**. Trismus, atlas, radiotherapy, contouring, inter observer variability*

4.2 Introduction

Radiotherapy to the head and neck is challenging due to complex anatomy and large number of organs at risk (OARs). Current radiotherapy techniques such as Intensity Modulated Radiotherapy (IMRT) increase dose conformity allowing improved loco-regional tumour control as well as reduced normal tissue effects (84,215). To fully exploit the advantages of IMRT, accurate and consistent target delineation is required. Manual target volume and OAR delineation are affected by clinician variability (107). Minimising interobserver variability will improve the accuracy of the dose delivered, maximise tumour control, limit toxicities and increase knowledge of organ at risk dose (111,216,217). Methods to standardise organ at risk volumes have been developed including superior imaging techniques, peer review and the development of contouring atlases (218).

Contouring atlases can help standardise volumes, reduce interobserver variability and improve normal tissue sparing in daily clinical practice (45,219–221). Atlases agreed by an expert panel may reduce inconsistencies between radiotherapy centres and facilitate multi-institutional clinical trials (40). There are a number of atlases for head and neck cancer: the current Danish Head and Neck Cancer Group (DAHANCA), European Organization for Research and Treatment of Cancer (EORTC) and Radiation Therapy Oncology Group (RTOG) (40). One limitation of current published atlases is the absence of delineation guidelines for the masticatory muscles.

Trismus is defined as a maximum inter-incisor distance of ≤35 mm and is caused by impaired function of the masticatory muscles (222). Trismus can manifest in poor dental hygiene, impaired chewing, malnutrition and psychological difficulties including low self-esteem, depression and suicidal intentions, which all reduce health-related quality-of-life (97,223). Clinical assessment of patients is challenging due to a restricted ability to assess disease status. There are several patient, tumour and treatment related factors for trismus of which radiotherapy is a known contributor with an incidence in advanced oropharyngeal cancers of 35-55% (91,103–105,224). Mouth opening is a complex action controlled by the synergistic actions of the paired muscles of mastication. These include: medial and lateral pterygoids (MP, LP), masseter (M), temporalis (T) as well as the

temporo-mandibular joint (TMJ) (225). The origin, insertion and function of each of the muscles of mastication are summarised in Supplementary Tables1a and 1b. Identifying the masticatory apparatus as an OAR with a view to avoidance planning will aim to reduce toxicities and improve quality-of-life. Dosimetric studies showed a relationship between the severity of trismus with dose and volume of muscle treated (226). Despite this there is no standardised defined OAR or dose threshold for the masticatory muscles for radiotherapy planning (209). Within the current literature, inconsistences exist regarding proposed dose parameters as summarised in Supplementary Table 2. For example, the largest most recent study by Rao et al of 421 patients suggested limiting the high dose volume of the ipsilateral MP to V_{68} Gy <10cm^3 (103).

Few studies have evaluated the use of delineation guidelines to improve interobserver variability in contouring OARs. A paper by Brouwer et al of 6 head and neck clinicians showed poor compliance with delineation guidelines for the spinal cord, parotid and submandibular glands was associated with an increase in interobserver variability (227). Currently there are no standardised delineation guidelines for contouring the masticatory muscles (40). Accurate delineation of these muscles with a validated atlas is a prerequisite for high quality radiotherapy planning to improve consistency, standardise contours and reduce radiation related trismus. This study aimed to evaluate a novel muscles of mastication atlas to aid clinician contouring, reduce interobserver variability, support training and the development of multi-institutional clinical trials.

4.3 Methods

A muscles of mastication atlas was developed by a multi-disciplinary expert team consisting of a consultant radiologist, maxillo-facial surgeon and clinical oncologists. Using the Pinnacle (Pinnacle version 9.6, Philips Radiation Oncology Systems, Andover, MA) treatment planning system, the paired muscles of mastication (medial and lateral pterygoid (MP), (LP), masseter (M), temporalis (T) and temporo-mandibular joint (TMJ)) were contoured on computed tomography (CT) slices. All muscles were delineated using the soft tissue window with the exception of the TMJ which was contoured on a bone window. The contours were

extracted as DICOM files and converted into an app using in-house software. The atlas app is shown in Figure 4.0 with a link attached https://bit.ly/trismusatlas (to access the webpage please open with google chrome version 56 or opera version 43). Included in the app is a table explaining the anatomical boundaries of each component of the muscles of mastication.

Atlas Images Overview

Soft tissue window   Bone window

Key

Colour	Structure
	leftLateralPterygoid
	leftMasseter
	leftMedialPterygoid
	leftTemporalis
	leftTMJ
	rightLateralPterygoid
	rightMasseter
	rightMedialPterygoid
	rightTemporalis
	rightTMJ
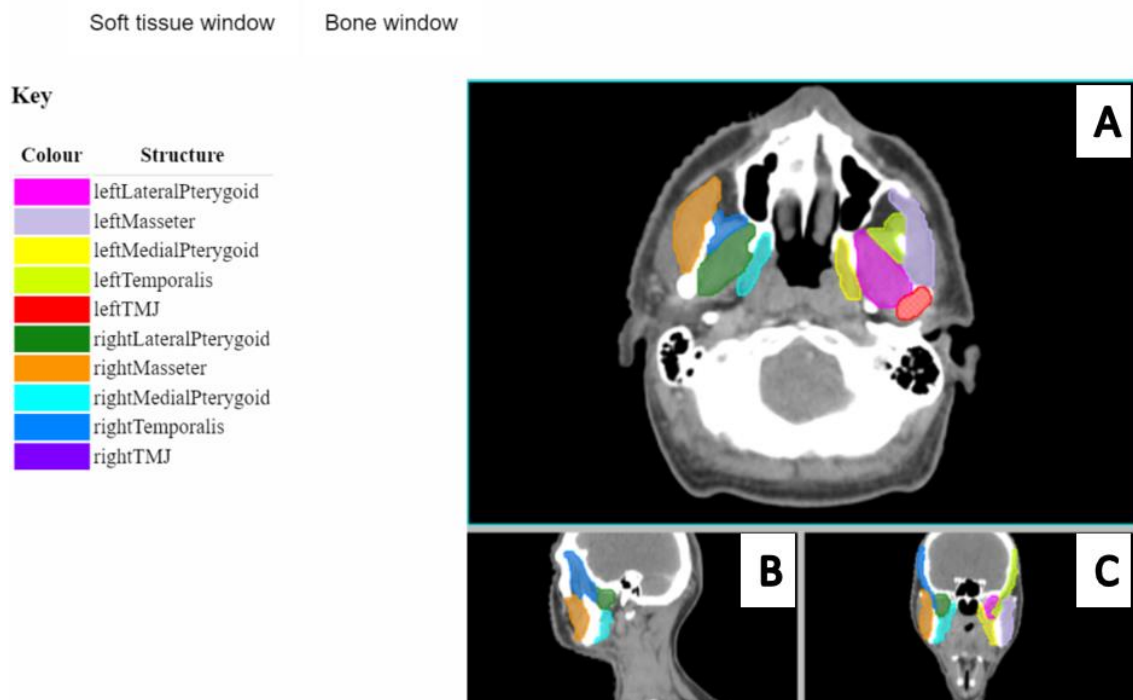
Figure 4.0. Overview of the masticatory muscles' atlas.
Axial (A), sagittal (B) and coronal (C) slices are shown. The right TMJ is not visible on the CT slice shown in Figure 4.0. https://bit.ly/trismusatlas (to access the webpage please open with google chrome version 56 or opera version 43).

Seven head and neck clinicians (five consultants and two trainees not included in the multi-disciplinary team) delineated the paired muscles of mastication on randomly selected CT scans from five patients without the atlas. After a minimum gap of five weeks each clinician was given the atlas and re-contoured the same structures on the same five CT scans.

Contours were created for each patient by the same multi-disciplinary team that produced the atlas and used as the reference. In-house software was used to compare clinician-drawn structures with the reference. Dice similarity coefficient (DSC), mean distance to agreement (DTA) and the centre of mass difference

(COM) were evaluated and compared to the reference for each volume without and with the atlas. DTA was calculated by measuring the distance from each point on the reference surface to the closest point on the clinician-drawn surface and combining into a DTA histogram. The mean DTA was then calculated from this DTA histogram. The mean and standard deviation (SD) across all patients were compared without and with the atlas to assess inter-observer variability. Comparison was also performed split by training grade. Standard deviation maps were produced to illustrate the variation in structure delineation between clinicians at each voxel without and with the atlas. The Standard deviation (SD) per voxel is calculated in 3D by combining all clinician contours. All voxels inside a contour are given a value of 1 and all outside the contour are given a value of 0. The standard deviation of each voxel is calculated to illustrate regions in which there is little agreement between clinicians: the larger the standard deviation the larger the disagreement.

4.3.1 Statistical analysis

Analysis was performed using GraphPad prism version 6 (Graph pad software) and Microsoft Office Excel 2010. A paired t-test was used to compare mean DSC, mean DTA and distance to COM without and with the atlas. Statistical significance was defined as $p \leq 0.05$.

4.4 Results

The median (range) time between contouring without and with the atlas across all seven clinicians was 66 (35-145) days. Using the atlas there was an increase in the mean delineated volumes for all muscles excluding the TMJ, as shown in Table 4.0. The SD of the contoured volumes significantly reduced using the atlas for the MP (p=0.01), T (p=0.05) and TMJ (p<0.01). The difference in distance between the COM and SD reduced significantly in all directions with the atlas for the T: anterior-posterior 2.1±1.4 vs 4.7±4.7 mm, p=0.03; left-right 4.5±3.0 vs 8.7±5.9 mm, p=0.03; superior-inferior 3.4±2.8 vs 7.0±4.7 mm, p=0.05. No significant difference in the COM distance was observed with the atlas for the LP, MP and M.

Table 4.0. Comparison of the contoured volumes of the muscles
of mastication without and with the atlas

	Volume Mean ± standard deviation (cm3)		P value
	No atlas	Atlas	
Lateral pterygoids	6.7±1.5	6.8±1.2	0.70
Medial pterygoids	7.7±1.9	8.5±1.3	0.02
Masseters	19.4±2.1	20.1±1.7	0.04
Temporalis'	20.3±6.6	28.0±4.3	<0.01
TMJs	1.7±1.1	1.3±0.3	<0.01

Abbreviations: DTA=distance to agreement; SD=standard deviation; TMJ=temporo-mandibular joint

Mean DSC significantly improved using the atlas for the LP (0.8±0.1 v 0.8±0.1 $p<0.01$), MP (0.7±0.2 v 0.7±0.2, $p<0.01$), T (0.7 ±0.2 v 0.8±0.1, $p<0.01$) and TMJ (0.6±0.2 v 0.8±0.1, $p<0.01$). No significant improvement in mean DSC was observed using the atlas for the M (0.9±0.1 v 0.9±0.0, p=0.27), Figure 4.1.
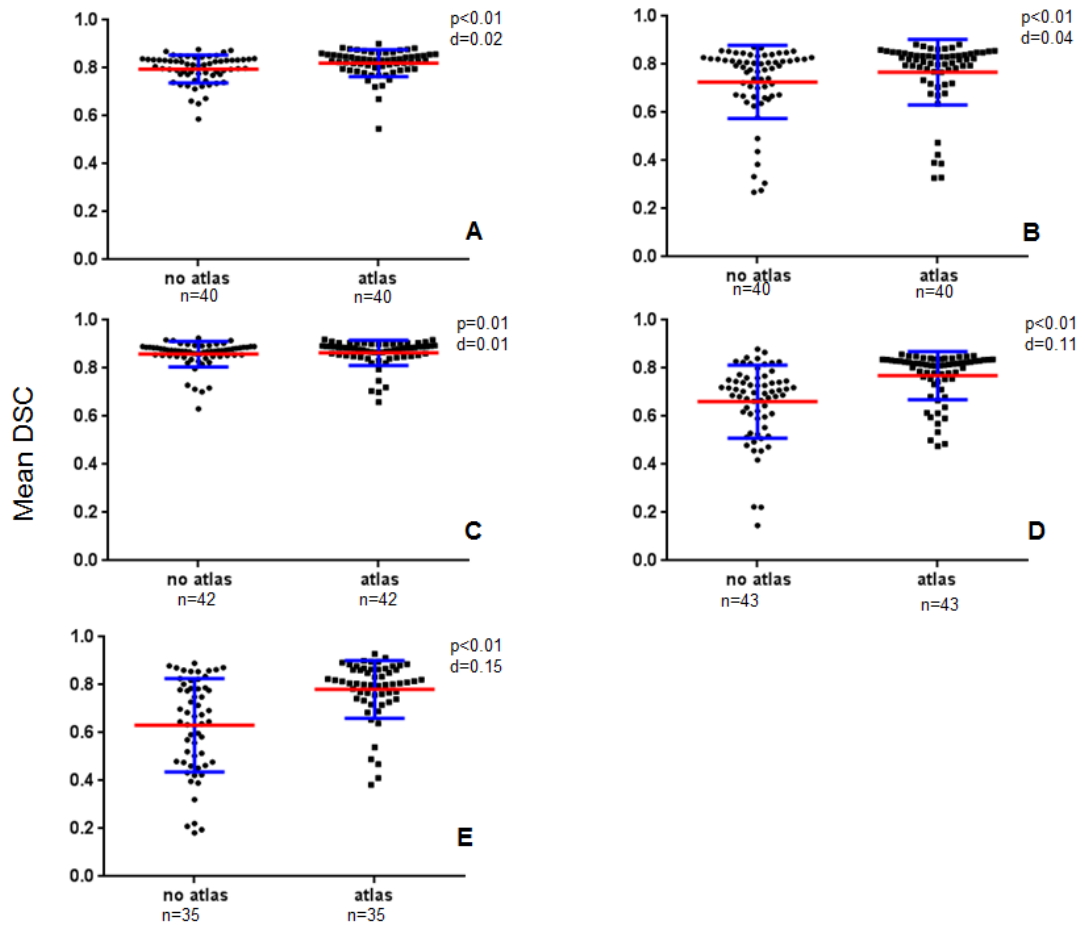
Figure 4.1. Comparison of the Dice similarity coefficient (DSC) for individual clinician manual contours without and with the atlas across the 5 pairs of masticatory muscles.

A. Lateral pterygoid, B. Medial pterygoid, C. Masseter, D. Temporalis, E. Temporo-mandibular Joint. Red horizontal bar illustrates the mean, blue bars illustrate the standard deviation. d equals the difference in absolute mean values.

Mean DTA improved using the atlas for all muscles, reaching significance for the MP (3.5 ±4.1 mm, 3.0 ±3.8, p=0.01), M (1.4 ±0.4 v 1.2 ±0.4, p<0.01), T (5.4±5.6 v 1.6±1.7 mm, p<0.01) and TMJ (1.7±1.2 v 0.9±0.7 mm, p<0.01), see Figure 4.2. Using the atlas, the mean DTA improved for the LP but the variability increased, however this was not significant (1.5 ±0.5 v 1.4 ±0.7 mm, p=0.09).
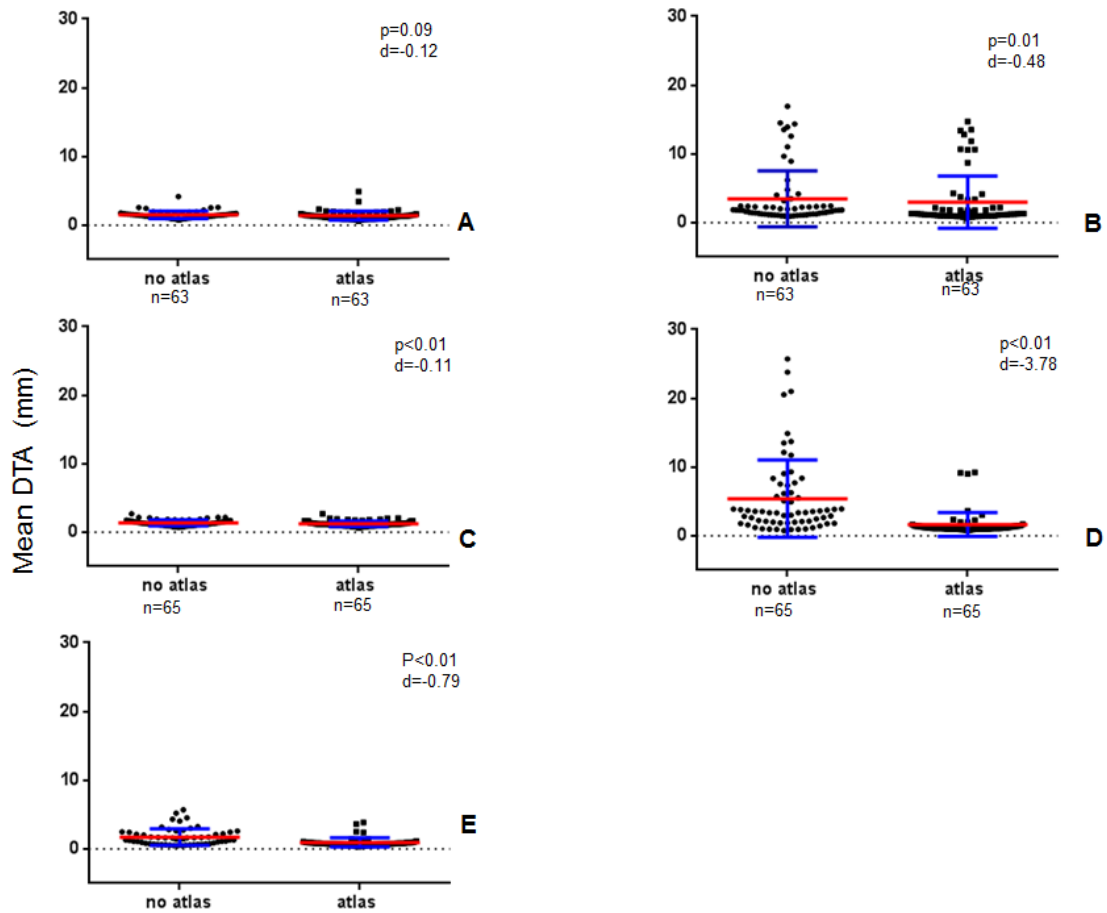
Figure 4.2. Comparison of the mean Distance to agreement (DTA) of all individual clinician manual contours without and with the atlas for the 5 pairs of masticatory muscles.
A. Lateral pterygoid, B. Medial pterygoid, C. Masseter, D. Temporalis, E. Temporo-mandibular Joint

Red horizontal bar illustrates the mean, blue bars illustrate the standard deviation. d equals the difference in absolute mean values.

Figure 4.3 illustrates standard deviation maps on representative slices for the M and T for a single patient. Regions in which there is variation between clinician contours are shown in varying degrees of blue- the darker the shade of blue the larger the variation. The atlas reduced the variation between clinicians for the T particularly at the cranial and caudal aspects of the muscle. The reduction in variability with the atlas was smaller for the M.

101
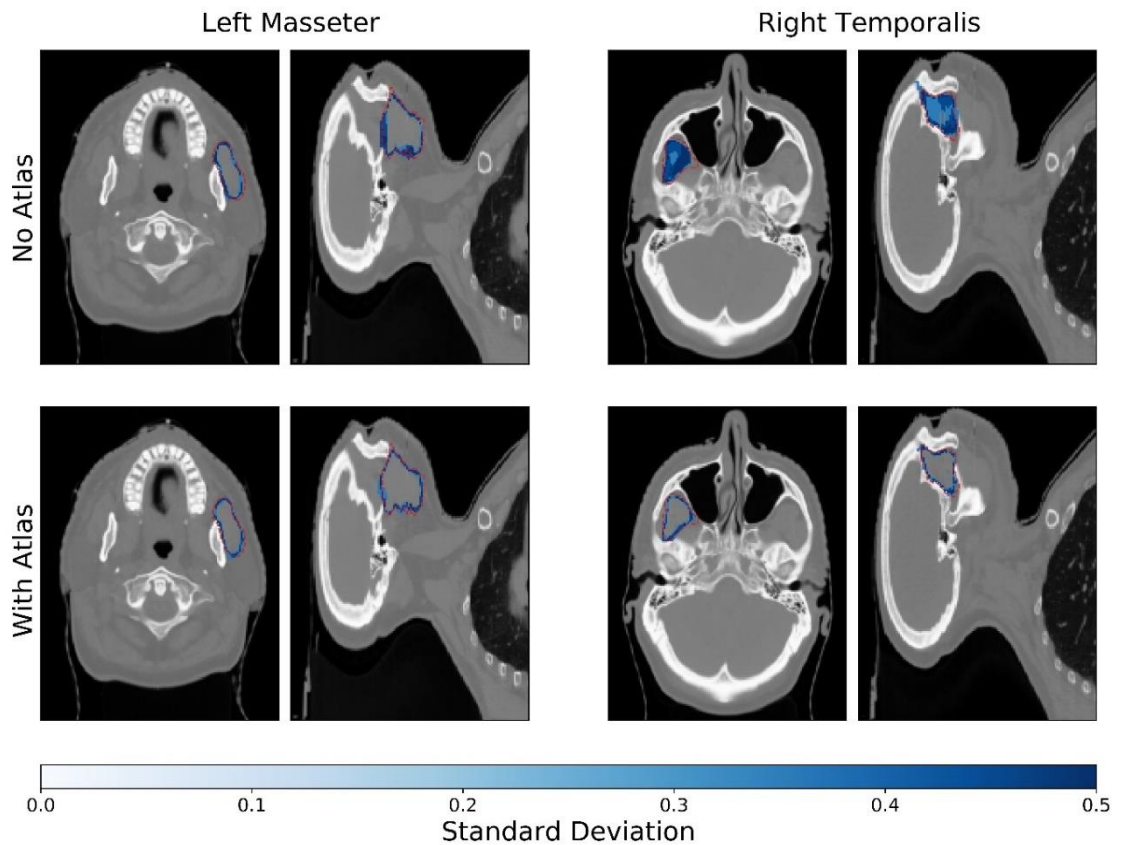
Figure 4.3. Comparison of standard deviation maps.
The variation in structure delineation between clinicians at each voxel without (top row) and with (bottom row) the atlas is shown. The left-hand column shows the standard deviation maps for the masseter, and the right-hand column shows the standard deviation maps for the temporalis. The red contours indicate the reference delineation which were agreed upon by the multi-disciplinary team.

Table 4.1 shows the analysis of the mean and SD DTA without and with the atlas performed according to clinician training grade. An improvement in mean DTA using the atlas was observed by the trainees across all masticatory muscles, with the largest improvement and reduction in variability noted for the T (4.3±7.1 v 1.2±0.4 mm, p=0.06) and TMJ (2.1±0.7v 0.8±0.3 mm, p<0.01).

Table 4.1. Comparison of DTA without and with the atlas between trainee and consultant grade

	Trainees			Consultants		
	Mean ±SD DTA (mm)		p value	Mean ±SD DTA (mm)		p value
	No atlas	Atlas		No atlas	Atlas	
Lateral pterygoids	1.1 (0.5)	1.1 (0.40)	0.06	1.5 (0.6)	1.5 (0.7)	0.68
Medial pterygoids	3.4 (3.9)	2.3 (3.1)	0.34	3.5 (4.2)	3.5 (4.2)	0.97
Masseters	1.3 (0.4)	1.2 (0.3)	0.46	1.4 (0.4)	1.2 (0.4)	0.12
Temporalis'	4.3 (7.1)	1.2 (0.4)	0.06	5.9 (4.8)	1.8 (2.0)	<0.01
TMJs	2.1 (0.7)	0.8 (0.3)	<0.01	1.6 (1.4)	1.0 (0.8)	<0.01

Abbreviations: DTA=distance to agreement; SD=standard deviation; TMJ=temporo-mandibular joint

4.5 Discussion

This prospective study is the first to test the feasibility of a novel atlas of muscles of mastication for contouring in head and neck radiotherapy. Feasibility was defined as a reduction in interobserver variability. This study showed the atlas: (i) improved spatial overlap and alignment of contours for the LP, MP, T and TMJ; and (ii) improved consistency in contouring of all masticatory muscles by the trainees.

Radiation induced trismus is a significant cause of treatment related morbidity (205). The main application of contouring the masticatory apparatus as an avoidance structure is for tumours not infiltrating the muscles in order to spare normal healthy tissue. Clinician variation in OAR contouring may lead to over dosage of the muscles of mastication and as a consequence trismus and poor quality of life. There is overlap between the development of trismus and other health-related quality of life variables. In a paper by Lee et al functional deficits such as taste disturbance, pain, dry mouth and social functioning were increased in patients with trismus (210).

The masticatory muscles are not routinely contoured as avoidance structures. This is partly due to insufficient understanding and consensus on dose-response relationships and a lack of standardized volumes for radiotherapy planning (94,103,104,222). Correct identification of the TMJ as an avoidance structure is important to prevent and delay development of trismus, particularly in nasopharynx cancers (228,229). Improving the standardisation of contouring the MP, T and TMJ will enhance consistency in the study of dose-response relationships (225).

In our study use of the atlas was shown to significantly reduce the mean and SD for DSC and DTA suggesting an improvement in interobserver variability for the MP, T and TMJ. Although the improvement in mean DTA was clinically significant, there was only a small improvement in SD (1.2±0.4 vs 1.4± 0.4 mm) with the atlas observed for the M. Clinicians are more experienced in contouring the M and anatomically it is easier to define. Using the atlas, consistency in contouring did not significantly improve for the LP. An increase in the mean volume was observed for all contoured muscles using the atlas excluding the TMJ. The TMJ is defined by bone limits and was thus contoured using the atlas on the CT bone window. The reduction in TMJ volume and significant improvement in contouring consistency (DTA and DSC) may be explained by the superior visualisation of the bone and joint structures. Using the atlas, improvements in mean DSC and mean DTA were noted for the TMJ (0.6±0.2 vs 0.8±0.1, p<0.01) and (1.7±1.2 vs 0.9±0.7 mm, p<0.01) respectively. The improvement in mean DTA is a large significant change for a small, flat structure. Furthermore, the increase in DSC was despite a reduction in volume with the atlas; DSC is correlated with volume so the large increase observed here was a direct result of the atlas. There is no consensus on which metrics should be used for assessing inter-observer variation, and so for the present study both a volume-based (DSC) and surface-based (DTA) approach were used.

The atlas significantly improved consistency of contours at the muscle extremities in particular the T as shown in Figure 4. There is currently no established dose-response relationship for the T muscle and no consensus as to which masticatory muscles are important in the development of trismus. The T muscle was identified by the MDT team as being an important masticatory muscle and therefore included in the atlas.

Using the atlas, there was a reduction in variability of contours of all muscles by the trainees, in contrast with the consultants that only more consistently contoured the T and TMJ. The improvement in consistency across all muscles by the trainees implies the benefit of the atlas as an educational tool for trainees including dosimetrists. The atlas may also be used by other radiotherapy centres to improve consistency, knowledge and establish collaborations to aid the development of multi-institutional clinical trials.

Development of NTCP models can be facilitated by generating agreement in dose constraint parameters, facilitated by a greater consistency in contouring the muscles of mastication. The reduction in variability in contouring the muscles of mastication may translate into a reduction in variability in reported dose to these structures (230). It is beyond the scope of the present study however to determine the dosimetric effects of reduced clinician inter-observer variation. Future studies to integrate the atlas into an auto-contouring model to reduce inter and intraobserver variability and minimise time constraints should also be considered (109). Improving consistency of contours of the MP, T and TMJ will help standardise volumes, develop more precise dosimetric parameters which can be implemented into avoidance radiotherapy planning to potentially improve radiation related trismus and quality of life (225).

4.5.1 Study limitations

Whilst this is the first paper to our knowledge that uses a novel atlas for the muscles of mastication to evaluate interobserver variability, the study did not explore intra-observer variability or time constraints. The study only involved CT scan images of five patients, however as seven clinicians contoured each plan, a good measure of interobserver variability was obtained. Established dose-response relationships will be facilitated by a more consistent clinician approach to contouring the masticatory muscles. Future studies with greater consistency in contouring and larger numbers are required to further evaluate dose constraints.

4.6 Conclusion

A novel atlas has been developed to contour the masticatory muscles during head and neck radiotherapy planning. The atlas has been shown to significantly reduce interobserver variability for the MP, T and TMJ. The atlas could be considered as an education tool to improve knowledge amongst trainees and provide contouring consistency to aid the development of multi-institutional clinical trials. The atlas has been developed into an app for wider distribution amongst radiotherapy centres. Reducing interobserver variability and standardising treatment volumes will improve the accuracy of avoidance planning and potentially reduce radiation related trismus.

5.0 Evaluation of a combined CT based deep learning auto-contouring model for organ at risk delineation for head and neck radiotherapy

C. Hague[1], A. McPartlin[1], L.W. Lee[1], C. Hughes[1], P. Whitehurst[1], W. Beasley[1], A. Green[2,3], N. Slevin[1], M. van Herk[2,3], C. West[4], R. Chuter[2,3].

[1] Department of Head and Neck Clinical Oncology, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, UK
[2] Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Heath, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
[3] Department of Radiotherapy Related Research, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, M20 4BX, UK
[4] Translational Radiobiology Group, Division of Cancer Sciences, Manchester Academic Health Science Centre, University of Manchester, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, UK

Manuscript in draft

Authors' contribution:

This paper evaluates the use of a novel deep learning CT based auto contouring model for OAR delineation to reduce time and inter observer variability in head and neck radiotherapy planning. For this study I performed OAR contouring, collected and analysed data and wrote the manuscript. Dr Andrew McPartlin, Dr Lip Lee and Dr Chris Hughes helped with contouring head and neck structures. Dr Robert Chuter (physicist) helped with patient selection and data analysis. The CT auto contouring model was developed by Mirada Medical UK.

5.1 Abstract

Introduction. Manually contouring organs at risk (OARs) on computed tomography (CT) images is labour intensive and subject to inter-observer variability. Auto-contouring models can help define volumes, reduce clinical workload and improve inter-observer consistency. This study aimed to develop and evaluate a combined deep learning auto-contouring model for OAR delineation for head and neck radiotherapy.

Methods. Two models trained on local data and one vendor supplied model for OAR delineation on CT images were reviewed by two independent observers. The OARs with the best goodness of fit scores from each model were combined into one model (model$_{CT}$) using deep learning contouring expert software. Model$_{CT}$ was used to produce OAR contours on ten radiotherapy CT planning scans. Four observers recorded the time taken to manually contour each OAR and after a gap of 4 weeks edit model$_{CT}$ contoured OARs. Dice similarity coefficient (DSC) and distance to agreement (DTA) values were compared between manual and edited model$_{CT}$ contours to evaluate inter-observer variability.

Results. Model$_{CT}$ reduced the time for contouring by the four observers by 16% to 63%. A ≥50% time reduction was recorded for the eyes, spinal cord, mandible and submandibular glands. The model reduced inter-observer variability for the left parotid gland (mean DTA 0.93 v 1.69 mm) left submandibular gland (mean DTA 0.78 v 1.53 mm), brainstem (mean DTA 1.30 v 3.42 mm) and larynx (mean DTA 1.92 v 1.96 mm).

Conclusion. An optimised CT deep learning auto-contouring model saves time and reduces inter-observer variability for head and neck OAR delineation for radiotherapy planning. Model$_{CT}$ should improve the adaptive radiotherapy workflow in the treatment of head and neck cancer.

5.2 Introduction

As the number of long term survivors with oropharyngeal cancer increases, there is a need for more research aimed at reducing long-term side-effects that reduce health related quality-of-life (231). To minimise toxicity with intensity-modulated radiotherapy (IMRT), structures need to be adequately delineated (119). The large number and close proximity of organs at risk (OARs) in the head and neck makes manual contouring both time consuming and at risk of significant inter-observer variability compared with other anatomical sites (232).

Auto-contouring models have been shown in published studies to reduce contouring time, clinical workload and inter-observer variability (106,107,110,111), but none have been adopted for use in routine clinical practice. This lack of adoption is partly due to clinicians' concerns about accuracy, automation bias and quality assurance. Novel methods of auto-contouring have been developed over the last decade, the most recent involving artificial intelligence (AI) approaches (115). One advanced form of machine-based AI is the deep learning contouring expert (DLCExpert) developed by Mirada Medical. DLCExpert like other types of deep learning models uses convolutional neural networks to imitate human contours, and it has proved to be more clinically acceptable and require less manual editing than atlas-based auto-contouring (233–235). Using qualitative scoring systems such as the 'goodness of fit' by Greenham et al. can help identify which models produce the most clinically acceptable contours for OARs. OARs can then be combined into one optimised deep learning model for evaluation, which has not previously been studied in head and neck cancer. This study aimed to develop systematically a combined DLCExpert model ($model_{CT}$) and test its ability to reduce time and inter-observer variability with a view to integration into the radiotherapy workflow including adaptive re-planning for the treatment of head and neck cancer.
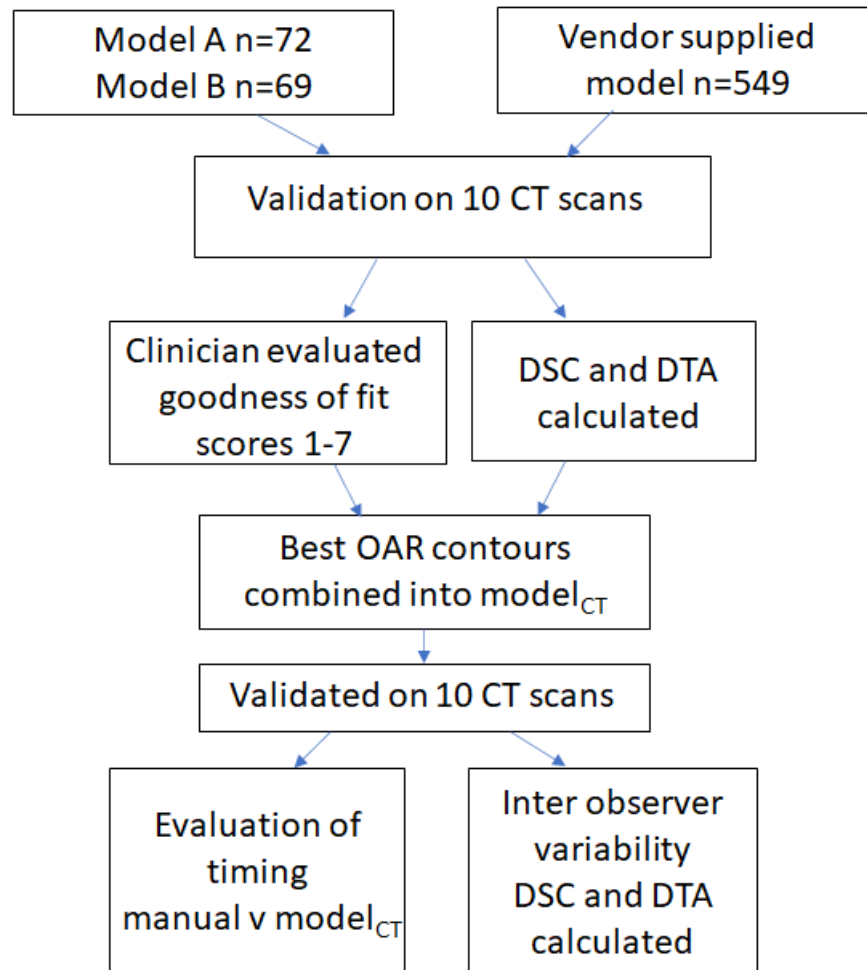
Figure 5.0. Study design schema

5.3 Methods

Figure 5.0 summarises the study design. The study was for service development and did not require ethical approval. Two independent head and neck CT based models (A and B) trained on local data and one vendor supplied (Mirada Medical, Oxford, UK) CT model trained on data from another centre were evaluated. The number of patient datasets used to train the models was 72 for model A, 69 for model B and 549 for the vendor supplied model. Models A and B were developed for other projects and used retrospectively; they were used separately for this project as each had different OARs delineated. OAR contours generated using each model for ten CT scans were reviewed. OARs included: bilateral parotid glands, submandibular glands (SMG), eyes, larynx, oral cavity, mandible, pharyngeal constrictor muscles, brainstem, spinal cord, cricopharyngeal

oesophagus, optic nerves and optic chiasm. Two independent observers scored the contours from 1-7 according to a 'goodness of fit' descriptive approach developed by Greenham et al. (202). The categories in the goodness of fit evaluation ranged from 'good agreement' (score of 1) to 'gross error' (score of 7). A score of 5 or lower was defined as clinically acceptable. A score of 5 was defined by Greenham et al. as moderate manual edits required in 20–50% of slices to meet clinical standards. Contours generated by the three independent models for each OAR were compared with the manual contours created by a consultant and subject previously to qualitative assurance. The agreement of the manual and model contours for each OAR were assessed using qualitative goodness of fit scores and standard quantitative metrics of dice similarity coefficient (DSC) and distance to agreement (DTA). Where goodness of fit scores, DSC and DTA scores conflicted for an OAR, three physicists reviewed the scores and a consensus obtained.

5.3.1 Comparison of timings between manual contours versus editing model$_{CT}$ contours

A combined model$_{CT}$ was developed in DLCExpert (Mirada Medical, Oxford, UK) using the OARs with the best goodness of fit, DSC and DTA scores from each of the three models. Contours generated by model$_{CT}$ were then applied to 10 CT radiotherapy plans from patients with squamous cell carcinoma of the larynx. The demographics of the 10 patients were: 100% T1N0M0 squamous cell cancer of the larynx, 80% male, median age 71.5 (range 60-90) years. T1 larynx plans were used as the tumour tissue is not visualised on T1 larynx scans removing the potential effect of the GTV on OAR visualisation.

Four independent clinicians (two clinical oncology consultants and two registrars) recorded the time taken to manually contour each OAR in Pinnacle v16. After a minimum gap of four weeks each clinician recorded the time taken to edit model$_{CT}$ contours. Mean absolute and percent difference in time were calculated and compared for each clinician and OAR.

5.3.2 Comparison of inter-observer variability between manual and model$_{CT}$ edited contours of OARs

To assess inter-observer variability, each clinician's manual and model$_{CT}$ edited contours were compared with those of the other clinicians. Twelve comparisons were calculated for each of the four OARs for each patient (n=48). Contours were compared for five plans producing a total of 240 comparisons for the four OARs. The manual contour was defined as the 'ground truth' and compared with model$_{CT}$ edited contours using admire and python software. The mean DTA and DSC values were compared between the manual contours and the model$_{CT}$ edited contours for the five CT scans, which demonstrated a proof-of-principle concept.

Table 5.0. Comparison of OAR goodness of fit scores for two observers and 3 models.

OAR	Model A		Model B		Vendor supplied model	
	Observer A	Observer B	Observer A	Observer B	Observer A	Observer B
Brainstem	4.9	4.9	n.a.	n.a.	4.7	4.9
Spinal cord	n.a.	n.a.	3.1	3.4	4.3	4.5
Left parotid	4.2	4.7	4.2	4.6	4.3	4.9
Right parotid	3.9	3.7	3.7	4.4	4.6	5.0
Oral cavity	4.8	5.2	n.a.	n.a.	4.7	4.4
Left SMG	n.a.	n.a.	5.4	5.7	4.7	5.2
Right SMG	n.a.	n.a.	5.8	6.0	5.7	5.6
Mandible	n.a.	n.a.	5.0	5.5	3.0	3.3
Larynx	5.3	5.6	4.6	5.0	n.a.	n.a.
Left eye	3.9	4.3	n.a.	n.a.	n.a.	n.a.
Right eye	4.0	4.2	n.a.	n.a.	n.a.	n.a.
Left optic nerve	7.0	7.0	n.a.	n.a.	n.a.	n.a.
Right optic nerve	6.8	6.8	n.a.	n.a.	n.a.	n.a.
Optic chiasm	7.0	7.0	n.a.	n.a.	n.a.	n.a.
Superior PCM	n.a.	n.a.	5.1	5.7	n.a.	n.a.
Middle PCM	n.a.	n.a.	6.0	6.4	n.a.	n.a.
Inferior PCM	n.a.	n.a.	5.0	5.4	n.a.	n.a.

Abbreviations: n.a.= not available; OAR=organ at risk; SMG=submandibular gland; PCM= pharyngeal constrictor muscle

5.4 Results

Table 5.0 shows that the average goodness of fit scores for models A, B and vendor supplied were similar between the two independent observers. The optic apparatus was only contoured by model A of which it met clinical standards for the left and right eye, but not for the optic nerves or chiasm, where an estimated >50% manual edits were required. Model B outperformed model A for the larynx but did not meet clinical standards for the pharyngeal constrictor muscles. The vendor supplied model met clinical standards for the mandible, oral cavity, brainstem, spinal cord and parotid glands based on the goodness of fit scores.

Table 5.1 compares the mean DTA values obtained for each OAR and model and shows that the vendor supplied model outperformed model B for the mandible (1.6 v 3.1mm), left SMG (3.9 v 4.5 mm) and right SMG (2.6 v 4.0 mm). For brainstem, spinal cord, larynx, bilateral parotid glands and eyes, mean DTA values were better for models A and B than the vendor supplied model. The vendor supplied model was clinically more acceptable on goodness of fit scores for the oral cavity, than model A but mean DTA was worse (9.9 v 5.2 mm). Based on the goodness of fit scores (Table 5.0) and the DTA values (Table 5.1), the best model for delineating each OAR was included in a combined model (model$_{CT}$) as shown in Table 5.2.

Table 5.1. Comparison of mean DTA values for each OAR for each of the 3 models.

OAR	Model A Mean DTA (mm)	Model B Mean DTA (mm)	Vendor supplied Mean DTA (mm)
Brainstem	2.6	n.a.	4.0
Spinal Cord	n.a.	2.1	8.7
Left Parotid Gland	3.3	3.5	3.1
Right Parotid Gland	2.8	2.5	2.5
Oral Cavity	5.2	n.a.	9.9
Left SMG	n.a.	4.5	3.9
Right SMG	n.a.	4.0	2.6
Mandible	n.a.	3.1	1.6
Larynx	3.5	4.9	n.a.
Left Eye	1.5	n.a.	n.a.
Right Eye	1.8	n.a.	n.a.

Abbreviations: n.a.=not available; SMG= submandibular gland; DTA= distance to agreement

Table 5.2. OARs taken from each model into the combined model$_{CT}$.
The OARs chosen are outlined in black.

OAR	Model A	Model B	Vendor supplied
Brainstem	■		
Larynx	■		
Left parotid	■		
Right parotid		■	
Oral cavity	■		
Spinal cord		■	
Mandible			■
Left SMG			■
Right SMG			■
Left eye	■		
Right eye	■		

Abbreviations: OAR=organ at risk; SMG=submandibular gland

5.4.1 Comparison of timings for manual versus model$_{CT}$ edited contouring

The mean absolute and percent differences in time taken for manual versus model$_{CT}$ edited contouring for the four independent observers were: 23 min 1s v 10 min 2s (54%); 19 min 1s v 7 min 1 s (63%); 14 min 9 s v 12 min 2 s (16%); and 21min 4 s v 10 min 2 s (52%). Table 5.3 summarises these differences in timings.

Table 5.3. Difference in times for manual versus model$_{CT}$ edited contouring for multiple OARs and four independent clinicians.

OAR	Absolute (%) reduction in time (s) for model$_{CT}$ edited v manual contouring				
	Observer 1	Observer 2	Observer 3	Observer 4	Mean
Brainstem	73 (72%)	75 (63%)	45 (42%)	9 (13%)	50 (46%)
Spinal cord	163 (72%)	100 (66%)	68 (58%)	56 (51%)	96 (62%)
Larynx	61 (41%)	39 (44%)	+11 (11%)	80 (39%)	47 (34%)
Left parotid	52 (46%)	71 (57%)	18 (20%)	85 (52%)	56 (45%)
Right parotid	50 (46%)	60 (47%)	+1 (1%)	103 (54%)	53 (37%)
Mandible	207 (80%)	88 (68%)	13 (13%)	93 (75%)	100 (59%)
Oral cavity	17 (11%)	39 (39%)	+53 (67%)	48 (29%)	39 (3%)
Left SMG	37 (47%)	63 (78%)	27 (41%)	54 (58%)	45 (56%)
Right SMG	30 (40%)	64 (74%)	22 (38%)	35 (49%)	37 (50%)
Left eye	41 (64%)	68 (89%)	17 (36%)	55 (90%)	45 (70%)
Right eye	44 (69%)	60 (90%)	17 (35%)	31 (74%)	38 (67%)

Abbreviations: OAR=organ at risk; SMG=submandibular gland
Yellow boxes highlight increased time for model$_{CT}$ edited versus manual contouring.

114

The largest mean % reduction in time obtained using model$_{CT}$ across all four observers was seen for the left eye (70%), right eye (67%), spinal cord (62%) and mandible (59%). Model$_{CT}$ saved less time to contour the oral cavity (3%) across all four clinicians. An increase in absolute (%) contouring time using model$_{CT}$ contouring was recorded for the larynx of 11 s (11%); right parotid gland 1 s, (1%) and oral cavity, 53 s, (67%) by Observer 3.

5.4.2 Comparison of inter-observer variability between model$_{CT}$ edited and manual contours for OARs

The reliability of contouring was improved using model$_{CT}$. There was a statistically significant increase in mean DSC for all structures except the larynx (Table 5). An improvement in mean DTA and reduction in standard deviation (SD) was observed for the brainstem, left parotid and left SMG, but only reached statistical significance for the left SMG (Table 5.4, Figure 5.1).

Table 5.4. Comparison of mean DSC and mean DTA between manual and model$_{CT}$ contours for OARs across the 5 CT plans.

	Mean DSC			Mean DTA (mm)		
	Manual	Model$_{CT}$	p value	Manual (±SD)	Model$_{CT}$ (±SD)	p value
Brainstem	0.78	0.93	0.00	3.42 (2.72)	1.30 (0.48)	0.09
Larynx	0.79	0.82	0.22	1.96 (0.64)	1.92 (0.62)	0.93
Left parotid	0.84	0.92	0.04	1.69 (0.51)	0.93 (0.35)	0.07
Left SMG	0.79	0.89	0.02	1.53 (0.84)	0.78 (0.28)	0.00

Abbreviations: SMG= submandibular gland, DSC= dice similarity coefficient; DTA= distance to agreement; SD= standard deviation; OAR=organ at risk
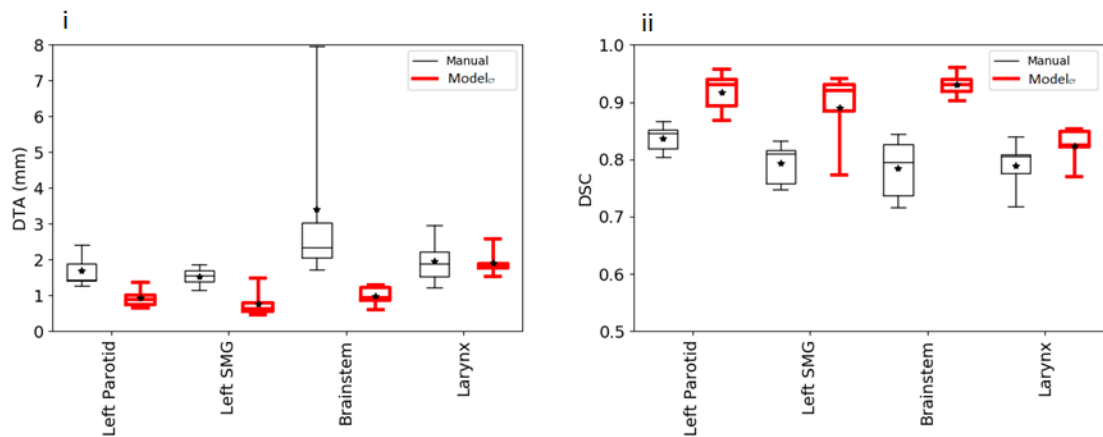
Figure 5.1. Comparison of manual and modelct edited contours across observers using measured (i) DTA and (ii) DSC values for the OARs.
Box and whisker plots illustrating the overall difference between the manual and modelct edited contours across all 4 observers, using DTA (i) and DSC (ii) values for the left parotid, left SMG, brainstem and larynx. Asterisks indicate the mean value of DTA and DSC measurements. Horizontal lines are the median, whiskers are the maximum and minimum values and box shows the interquartile ranges.

5.5 Discussion

This study is the first to develop an optimised deep learning auto contouring model modelct that combines the best OAR contouring obtained using three independent models. A combination of qualitative and quantitative scoring methods assessed the clinical usefulness of each model in line with previous studies (117). Using only quantitative metrics for volume (DSC) and distance (DTA) to assess the similarity of contours may not truly reflect a model's performance. For example, in this study, the vendor supplied model generated auto contours that were clinically acceptable for the oral cavity based on the goodness of fit scores (≤5) but not based on mean DTA (9.9 mm). This discrepancy highlights the increased subjectivity of qualitative analysis, which if used alone may introduce bias (236).

In this study, modelct reduced contouring time and inter-observer variability in head and neck radiotherapy planning. Editing the modelct assisted contours reduced the average contouring time by up to 63%, equating to an absolute time saving of 12 minutes per patient by the most experienced clinician. When divided by OAR, the largest percentage time saving using the edited modelct contours

116

was observed for the left (70%) and right (67%) eyes with time savings of ≥50% recorded for the spinal cord, mandible and SMGs. These findings are supported by a study by Wong et al. where unedited auto contours generated by deep learning commercial software were compared with manual contours for selected head and neck OARs and took 0.6 v 26.6 minute per plan (106). Adaptive re-planning is time consuming and has the potential to cause treatment delays. The improved speed and reduction in variability of edited auto contours versus manual has a potential role in adaptive radiotherapy to reduce workload, standardise volumes and improve the accuracy of the dose delivered.

Despite the large amount of published data highlighting the benefits of auto contouring models to improve contouring time, they are not routinely used in clinical practice. Most of the auto-contouring models studied are atlas based, trained on small data sets within a single institution. The models do not, therefore, represent the natural variation seen amongst patients. ABAS uses edge detection to define a structure, which is difficult if the grayscale contrast is low and means time-consuming post-segmentation manual edits are often required (237). In a study by Teguh et al. ABAS generated auto contours for 30 structures took on average 7 minutes per patient, compared with up to 66 minutes per patient for manually edited auto-contours (109). Model$_{CT}$ developed in this study is the first deep learning model developed by combining information from selecting the best OARs using models and data generated from more than one institution. Using deep learning-based auto contouring models reduces the effects of artefacts such as from metal implants and allows for adaptation to changes in anatomy secondary to tumour volume, shape and treatment position. In a recent comparative study of ABAS versus DLCExpert for head and neck OAR delineation, DLCExpert contours were superior for all glandular structures and less prone to error. This superiority compared with manual contouring was due to reducing differences in doses and inter-observer variability for all OARs (117).

Auto contouring models do not remove operator variability, however, and are still prone to clinician bias and personal preference. In this study, observer 3 recorded the smallest percent reduction in time (16%) equating to a mean absolute time saving of 2.4 minutes per patient using model$_{CT}$ contours. Observer 3 spent longer editing model$_{CT}$ contours for the larynx, oral cavity and right parotid gland, but

overall was the quickest to manually contour OARs compared with the other observers. Observer 3 was a consultant clinical oncologist appointed within the last 2 years and so was more experienced than observers 1 and 4 who are registrars, but less experienced than observer 2. The benefit of auto contouring models to reduce contouring time may be attenuated dependent on user experience but are still better due to improved consistency of contouring. The model's performance also varied by OAR. Overall, the time taken to edit model$_{CT}$ contours was similar to the time taken to manually contour the oral cavity with an average percentage saving of only 3%. This highlights the variability of contouring the oral cavity on CT by the model and manually, requiring more than the average number of edits. The effects of dental artefact on image quality may affect the sensitivity of the model to recognise tissue boundaries leading to delineation inaccuracy. Proper selection and validation of an auto contouring model for use in clinical practice is therefore needed to minimise this risk.

In this study, using model$_{CT}$ edited contours reduced inter-observer variability. High mean DSC values were obtained for the brainstem (0.92), left parotid gland (0.93) and left SMG (0.89). These results are supported in a study by Van Dijk et al. where a similar mean DSC using a deep learning auto contouring model for the parotid and SMGs (0.81) was reported (117). Auto contouring models can help reduce the large inter-observer variability observed at the extremities of structures (227). A large reduction in variability using the model$_{CT}$ edited contours for the brainstem were reported with a mean and standard deviation for DTA (1.30 ± 0.48 v 3.42 ± 2.72 mm, p=0.09) compared with manual contouring. Future research should explore how improvements in the quality of imaging obtained with MRI might improve recognition of tissue boundaries and reduce the variability at the extremities of structures. Delineation atlas based guidelines may also help reduce variability at the extremities of structures, but they rely on clinician compliance and using a single validated guideline (238). Improving consistency of OAR contouring will help standardise volumes, improve the accuracy of the dose delivered to maintain adequate tumour coverage whilst minimising the risk of normal tissue toxicity. Further work is needed to assess the quality of the contours from model$_{CT}$ against manual contours with a larger library of OARs including the pharyngeal constrictor muscles and optic apparatus. This will increase the robustness of validating the model for use in clinical practice.

The study has limitations. Although trained on 690 data sets, the ability of model$_{CT}$ to improve inter-observer variability needs evaluating in a larger number of patients and with an increasing library of OARs. OARs such as the pharyngeal constrictor muscles and optic apparatus were excluded from model$_{CT}$, due to the poor performance of the three individual models to generate adequate contours. The assessment of goodness of fit scores was only performed by two observers which may introduce clinician bias and fail to capture the variability of OAR segmentation. In this study where the goodness of fit and mean DTA scores conflicted for an OAR, three physicists reviewed the contours and relevant scores and a consensus decision was made. This is a potential limitation of the study which may have been improved by agreeing which scoring method, based on its clinical importance should be used, prior to generating the results. The model with the best overall score could then be used. To evaluate inter-observer variability, manual and model$_{CT}$ edited contours for the left parotid gland, left SMG, brainstem and larynx were compared between the four observers. This resulted in 48 comparisons per patient which across 5 plans produced a total of 240 comparisons to be analysed. This analysis provided a proof of principle concept which requires validation in a larger cohort. Further work with a larger set of OARs including the pharyngeal constrictors muscles and optic apparatus will help validate model$_{CT}$ prior to its use in clinical practice.

In this study model$_{CT}$ contours required manual editing prior to being accepted clinically. Post-segmentation manual editing is similar to that used in ATLAS based auto segmentation studies (113,119). From a review of the literature there are currently no fully automated contouring models which do not require human quality assurance. Without human input, auto contouring models may interpret information differently, which is clinically irrelevant and may introduce automation bias as described by Skitka et al. both of which have the potential to increase the risk of missing errors (239). The need to manually edit contours generated by an auto contouring model is time consuming and may disrupt the adaptive re-planning workflow. To improve this potential limitation of auto contouring models, better delineation accuracy using MR imaging may be an option (133). The superior soft tissue visualisation of MR compared with CT can recognise tissue boundaries and improve the model's performance to define smaller, less well defined structures, such as the optic apparatus and pharyngeal constrictor muscles as described

above (240). The combined advantages of MRI and "deep learning methods" are a potential area of future work to improve the integration of auto contouring models both into treatment planning and the adaptive radiotherapy workflow of head and neck cancer.

5.6 Conclusion

In this feasibility study, an optimised deep learning CT based auto contouring model was shown to reduce time and inter-observer variability. The model is worth further validation. Integration of such models into clinical practice should improve clinician workload, help standardise volumes and may have a role in improving the adaptive radiotherapy workflow in the treatment of head and neck cancer.

6.0 An MR based deep learning auto-contouring model for planning head and neck radiotherapy

C. Hague[1], A. McPartlin[1], L.W. Lee[1], C. Hughes[1], D. Mullan[2], W. Beasley[1], A. Green[3,4], G. Price[3,4], P Whitehurst[1], N. Slevin[1], M. van Herk[3,4], C. West[5], R. Chuter[3,4].

[1] Department of Head and Neck Clinical Oncology, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, UK
[2] Department of Radiology, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, UK
[3] Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Heath, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
[4] Department of Radiotherapy Related Research, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, M20 4BX, UK
[5] Translational Radiobiology Group, Division of Cancer Sciences, Manchester Academic Health Science Centre, University of Manchester, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, UK

Manuscript in draft

Authors' contribution:

This paper highlights the systematic development of the first MR deep learning auto contouring model for OAR delineation in head and neck radiotherapy planning. I coordinated the study, identified the patients, transferred data onto a planning platform and performed contouring. I collected, analysed and interpreted the data and drafted the manuscript. Dr Gareth Price and Dr Andrew Green helped with identification and transfer of data via the theragnostics system. Dr Damian Mullan (consultant radiologist with 15 years' experience) peer reviewed all manual contours prior to model development. Mirada medical software, UK developed the model and liaised with Mr Andrew Green to import the model into workflow box for further analysis. Dr Andrew McPartlin identified MR-Linac scans and with Dr Robert Chuter supported the data analysis and drafting the manuscript.

6.1 Abstract

Introduction. Accurate organ at risk (OARs) delineation improves radiation dose delivery and minimises normal tissue toxicity. Auto contouring models help define volumes and reduce clinical workload. This study aimed to develop and evaluate a novel Magnetic Resonance (MR) deep learning auto contouring model for OAR delineation in head and neck radiotherapy.

Methods. Two auto contouring models were developed in deep learning contouring expert (DLCExpert) for OAR delineation: an optimised CT model (model$_{CT}$) and a novel MR model (model$_{MRI}$). The models were trained to generate auto contours for the bilateral parotid glands and submandibular glands. Auto-contours were generated by model$_{MRI}$ on 10 T2 weighted (W) 2-dimensional (2D) diagnostic, 10 T2W dixon 2D MR radiotherapy planning, eight 3D MR Linac (MRL) T2 scans and, by model$_{CT}$, on 10 CT planning scans. Comparisons were made with the ground truth, defined as manual contours from the same clinician peer reviewed by a consultant radiologist. Dice similarity coefficient (DSC) and distance to agreement (DTA) were calculated for comparison.

Results. Model$_{MRI}$ contours improved the mean DSC and DTA compared with manual contours for the bilateral parotid glands and submandibular glands on the T2W 2D diagnostic and T2W Dixon 2D MR radiotherapy planning scans. Due to contrast differences with the training set, improvements are needed in the 3D MRL sequence. There was a trend to improvement with statistically significant differences seen for model$_{MRI}$ compared to model$_{CT}$ for the left parotid (mean DTA 2.3 v 2.8mm), right parotid (mean DTA 1.9 v 2.7 mm), left submandibular gland (mean DTA 2.2 v 2.4 mm) and right submandibular gland (mean DTA 1.6v 3.2 mm).

Conclusion. A novel deep learning MR auto-contouring model shows promise for OAR auto-contouring but needs validating in a larger cohort. Performance is affected by the method of MR acquisition and further work is needed to improve its use with MRL images.

6.2 Introduction

Accurate volume delineation is a prerequisite to delivering conformal doses to target structures whilst maintaining sufficient organ sparing (117). The large number and complex anatomy of structures in the head and neck makes manual contouring labour intensive and subject to inter and intra-observer variability (227). Atlas based auto contouring models reduce time and inter-observer variability, but none have been adopted for widespread use (241). One reason for the lack of adoption is the poor imaging contrast of computer tomography (CT) that requires clinicians to manually edit auto contours before they are clinically acceptable (113,119). Magnetic resonance (MR) is expected to be superior to CT for auto contouring in the head and neck and improve the accuracy of OAR delineation (131).

Auto-contouring tools have developed over the last decade, the most recent involving artificial intelligence (AI) approaches. Compared with atlas-based auto-contouring, deep learning using convolutional neural networks to imitate human contours that are found to be more acceptable clinically and require less manual editing (233–235). CT based deep learning models perform well for structures with well-defined boundaries, such as the spinal cord and mandible, but less so for glandular structures such as the parotids and submandibular glands (SMG) (242). MR imaging is a promising addition to AI models, due to its superior soft tissue discrimination (243–245). Wardman et al. highlighted the improved contouring accuracy using MR atlas based auto contours compared with CT for OAR delineation(133). There is currently, however, no MR deep learning model available for use in head and neck OAR contouring.

In head and neck radiotherapy planning different types of MR image acquisitions are used, which makes the development of a generic MR auto contouring model challenging. One of the areas where AI auto-contouring may have a role to improve contouring accuracy and clinician workload is MR image guided radiotherapy (MRIgRT). Anatomical and geometric changes throughout a patient's treatment course can result in over and under dosage of normal tissue and tumours. MRIgRT is able to capture inter and intra fraction motion due to tumour response or anatomical change and facilitate real time adaptation to optimise dose

delivery and improve outcome (246,247). Repeatedly re-contouring structures, however, is labour intensive and subject to error. Integration of auto-contouring models into the adaptive radiotherapy workflow for use in MRIgRT may save time and improve contouring accuracy by producing a more standardised volume (109). This study aimed to evaluate a novel MR based model for OAR delineation for future use in adaptive radiotherapy to improve the efficiency of the online workflow.
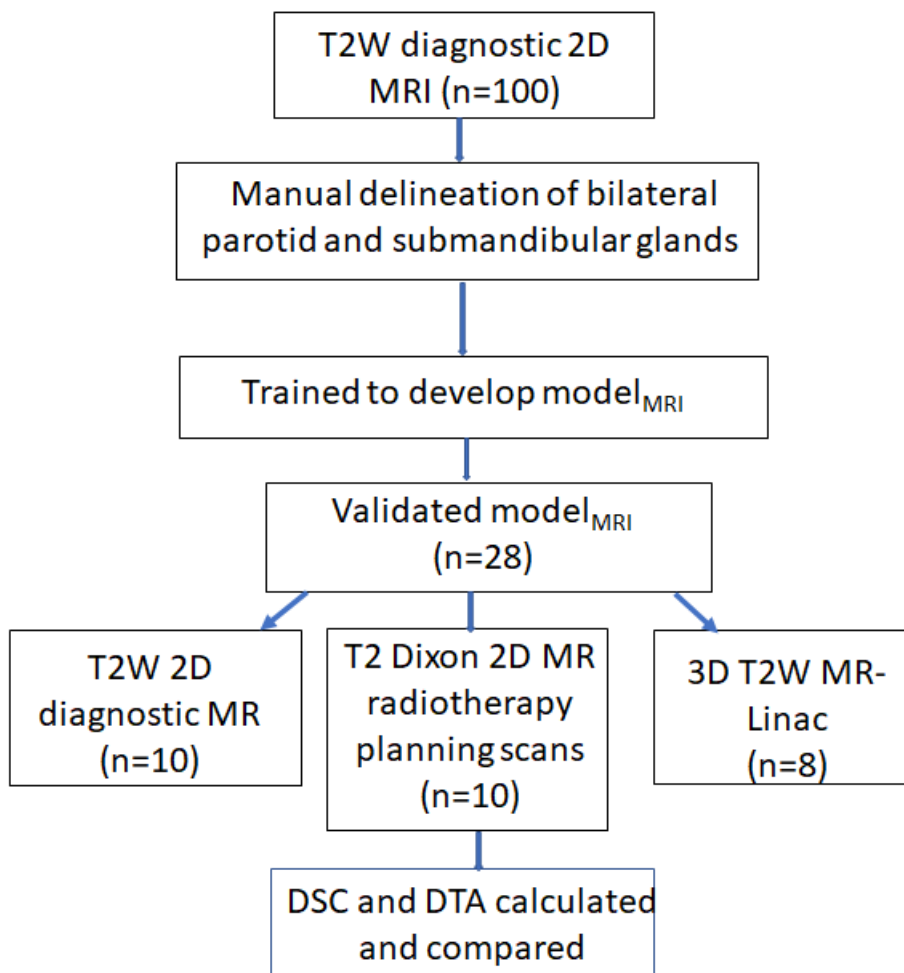
Figure 6.0. Study design schema

6.3 Methods

Figure 6.0 summarises the study design. The study was for service development and did not require ethical approval. A novel deep learning MR auto contouring model was developed, termed model$_{MRI}$. Bilateral parotid glands and submandibular glands (SMGs) were contoured by the same clinician on one hundred anonymised oropharynx T2 weighted (T2W) 2 dimensional (2D) diagnostic MR scans. To ensure adequate visualisation of structures, oropharynx scans were excluded if the images contained artefacts or there was tumour involvement of the parotid or SMGs. Oropharynx scans were used due to the ease of obtaining enough numbers for model development. Contours were peer reviewed by a consultant radiologist prior to being exported to deep learning contouring expert (DLCExpert) imaging software for model development. Model$_{MRI}$ was subsequently developed by Mirada Medical, Oxford, UK. The performance of model$_{MRI}$ was tested on three different types of MR scans: 10 T2W 2D diagnostic MR scans, 10 T2W Dixon 2D MR radiotherapy planning scans and eight 3D MR-Linac (MRL) T2 scans (Elekta, Stockholm, Sweden) scans. The MR parameters for each scan are summarised in Table 6.0. An optimised CT auto contouring model reported in Chapter 5 was also tested termed model$_{CT}$. Model$_{CT}$ was developed using the best OAR contours taken from two in house models and one vendor supplied model. Model$_{CT}$ automated contours were generated for CT radiotherapy planning scans for the same 10 patients who were automatically contoured using model$_{MRI}$ on 10 T2W Dixon 2D MR radiotherapy planning scans.

The same clinician manually contoured bilateral parotid glands and SMGs on all the MR and CT scans – defined as the 'ground truth'. OAR contours were compared using goodness of fit scores, Dice similarity coefficient (DSC) and distance to agreement (DTA) measurements. The statistical significance of t-tests was defined as $p < 0.05$.

Table 6.0. Summary of MR parameters.

Parameter	T2W 2D diagnostic			T2W Dixon 2D planning scan		T2W 3D MR-Linac	
Manufacturer	Siemens	Siemens	Philips	Philips	Philips	Philips	Philips
Model	Aera	Amira	Achieva	Ingenia	Ingenia	Marlin	Marlin
Series description	t2_tse_tra 320	t2_tse_tra_3 84	T2W_TSE	T2m Dixon 3 mm	T2m Dixon 3 mm	T2 3D Tra 2 min	T2 3D Tra 2 min
Slice thickness (mm)	3	4	3	3	3	2	2
Repetition time (s)	4630	6060	4977	3352	3352	1535	1535
Echo time (s)	80	89	100	70	70	278	278
Field strength (T)	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Pixel size (mm)	0.9	0.3	0.5	0.6	0.6	0.8	0.8

Abbreviations: T2W=T2 weighted; 2D= 2 dimensional; 3D= 3 dimensional; MR= magnetic resonance

6.4 Results

6.4.1 Comparison of manual contours and model$_{MRI}$

Model$_{MRI}$ automated contours showed good agreement with manual contours defined on T2W 2D diagnostic and T2W Dixon 2D radiotherapy planning scans for bilateral parotid glands and SMGs (mean DSC ≥0.80). The agreement was lower for model$_{MRI}$ automated contour versus manual contours for the T2W 3D MRL for the bilateral parotid glands (mean DSC 0.70). There was a lack of overlap between model$_{MRI}$ automated contours and manual contours on the MRL images for the left and right SMG (mean DSC of 0.10 and 0). The overlap of manual and model$_{MRI}$ automated contours are illustrated in Figure 6.1.
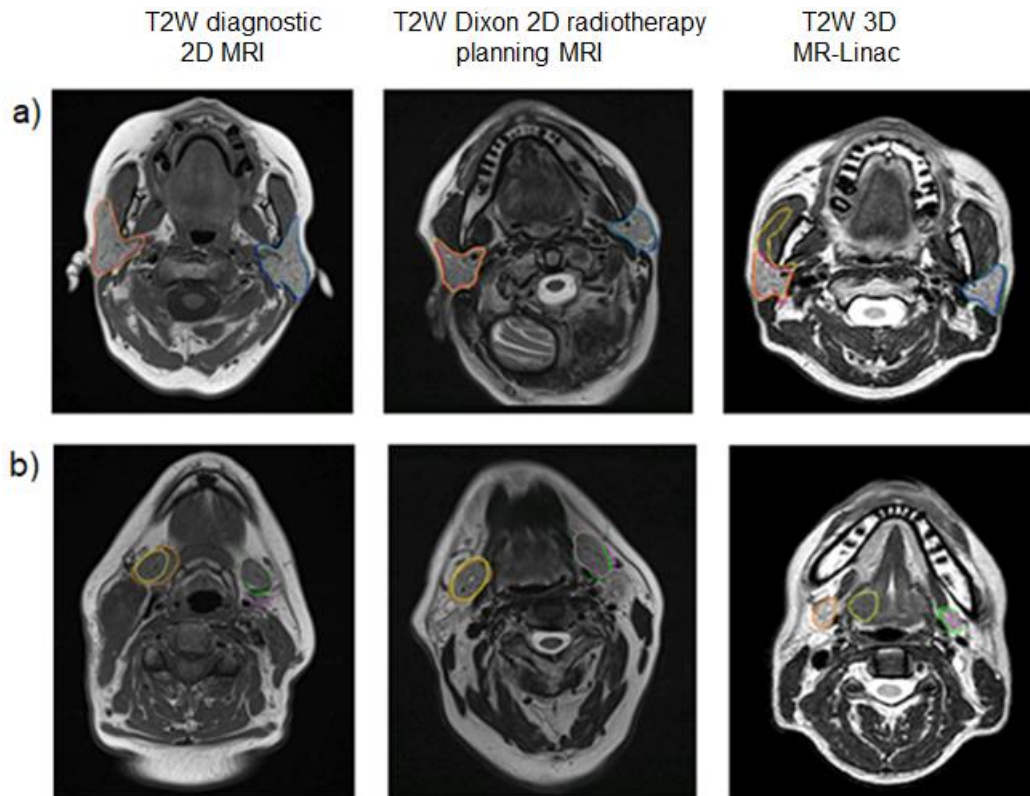
Figure 6.1. Spatial overlap of contours between model$_{MRI}$ automated contours and manually drawn contours for the 3 types of MR. (a) right and left parotid glands (b) and submandibular glands
Manual contours: left parotid gland =blue, right parotid gland =pink, left submandibular gland = green, right submandibular gland= orange.
Model$_{MRI}$ auto contours: left parotid gland=green, right parotid gland= yellow, left submandibular gland = pink, right submandibular gland = yellow.
Abbreviations: T2W= T2 weighted; 2D= 2 dimensional; 3D= 3 dimensional; MR= magnetic resonance

Table 6.1. Performance of model$_{MRI}$ automated versus manual contours.

OAR	T2W 2D diagnostic n=10		T2W 2D planning n=10		3D MR-Linac n=8	
	DSC	DTA (mm)	DSC	DTA (mm)	DSC	DTA (mm)
Left parotid gland	0.83	1.80±0.40	0.82	2.30±0.50	0.72	2.70±1.50
Right parotid gland	0.83	1.80±0.50	0.81	1.90±0.50	0.69	5.60±8.80
Left SMG	0.79	2.10±1.90	0.79	2.20±3.40	0.09	21.40±22.90
Right SMG	0.79	1.60±0.60	0.81	1.60±0.50	0	18.30±13.80

Abbreviations: Values are mean ± standard deviation for the comparisons between the 3 types of MR. 2D=2 dimensional; 3D= 3 dimensional; MR-magnetic resonance; DSC=dice similarity coefficient; DTA= distance to agreement; OAR= organ at risk; SMG=submandibular gland.
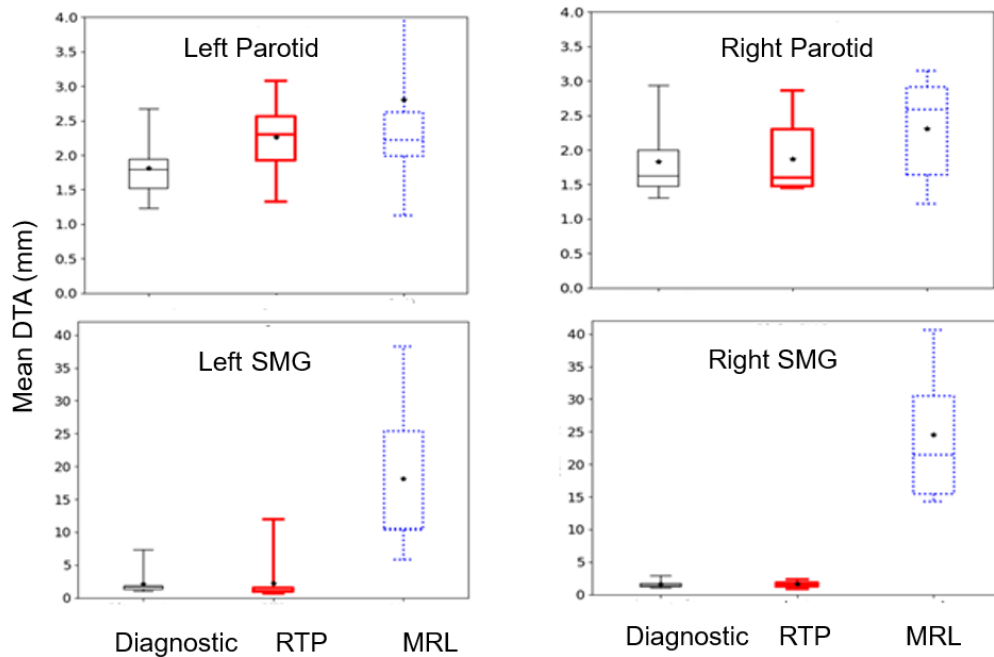
Figure 6.2. Comparing model$_{MRI}$ automated and manually drawn contours for the bilateral parotid and submandibular glands using DTA values. Model$_{MRI}$ data were obtained on three types of MR imaging: diagnostic, radiotherapy treatment planning (RTP) and MR-linac (MRL).
Asterisks show mean values, horizontal lines the median, whiskers the maximum and minimum values and boxes show the interquartile range.
Abbreviations: RTP= radiotherapy treatment planning; MRL=magnetic resonance linear accelerator; DTA= distance to agreement; SMG= submandibular gland

6.4.2 Comparison of model$_{MRI}$ and model$_{CT}$

Model$_{MRI}$ contours showed better agreement with manual contours on MR radiotherapy planning scans compared with model$_{CT}$ contours on CT radiotherapy planning scans. An improvement in mean DTA and reduction in SD for bilateral parotid and SMGs was observed except for the left SMG which was similar between the two modalities (Table 6.2, Figure 6.3). The mean DSC values to compare the model and manual contours were similar for the bilateral parotid and left SMGs irrespective of the model used (mean DSC >0.7). The only exception was the right SMG, which had a lower mean DSC when using model$_{CT}$ versus model$_{MRI}$ assisted contouring (0.6 v 0.8).

Table 6.2. Comparison between manual versus model$_{CT}$ contouring with manual versus model$_{MRI}$ contouring on radiotherapy planning scans.

OAR	Model$_{CT}$ v manual CT (n=10)		Model$_{MRI}$ v manual MR (n=10)	
	DSC	DTA (mm)	DSC	DTA (mm)
Left parotid gland	0.75	2.80±1,00	0.82	2.30±0.50
Right parotid gland	0.73	2.70±0.90	0.81	1.90±0.50
Left SMG	0.74	2.40±1.10	0.79	2.20±3.40
Right SMG	0.64	3.20±2.60	0.81	1.60±0.50

Abbreviations: Values are mean ± standard deviation for 10 comparisons
OAR= organ at risk; DSC=dice similarity coefficient; DTA= distance to agreement;
SMG=submandibular gland.
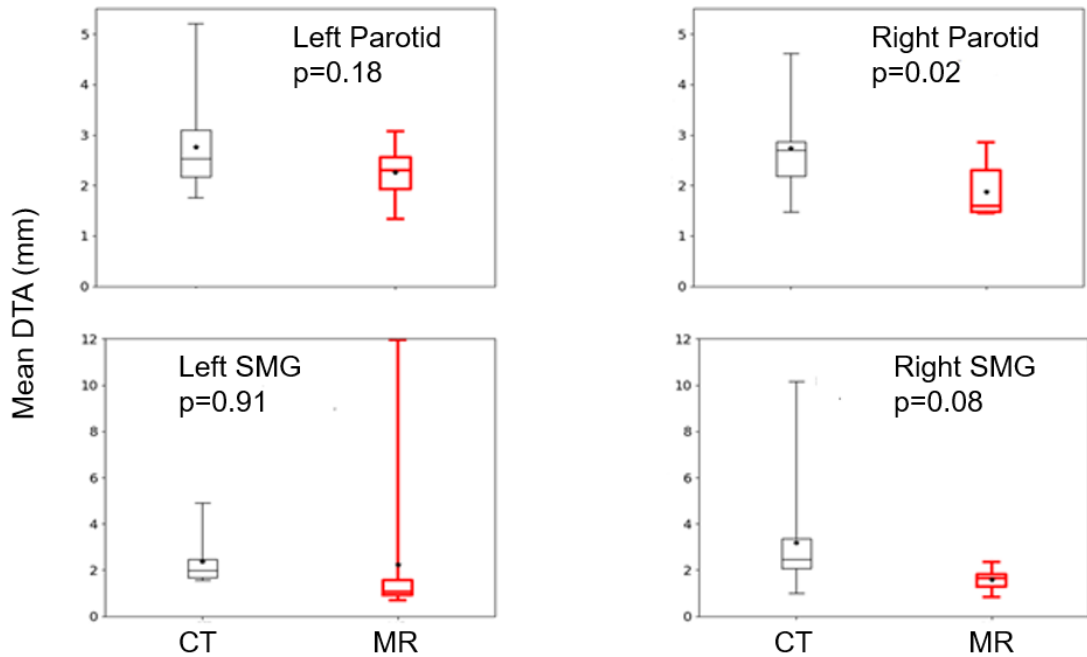
Figure 6.3. Box and whisker plots comparing model_MRI versus manual contours on ten T2W Dixon 2D MR radiotherapy planning scans and model_CT versus manual contours on ten CT radiotherapy planning scans.
Asterisks show mean values, horizontal lines the median, whiskers the maximum and minimum values and boxes show the interquartile range.
Abbreviations: DTA= distance to agreement; CT= computer tomography; MR= magnetic resonance; SMG= submandibular gland

6.5 Discussion

Auto contouring models have been studied to reduce contouring time, resources and inter-observer variability (40,241). The most widely studied models are atlas-based (AB), which have limitations. AB auto contouring models use contours from a 'reference library' and apply them to a new patient image. They are limited by the range and number of contours in a reference library and are unable to adapt to anatomical differences between the atlas and the structure to be delineated. More recently the use of AI is a promising approach for overcoming these limitations. Deep learning convolutional networks such as DLCExpert are less affected than AB by anatomical variations between images and artefacts such as metal (248,249). AI models do not rely upon edge detection to identify structures and therefore have the potential to perform better in areas of low grayscale contrast such as in the head and neck (237). Several studies have shown the potential benefits of using CT based

deep learning convolutional networks for head and neck contouring, but as yet these approaches have not have gained widespread adoption (106,117,250).

This study is the first to combine and test the validity of MR imaging with deep learning convolutional networks (DLCExpert) for head and neck OAR auto contouring. Model$_{MRI}$ performed well compared with manual contours when applied to T2W 2D diagnostic and radiotherapy planning scans, with a mean DSC score of 0.80, and DTA of ≤2mm. A similar solution trained on 713 multiparametric MR images by Lin et al. showed edited model-generated auto contours improved contouring accuracy (mean DSC of 0.79) and reduced within (36.4%) and between (54.5%) observer variability compared with manual delineation for GTV contouring in nasopharynx cancer (132). The current study was trained on a smaller number of image sets and yet performed well. The optimal number of datasets to train and validate a deep learning convolutional network model is not well understood. In some of the published studies to date, the size of training datasets range between 142-713 cases with validation sets of over 100 scans (117,132,233). Model$_{MRI}$ performed well with similar geometric accuracy to manual contours and yet was trained and tested on 100 training and 28 validation datasets. The ability to generate an effective model with a relatively small training cohort may be explained by the high standard of contouring and consistent MR sequences used to produce the model. The OARs were prospectively contoured for this study following consensus guidelines and independently reviewed by a consultant radiologist with twelve years' experience of head and neck radiology. The training dataset images involved the same sequence (t2-tse-tra). The potential effects of factors on image quality such as patient positioning, artefacts and contrast were reviewed with images excluded from the dataset if heavily distorted either by artefacts or tumour infiltration. The size of the dataset needed to train a model is an area of current research, but as illustrated by this study there may be a trade-off between dataset size and consistency.

Model$_{MRI}$ performed well when applied to T2W diagnostic and radiotherapy planning scans, due to the training and testing datasets being of similar image acquisition parameters. When applied to the T2W 3D MRL images, model$_{MRI}$ generated contours were acceptable for both parotid glands but showed partial

or complete lack of overlap with manual contours for the left and right SMGs (mean DSC 0.1, 00) and mean DTA (18 mm and 24 mm), respectively. One potential explanation for this is the difference in image sequence and grayscale contrasts for the MRL compared to the dataset the model was trained on. Deep learning models are trained by extracting the image features it thinks are most important based on voxel intensities. Compared with CT, there is no quantification metric such as the Hounsfield unit used in MR to measure voxel intensity, which may be a potential limitation when using MR based models. The deep learning model will predict structures relative to the data it was trained on, therefore if this data (or intensity) differs, the model's performance will also be affected. In addition, if the structure to be delineated is small the model may fail to delineate it.

In our study, model$_{MRI}$ failed to generate contours for the right SMG on one of the 3D MRL scans. Before generating a contour, the model estimates the probability that each voxel belongs to a structure which has also been termed as an "activation threshold" as described by van Dijk et al. (117). A threshold probability or activation threshold of >0.5 is used in standard practice. A voxel with a score of >0.5 is more than likely to be from the structure and will "activate" the model to produce an output. In our study there were no voxels on the 3D MRL scan with a probability of >0.5 for the right SMG, therefore the model failed to generate any contours. There is a suggestion the differences in anatomical contrast on the 3D MRL scans and those that the model$_{MRI}$ was trained on (t2_tse_tra) are enough to cause the model to underperform and fail to produce an output. Such differences may not be obvious to the human eye but may be related to the differences and inter-tissue contrast and textures in the images. In addition, the ability of the model to contour may have been improved by using a larger testing dataset. Our findings are supported by a study by van Rooij et al. whereby the right SMG was not contoured due to its density being lower than in the training set. The deep learning model in this study incorrectly contoured a neighbouring structure as the "right SMG" of similar density to what is was trained on (233). This highlights the sensitivity of deep learning convolutional neural networks to the input data they receive and the potential challenges when developing a MR based model for generic use.

Model$_{MRI}$ outperformed Model$_{CT}$ for the bilateral parotid glands and SMGs. Improved mean DTA values were observed for the right parotid gland and right SMG compared with the left sided structures. This could be due to the relatively small numbers used in this study, and may improve in larger cohort. The superior soft tissue visualisation and ability to detect the edges of a structure with MR compared with CT may improve the sensitivity of model$_{MRI}$ to produce contours closer to the ground truth (14,25). The accuracy of the 'ground truth' contours are not well defined and can be influenced by imaging modality. Chen et al. in a study of 26 patients highlighted the differences in GTV when contoured on CT or MRI Viewray images (a 0.35T MRI guided radiotherapy machine) with MRI-derived GTV contours being significantly larger (252). To exploit the benefits and minimise the shortfalls of both imaging modalities for OAR delineation, it would be of interest to explore combining CT and MR images into an auto contouring solution using co-registration software. For example, Brainlab auto contouring software used in CNS treatment planning co-registers cranial MR datasets to a reference CT or MR for better target delineation and improved quality assurance. Mocnik et al. developed a multimodality auto contouring tool by co-registration of CT and MR images and compared its performance against a CT only model for parotid gland delineation. The mean DSC was similar between the two groups (0.79 v 0.76) and improved slightly to 81% of the cases using the co-registered model (253). The better overlap distance using model$_{MRI}$ auto-contours applied to the T2W Dixon 2D radiotherapy planning scans may have a role in replacing CT based models for use in treatment planning but requires evaluation with a larger library of OARs and a more heterogeneous training patient cohort.

Adaptive re-planning with MRIgRT has the potential to improve dose to the target and minimise normal tissue toxicity by facilitating real time adaptation in response to tumour reduction or anatomical change during a course of radiotherapy (251). Real time adaptation of contours is time consuming and labour intensive. AI tools such as model$_{MRI}$ have the potential to reduce workload, inter observer variability and may provide dosimetric evaluation of OARs (233). The performance of model$_{MRI}$ on the 3D MRL needs further optimisation prior to its use in MRIgRT. Potential strategies include: (1) training

the model on similarly acquired data, e.g., using data from the 3D MRL and testing the model on a similar dataset; and (2) developing a novel 2D sequence which could be tested on the MRL. Using model$_{MRI}$ assisted contours on a novel 2D MRL sequence will help assess its suitability to contour OARs in adaptive re-planning. In future, MRI based auto contouring should also be tested on different sequence settings such as T1, which may perform better by improving the visualisation of OARs or lymph node levels (133).

This study is the first to develop a novel MR deep learning auto contouring tool for OAR delineation in head and neck radiotherapy planning but has some limitations. First, model$_{MRI}$ was trained and tested on data from a single institution which potentially may limit its use in other institutions and in multi-centre clinical trials. Second, the training set also excluded problematic images that contained artefacts or tumour infiltration. Analysis of the relative performance of each model for patients with more image artefacts is an area of future work. Third, the analysis of the model did not include qualitative scoring metrics but used quantitative assessments of DSC and DTA similar to those used in others studies (250,254). DSC and DTA are limited as only assess the geometric similarities between contours, which do not always correlate with clinical criteria. In addition, they do not account for the time taken to manually adjust auto contours. The accuracy of DSC is also affected by the volume of the structure. Using a combination of quantitative and qualitative metrics to determine the accuracy of auto contours, such as goodness of fit scores described by Greenman et al. or grades of volumetric revision as studied by Lin et al may improve the clinical application of the model by including subjective evaluation of contours (132,202).

Model$_{MRI}$, which was tested on a small number of data sets (n=28), requires further training and validation in a larger multi-centre cohort to increase its robustness and address the limitations when used on the MRL, prior to clinical use. Model$_{MRI}$ also needs to be expanded to cover other OARs within the head and neck. The potential benefit of model$_{MRI}$ to save contouring time as shown by model$_{CT}$ in Chapter 5 and in other studies is also an area of future work (250). There is also a need to compare a fully automated contouring model with one that is manually edited for clinical acceptability.

6.6 Conclusion

MR based deep learning auto-contouring models show promise as an aid to clinician OAR contouring but are sensitive to the MR sequence used. Further work is needed to optimise the MRL image acquisitions and evaluate the model in a larger cohort prior to evaluating its use in adaptive re-planning.

7.0 Inter-fraction Robustness of Intensity-Modulated Proton Therapy in the Post-operative Treatment of Oropharyngeal and Oral Cavity Squamous Cell Carcinomas

C. Hague[1], MC. Aznar[2,3], T. Li[4], S. O'Reilly[4], L. Dong[4], A. Fotouhi-Ghiam[4], LW. Lee[1], A. Lin[4], M. Lowe[1], JN. Lukens[4], A. McPartlin[1], N. Slevin[1], S. Swisher-McClure[4], D. Thomson[1], M. van Herk[2,3], C. West[5], W. Zou[4], BK. Teo[4]

[1]Department of Head and Neck Clinical Oncology, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, UK
[2] Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Heath, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
[3] Department of Radiotherapy Related Research, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, UK
[4] Department of Radiation Oncology, University of Pennsylvania, Philadelphia, USA[5] Translational Radiobiology Group, Division of Cancer Sciences, Manchester Academic Health Science Centre, University of Manchester, The Christie NHS Foundation Trust, Wilmslow Road, Manchester, UK

Authors' contribution:

The study presented in this chapter was conducted at the Roberts Proton Therapy Centre, University of Pennsylvania, USA. This chapter reports the results of a study evaluating the robustness of multi-field optimisation plans to uncertainties in range and set up in post-operative oropharyngeal and oral cavity cancers. I performed contouring, collected and analysed the data and drafted the manuscript. Dr Alex Lin selected the patients and provided information about patient and tumour characteristics. The medical physics department provided radiotherapy planning and CT fusion. Dr Kevin Teo and Dr Taoran Li performed deformable registration, propagation of contours and dose recalculation on weekly cone beam CT scans. I received input from Dr Kevin Teo, Dr Matthew Lowe and Dr Alex Lin when drafting the manuscript.

7.1 Abstract

Introduction. To evaluate dosimetric consequences of inter-fraction set up variation and anatomical changes in patients receiving multi-field optimised (MFO) intensity modulated proton therapy for post-operative oropharyngeal (OPC) and oral cavity (OCC) cancers.

Methods. Six patients receiving MFO for postoperative OPC and OCC were evaluated. Plans were robustly optimised to clinical target volumes (CTVs) using 3 mm setup and 3.5% range uncertainty. Weekly online cone-beam computed tomography (CBCT) were performed. Planning CT was deformed to the CBCT to create virtual CTs (vCTs) on which the planned dose was recalculated. vCT plan robustness was evaluated using a set up uncertainty of 1.5 mm and range uncertainty of 3.5%. Target coverage, $D_{95\%}$, and hotspots, $D_{0.03cc}$, were evaluated for each uncertainty along with the vCT-calculated nominal plan. Mean dose to organs at risk (OAR) for the vCT-calculated nominal plan and relative % change in weight from baseline were evaluated.

Results. Robustly optimized plans in post-operative OPC and OCC patients are robust against inter-fraction set up variations and range uncertainty. $D_{0.03cc}$ in the vCT-calculated nominal plans were clinically acceptable across all plans. Across all patients D95% in the vCT-calculated nominal treatment plan was at least 100% of the prescribed dose. No patients lost ≥10% weight from baseline. Mean dose to the OARs and max dose to the spinal cord remained within tolerance.

Conclusions. MFO plans in post-operative OPC and OCC patients are robust to inter-fraction uncertainties in set-up and range when evaluated over multiple CT scans without compromising OAR mean dose.

Advances in knowledge. This is the first paper to evaluate inter-fraction MFO plan robustness in post-operative head and neck treatment.

7.2 Introduction

Radiotherapy for head and neck cancers is challenging due to the proximity of normal structures. Irradiation of healthy tissue can lead to long-term toxicities such as xerostomia, dysphagia and dysgeusia that have a negative impact on quality-of-life (255). The finite range of high energy protons makes proton beam therapy an attractive treatment option for squamous cell cancers of the head and neck by limiting dose to normal tissues beyond the target volume. However, the sharp distal fall-off of proton beams makes the dose distribution sensitive to setup, motion and range uncertainties. In proton treatment planning, potential sources of error resulting from uncertainty in the location of the Bragg peak need to be considered. Potential sources of error include: daily changes in the patients' position or anatomy, organ motion, delineation variation, image artefacts, inaccurate conversion of CT Hounsfield units to the proton stopping power and changes in the beam path due to variable tissue densities (256,257). Patient immobilisation, image guidance, expansion margins and use of a dual energy CT may reduce errors but in clinical practice centre-specific setup uncertainties and ±3.5% uncertainty on stopping powers are typically used to account for residual uncertainties (156,258). Despite attempts to reduce setup uncertainties, some daily variations are unavoidable and may result in under- or over-dosage of the target volume and organs at risk (OARs) respectively.

Robustness to uncertainty may be attained in many ways. One such approach is through robust plan optimisation whereby uncertainties are accounted for in the optimisation process. The optimisation process can reduce dose gradients within the plan and consequently the occurrence of hot and cold spots due to uncertainties in the proton beam range. In robust optimisation, patient set up errors and the corresponding perturbations in the dose distribution are simulated by shifts in the isocentre position on a single CT scan. Range uncertainties are simulated by scaling the CT to stopping power ratio calibration. Setup variation, however, is more complex than a rigid translation of the CT image and can involve soft tissue deformation, variations in spine alignment, neck tilt and shoulder position. In the literature, different methods of robust plan optimisation exist, including probabilistic treatment planning, the use

of worst-case scenarios, selective robust and minimax optimisation (259–261). Probabilistic treatment planning described by Unkelbach incorporates both range and setup as random variables into the optimisation process. This contrasts with worst-case optimisation described by Liu at al whereby the optimisation is based on the worst case scenario (152,155). However, the robust optimisation process only accounts for positional and range uncertainties, and does not evaluate robustness against anatomical changes during the treatment course. Commercial systems using multiple patient images prior to treatment are being studied for anatomical robust optimisation but are yet to be adopted in routine clinical practice (262,263). Using pencil beam scanning, two methods for plan optimisation can be used: single field optimisation (SFO) and multi-field optimisation (MFO). In the latter, all beam spots are optimised together using multiple fields and beam angles to modulate and shape the dose. A superior dose distribution can be delivered compared with passively scattered proton beams or SFO, in particular in cases of complex geometry. MFO, however, can be prone to inter-fractional uncertainties due to anatomic changes, such as from weight loss or tumour response, resulting from increased in-field dose gradients. These non-linear changes in patient's anatomy due to tissue deformations are not explicitly modelled in robust optimisation; it is unknown if robustly optimised MFO can still accommodate these unplanned range variations.

Conventionally, uncertainties are accounted for with the use of a planning target volume (PTV). In a study of fourteen patients by Liu et al, a robust clinical target volume (CTV) based optimisation approach produced a superior plan for targets and OARs compared with a PTV approach (156). Stutzer et al showed MFO to be superior to SFO in original plan robustness in the treatment of oropharyngeal cancer (264).

MFO is potentially more sensitive to tissue deformation during the treatment course, which is likely to be different along different beam angles. The effectiveness of robust plan optimisation in MFO has not been studied in post-operative head and neck treatment planning. There is also no agreed consensus as to which optimisation method is superior in evaluating MFO plans partly due to variable literature and lack of standardised protocols. In this study,

we aim to evaluate how robust MFO plans are to uncertainties in set up and range error for patients treated post-operatively for oropharyngeal and oral cavity cancers using weekly cone beam computed tomography (CBCT) scans. Such CBCT scans capture realistic inter-fraction setup variations as well as tissue deformations during the entire treatment course.

7.3 Methods

7.3.1 Patient selection

Six head and neck patients treated between July 2017 and April 2018 and planned with multi-field robust optimisation in Eclipse (Eclipse v 13.7, Varian Medical systems) were retrospectively selected. Inclusion criteria were: patients with oropharyngeal or oral cavity cancers requiring post-operative proton beam therapy to the primary site and elective neck, and the availability of weekly CBCT images. All patients were immobilised in a five-point thermoplastic mask and positioned using daily orthogonal kV imaging. The relative percentage change in each patient's weight was recorded weekly during treatment which was subsequently correlated with CTV and OAR coverage.

7.3.2 Planning approach

Each patient was treated with two or three CTV dose levels. CTV1 was defined as the surgical bed with CTV2 and CTV3 defining "at-risk sites". CTV1 received 60-63 Gy(RBE), CTV2 received 54-60 Gy(RBE) and CTV3 received 54 Gy(RBE) all treated with a dose of between 1.8 to 2.1 Gy(RBE) per fraction. Each treatment plan was based on a three-field beam arrangement consisting of 2 posterior obliques and 1 anterior field with range shifters (7.2 cm anterior 8.0 cm posterior) in situ as shown in Figure 7.0 (a) and 7.0 (b). The posterior oblique fields cover the superior portion of the target while the anterior field covers the inferior target. The posterior oblique and anterior fields overlap in the superior-inferior direction over a 2 cm region and are optimized to be robust against 3 mm longitudinal isocentre perturbations between the posterior oblique and anterior fields. In this way, smooth dose gradients are achieved in the superior-inferior directions for each field at the overlap region without the need

to feather the match line. Range shifters are positioned close to the patient to minimise the post-range shifter airgap and subsequent increase in spot size.

The total prescribed dose to the original treatment nominal plan was defined such that 95% of each target volume received a minimum of 100% of the prescribed dose to this CTV level. Clinically acceptable plans require 95% of the target volume in the worst-case scenario to receive at least 95% of the prescribed dose and D0.03cc to receive ≤110% of the prescribed dose.

Figure 7.0. Three field IMPT beam arrangement (a) and (b) with two posterior oblique fields and an anterior field.
The fields are indicated by red arrows. In situ range shifters located anterior and posterior to the patient are used to minimize the air gap between the range shifter and the patient. Representative lateral (c) and anterior (d) DRRs of one planning CT and maximum intensity projection of 5 vCTs (e) and (f) with red contours representing inter-fraction setup variation of cervical vertebrae.

7.3.3 Analysis of Robustness

Plans were optimised using MFO. CTVs were robustly optimised using a setup uncertainty of 3 mm and a range uncertainty of 3.5%. Set up perturbations were simulated along three orthogonal directions and combined with range uncertainties in each position to produce twelve scenarios used in the robust optimization, (x±3mm), (y±3mm), (z±3mm) and CT Hounsfield Unit (HU) scaling of ±3.5%.

Daily positioning and setup correction were accomplished using orthogonal kV imaging. Each patient underwent weekly online CBCT, which were then used to create reliable virtual CTs (vCTs). If a CBCT was acquired, a 3D-3D match was performed followed by a confirmatory kV imaging without additional couch motion. The vCTs were generated by deforming the planning CT onto the CBCT using a diffeomorphic implementation of the Morphons algorithm (265). The deformable registration provided a high-quality image in the treatment geometry on which the planned dose could be evaluated to estimate the delivered dose. Each target CTV on the vCT was compared with the original treatment nominal plan and manually edited to modify the superior and inferior extension to ensure consistency of anatomical landmarks. The vCT plan robustness evaluation was performed using a residual set up uncertainty of 1.5 mm and range uncertainty of 3.5%. The 1.5 mm setup robustness evaluation considers the uncertainty in the coincidence between the imaging and radiation isocentres, intrafraction motion, as well as variations in user dependent choice of region of interest for evaluating image registration between the vCT and the planning CT. Actual setup variability of the cervical neck vertebrae is illustrated in the digitally reconstructed radiographs (DRRs) depicted in Figures 7.0 (e) and 7.0 (f) for patient 5. A video depicting the setup variation is available in the supplementary data.

7.3.4 Dose calculation

All doses and dose volume histograms (DVHs) were calculated using Proton Convolution Superposition (PCS) v13.7 within the Eclipse™ treatment planning system (Varian Medical Systems, Palo Alto, CA). Plan metrics were extracted

from the calculated DVHs using a MATLAB (Mathworks, Natick, MA) script. Maximum and minimum values under uncertainty for $D_{95\%}$ for each CTV dose level and $D_{0.03cc}$ for the high dose CTV were extracted and compared with the original treatment nominal value. Mean dose in the vCT-calculated nominal case to ipsilateral and contralateral parotid glands, oral cavity, pharyngeal constrictor muscles, larynx and maximum dose to the spinal cord were calculated. Relative % change in weight from baseline to end of treatment was recorded.

7.4 Results

The demographics of the six patients are outlined in Table 7.0. In all patients the vCT-calculated nominal treatment plan $D_{95\%}$ across the vCTs was at least 100% of the prescribed dose. The median time from baseline radiotherapy planning scan to first CBCT within week 1 of treatment was 29 (range 22-52) days.

Across the six patients, robustness evaluation of CTV coverage using inter-fraction uncertainties of 3 mm set up and 3.5% range error for the initial plan (day 0) and 1.5 mm for the residual setup and 3.5% range for the vCT-calculated plans are shown in Figures 7.1a and 7.1b. The dose to the vCT-calculated nominal plan is close to the maximum CTV dose band across all six plans. Clinically acceptable $D_{0.03cc}$ values in the high dose CTV levels are shown in Figures 7.1a and 7.1b with the exception of patient 6. In patient 6, the highest $D_{0.03cc}$ value for the vCT-calculated nominal plan is 113% (71 Gy v 63 Gy) of the target dose. The hot spot in patient 6 is shown in Figure 7 and can be explained by a shift in the jaw position and variation of the posterior neck. When uncertainty analysis is included, the highest prescribed $D_{0.03cc}$ for this fraction was 116% of the target dose (73 Gy v 63 Gy if scaled to the full treatment dose). Analysis of daily kV setup image revealed that this relatively large variation in jaw position occurred only once during the entire course of treatment. The $D_{0.03cc}$ values for the other two vCT-calculated nominal plans of patient 6 in Figure 7.1b were within the uncertainty bands of the initial plan.

None of the six patients lost >10% weight relative to their baseline throughout treatment. The largest absolute (relative %) weight loss were noted in patients 4 and 6 both of whom lost 6% by the end of treatment. Two patients lost no weight and in the remaining two patients their relative % weight loss remained within 5%.

Table 7.0. Patients' demographics.

Patient number	1	2	3	4	5	6
Age (years)	73	50	67	70	55	74
Gender	M	M	M	F	M	M
Site	Base of tongue	Tonsil	Oral Tongue	Tonsil	Vallecula	Oral Tongue
Laterality of tumour	Right	Right	Right	Right	Left	Left
Stage	T2N2bM0	T2N2aM0	T2N1M0	T2N1M0	T4aN1M0	T2N0M0
HPV status	Positive	Positive	Negative	Positive	Positive	Negative
Chemotherapy	None	Weekly Cisplatin	None	Weekly Cisplatin	Weekly Cisplatin	Weekly Cisplatin
Prescription dose/ Gy RBE	60	63	60	63	63	63
Baseline weight (kg)	78	92	79	78	99	109
Weight loss in kg at the end of treatment (relative % change)	0	3 (3%)	0	5 (6%)	1 (1%)	7 (6%)

Abbreviations: M= Male, F=Female; RBE= Relative Biological Effectiveness; HPV=human papilloma virus
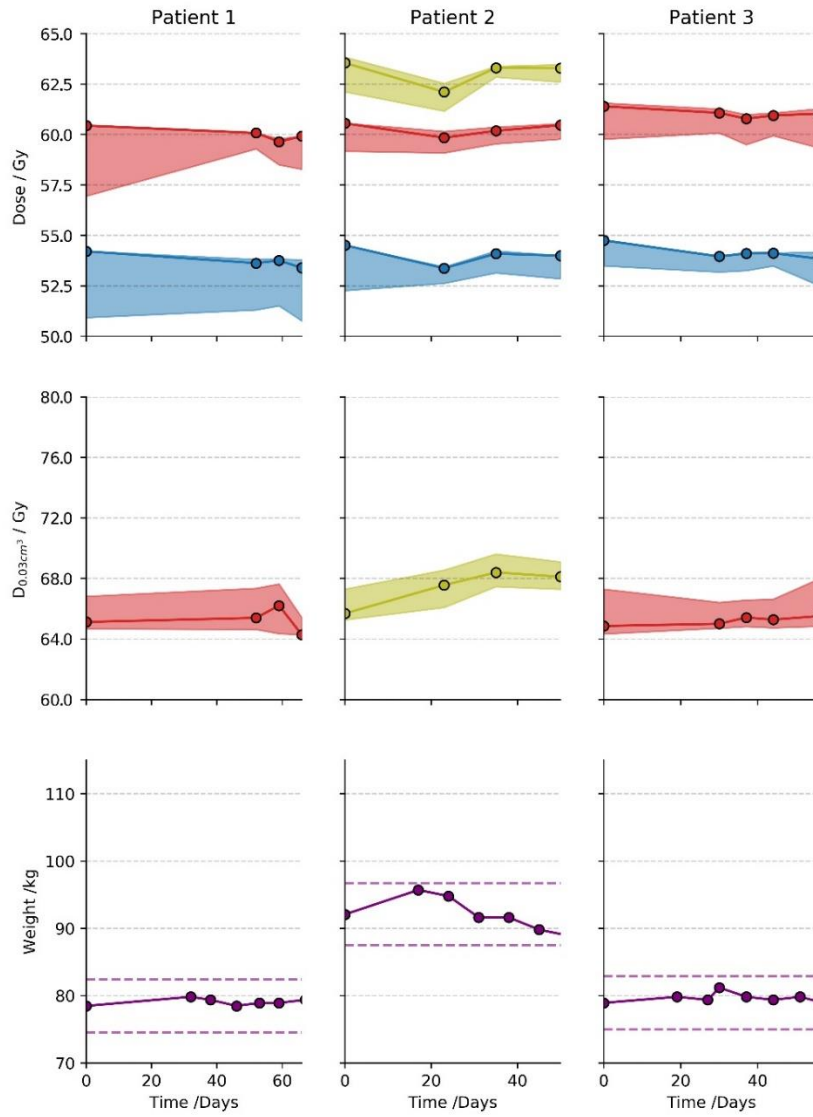
Figure 7.1a.
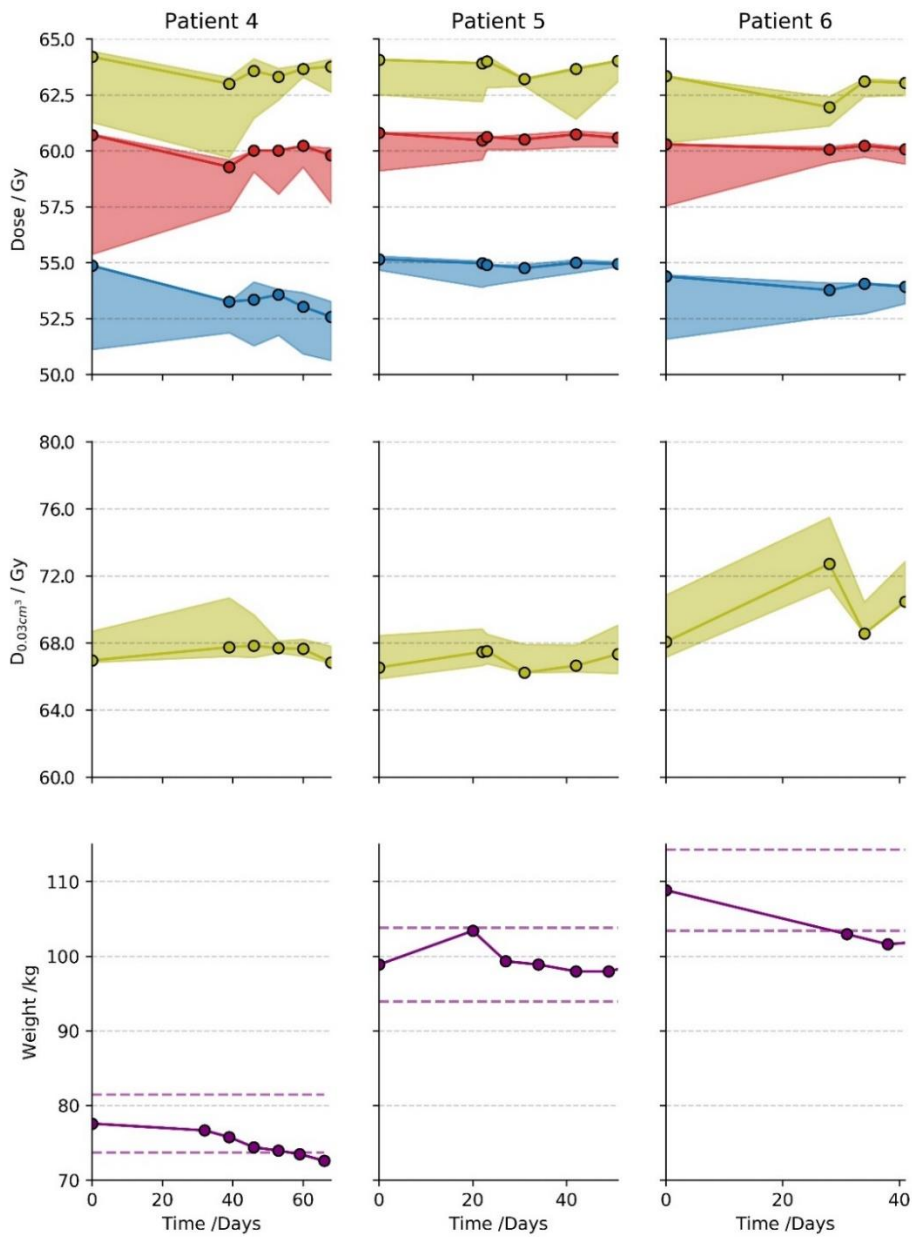
Figure 7.1b.

Figures 7.1a and 7.1b illustrate the changes in D95%, D0.03cc in the high-risk CTV and patients' weight from baseline to the end of treatment.
Uncertainties over all error scenarios are shown as a light band, CTV 63 (green), CTV 60 (red), CTV 54 (blue). The solid line shows the nominal dose.
Dashed lines indicate weight changes of ±5% from the baseline weight recorded at the time of the planning CT scan.
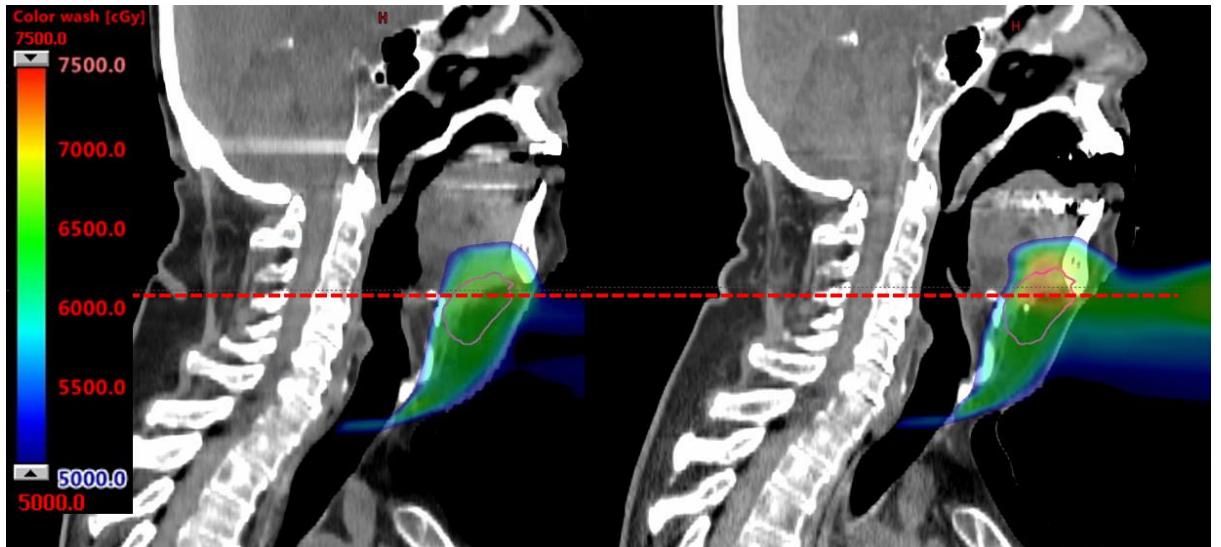
Figure 7.2. Sagittal view of the planning CT for patient 6.
CTV1 contour and corresponding slice of the first vCT-calculated nominal plan are shown.

The hotspot is attributed to differences in the jaw position (shown by the dashed red reference line) and the setup variation of the posterior neck region.
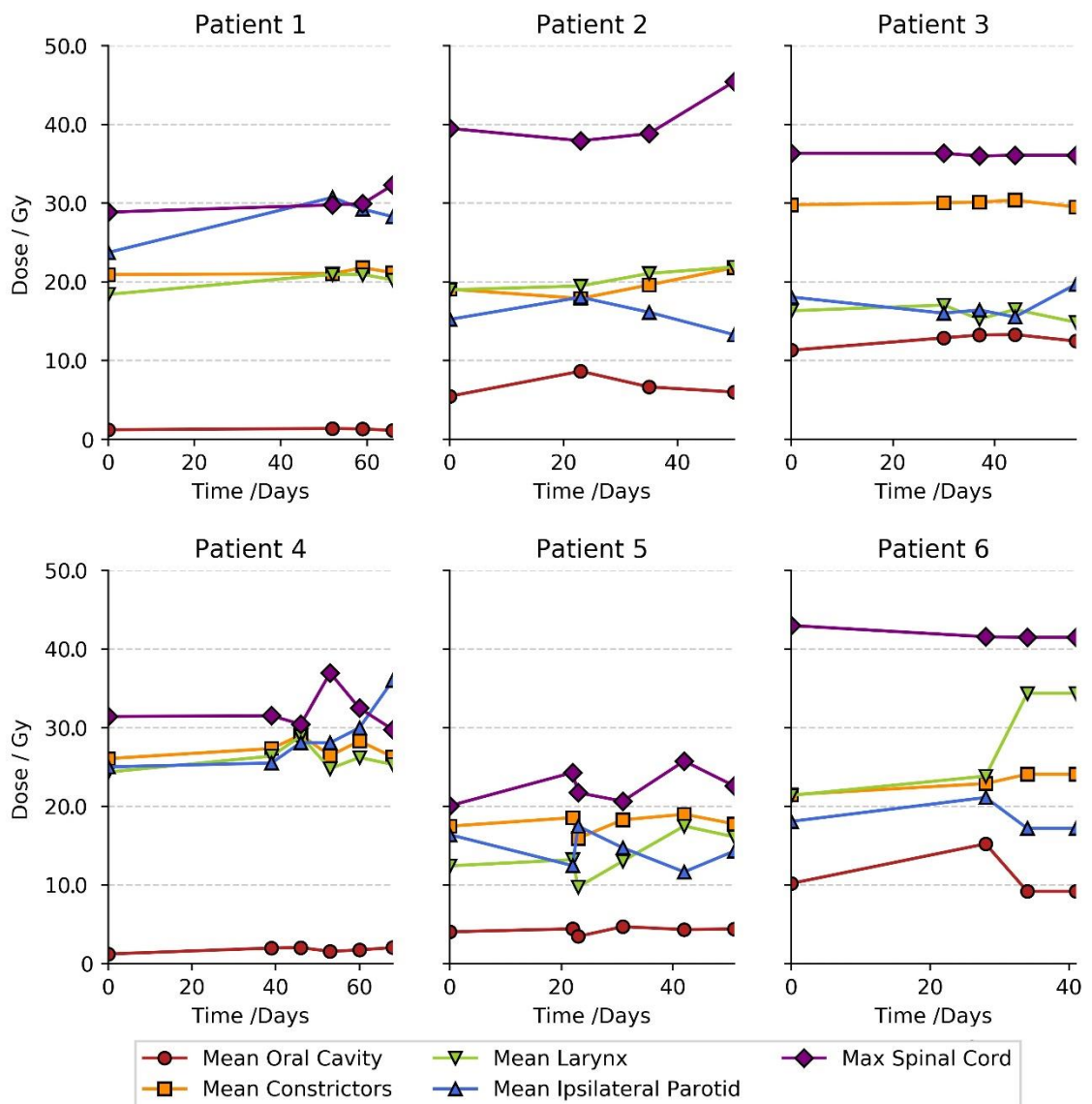
Figure 7.3. Changes in mean dose to the OARs and maximum dose to the spinal cord with time.
OARs: oral cavity, pharyngeal constrictor muscles, larynx, ipsilateral parotid gland. Dose constraints to each OAR: oral cavity mean dose 20 Gy or ALARA, pharyngeal constrictor muscles mean dose 50 Gy, larynx mean dose 20 Gy or ALARA, ipsilateral parotid gland mean dose 20 Gy or ALARA, maximum spinal cord 45 Gy.

Abbreviations: ALARA as low as reasonably achievable.

The impact of inter-fraction setup, anatomical variation and changes in patient weight on the OAR doses is depicted in Figure 7.4. While vCT-calculated

nominal doses do vary for each patient, dose constraints did not exceed planning objectives for any OARs.
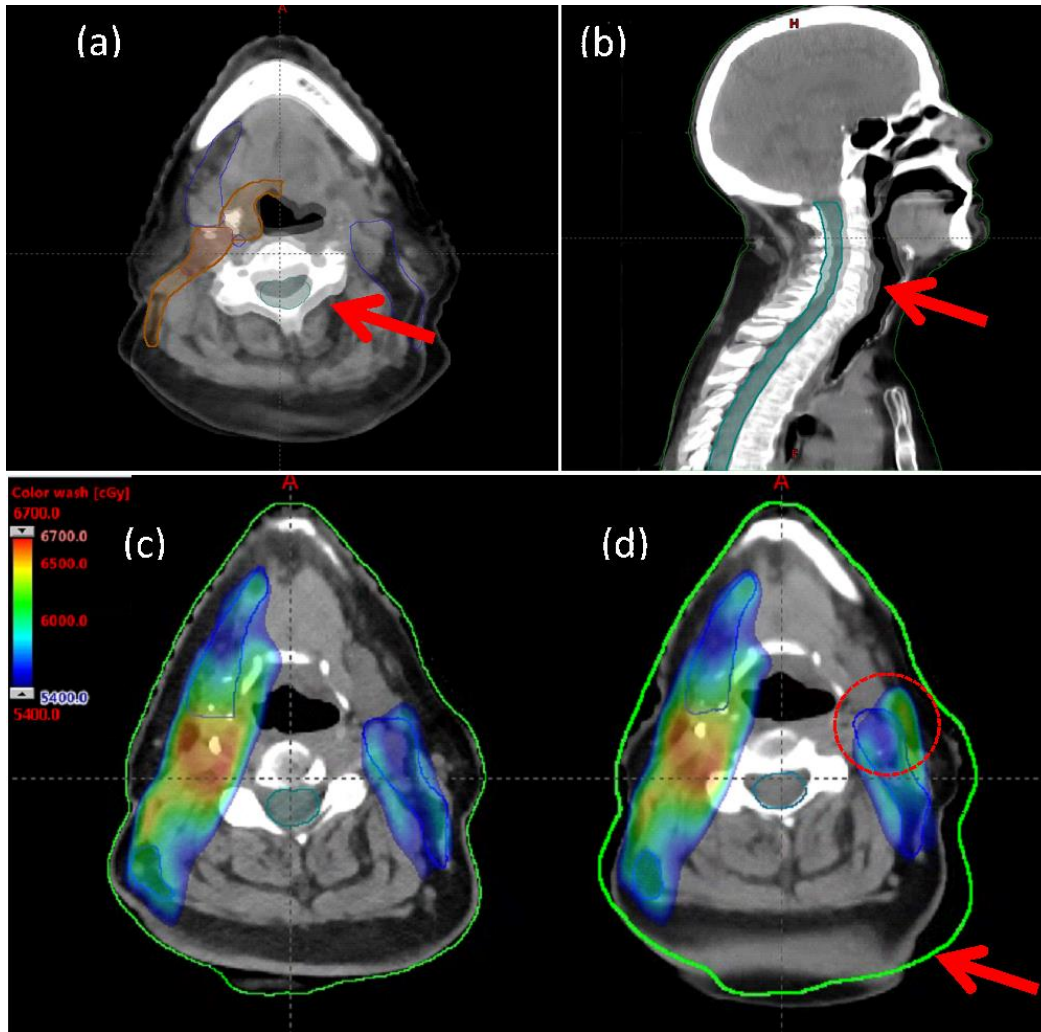
Figure 7.4. Setup variation in the neck region between the planned and actual treatment positions for patient 4.
Axial (a) and sagittal (b) views of setup variation in the neck region between planned and actual treatment positions for patient 4.

Blended images of planning CT and vCT shown demonstrating differences in the neck angle as indicated by red arrows. Nominal dose distribution on the planning CT is shown in (c) with CTV 54 contour in blue and the corresponding vCT dose. Green body contour of the planning CT is shown in (c) and (d). Red dotted circle in (d) indicates region of hotspot outside of CTV 54 and cold spot within CTV 54. Red arrow in (d) indicates region with variation in setup position.

7.5 Discussion

This novel pilot study to evaluate dosimetric consequences of anatomical variations and range errors in MFO plans in the post-operative setting has demonstrated:

i.   plans are robust to uncertainties in set up and range,

ii.   weight loss within 6% does not negatively impact dosimetric coverage,

iii.   OAR exposure for all vCT-calculated nominal plans are within tolerance.

MFO plans have been shown to be robust to uncertainties early in the treatment course enabling plan adaptation if necessary. Weight loss of >10% from baseline in head and neck radiotherapy is influenced by radiation related toxicities and significantly associated with global poor health-related quality of life (266). Changes in weight can affect the positioning of OARs such as the parotid gland resulting in potential over-dosage. The use of proactive enteral feeding is the cause of much debate. In this study none of the six patients lost ≥10% weight throughout treatment and required enteral feeding. Of the two patients who lost the largest amount of weight (6%) from baseline, plans remained robust to uncertainties. This highlights the potential importance of aggressive and proactive symptom and supportive management during treatment, so as to minimize weight loss.

Proton planning, particularly in head and neck radiotherapy is challenging due to uncertainties in setup and range on a background of complex anatomy and close proximity of tumours with normal tissues. Optimisation techniques may help minimise the effect of uncertainties, accepting the trade-off between delivering an acceptable integral dose and normal tissue sparing (267). In this study, dose constraints to the oral cavity, pharyngeal constrictor muscles, larynx and parotid glands were achieved without compromising plan robustness.

MFO proton beam plans in head and neck radiotherapy are susceptible to the development of hot and cold spots due to anatomical motion and changes in the

position of the head and neck. During radiotherapy, as patients relax, the neck and jaw position can move, which potentially alters the dose distribution. In all six patients the distribution of hot spots in the vCT-calculated nominal plan as defined by $D_{0.03cc}$ were clinically acceptable (within 110% of the prescribed dose and 115% for the worst case). In patient 6, $D_{0.03cc}$ of 113% in the vCT-calculated nominal plan was noted at day 34 from baseline. The development of a hot spot correlated with a shift in the jaw position as shown in Figure 7.2. The dose distribution corrected to within the clinically acceptable range on subsequent weekly CBCTs. Changes in the patients' position is a random variable unlike weight loss which is more systematic throughout treatment. More regular imaging with daily CBCTs could be considered in those who experience hot spots, $D_{0.03cc}$ >110% and weight loss >10% to ensure an adequate dose distribution to the target and maintain OAR constraints. Another example of the impact of setup variation of the neck and its impact on dose distribution of MFO plans is depicted in Figure 7.4 where the hotspot is seen outside of the CTV. Current robust optimisation techniques do not take into account anatomical deformation.

Although this is the first study to evaluate optimisation of MFO plans in post-operative oropharyngeal and oral cavity patients, the sample size is small and not all patients underwent the same number of CBCT scans during treatment. In addition, the mean doses at each vCT-calculated nominal plan reflect dose to the full prescription but the measurement point is for a single fraction. A more thorough analysis could be performed with daily CBCT using deformable dose accumulation on the vCTs. Unfortunately, daily CBCT was not available for this study. Challenges of maintaining robustness may be different in other anatomic sites in the head and neck, such as in the paranasal sinuses and skull base, where random, daily changes in sinus filling may occur. Despite these limitations this is the first paper to our knowledge to evaluate MFO robust optimisation in postoperative oropharyngeal and oral cavity cancer patients and may help in the process of developing optimisation protocols with agreed parameters to help identify plans that require additional individualisation.

7.6 Conclusion

In this novel pilot study, MFO plans in post-operative oropharyngeal and oral cavity patients were observed to be robust to uncertainties in set up and range in regard to CTV coverage when evaluated over multiple CBCT scans. Dose constraints to OARs remained within tolerance and changes in weight from baseline did not appear to affect CTV coverage or plan quality. Development of a robust analysis protocol for MFO plans may improve consistency of reporting and plan evaluation amongst radiotherapy centres.

7.7 Acknowledgements

8.0 Patient involvement in the design of a phase III trial comparing IMPT and IMRT for oropharyngeal cancer

C. Hague[1], B. Foran[2], S. Guild[3], E. Hall[4], O. Joseph[5], L. Lee[1], A. McPartlin[1], R. Moule[6], C. Nutting[7], M. Ofuya[4], S. Parsons[5], R. Prestwich[8], N. Slevin[1], M. VanHerk[9], C. West [9], D. Thomson[1]

[1]The Christie NHS Foundation Trust, Clinical Oncology, Manchester, United Kingdom.
[2]Weston Park Hospital, Clinical Oncology, Sheffield, United Kingdom.
[3]Patient representative, Leeds, United Kingdom
[4] The Institute of Cancer Research, Clinical Trials and Statistics Unit, London, United Kingdom
[5]The University of Manchester, Public Programmes Team, Manchester, United Kingdom
[6]The University College London Hospital, Clinical Oncology, London, United Kingdom
[7]The Royal Marsden NHS Foundation Trust, Clinical Oncology, London, United Kingdom.
[8]St James' University Hospital, Clinical Oncology, Leeds, United Kingdom.
[9]The University of Manchester, Division of Cancer Science, Manchester, United Kingdom

Authors' contribution:

The editorial presented in this chapter reports the results of a patient and public involvement study early in the trial design of TORPEdO, the first UK proton beam therapy trial in oropharyngeal cancer. Patients' views on the study design, patient pathway and willingness to travel and stay away from home were sought to better shape and inform the trial design. I selected patients, organised and led three focus groups with the help of Mrs Suzanne Parsons, Miss Olivia Joseph and Dr David Thomson. Dr Robin Prestwich and Dr Bernadette Foran helped with the focus groups at Leeds and Sheffield, respectively. Suzanne Parsons and Olivia Joseph helped to collect the data. I analysed the data and wrote the editorial with input from the co-authors.

8.1 Editorial

For patients with favourable risk human papilloma-virus associated oropharyngeal cancer, local control and survival outcomes are excellent (11). However, despite the use of highly conformal intensity-modulated radiotherapy (IMRT) severe acute and late side effects are common, adversely impacting quality-of-life. Compared with photons, the superior dosimetric properties of protons with sharp lateral penumbra and distal fall-off reduce the radiation dose beyond the target volume and may lessen treatment-related toxicities such as: oral mucositis, dryness, taste disturbance, swallowing dysfunction and osteoradionecrosis. However, there are only preliminary observational data to support the clinical advantage of proton beam therapy for oropharyngeal cancer (134,135,268,269), and prospective randomised trials are needed. An on-going phase II/III study (NCT01893307) from the MD Anderson primarily aims to compare rates of late grade 3-5 toxicities between IMRT and intensity-modulated proton therapy (IMPT) for oropharyngeal cancer (https://clinicaltrials.gov/ct2/show/NCT01893307, accessed December 2017). The UK proposes to open a multi-centre phase III study (TORPEdO, TOxicity Reduction using Proton bEam therapy for Oropharyngeal cancer) to assess the benefit of IMPT in terms of patient reported toxicities and quality-of-life and, as a secondary objective, cost-effectiveness.

Approximately 700 of the 1500 funded annual capacity for two planned NHS UK proton beam centres (The Christie Hospital in Manchester opening August 2018, and University College London Hospital opening 2020) will be available for either clinical trials or evaluative commissioning. The UK, with an established strong track record in delivering major practice changing clinical trials in radiotherapy e.g., PARSPORT, START and CHHiP (28,270,271) aims to be at the forefront in establishing the evidence-base for the use of proton beam therapy.

Patient and public involvement in the early stages of trial design increases the success of a study in terms of its feasibility and acceptability to patients, thereby supporting recruitment (272–274). We conducted three focus groups in Manchester, Leeds and Sheffield to understand patients' views about the

proposed TORPEdO trial, including acceptability of randomisation, the patient pathway when enrolled in the trial, willingness to travel and stay in Manchester or London for proton beam therapy and the trial design and endpoints. Fifteen patients with favourable risk oropharyngeal cancer who had completed radiotherapy ≥1 year ago were identified and invited to participate from each centre. Overall, 33 out of the 45 invited patients and eight relatives attended the focus groups between September and October 2017. Each session lasted two hours and consisted of presentations and discussions structured around a series of questions (Table 8.0). Information was recorded on pre-prepared laminates, patient questionnaires, audio recordings and by telephone or email contact with a sample of patients following each meeting. Data were interpreted using thematic analysis.

Table 8.0. Questions asked during the focus group

All Centres	Additional questions for Leeds and Sheffield
What do you think are the differences between standard radiotherapy and proton therapy?	What are your views on travelling and staying in Manchester if you are offered proton beam therapy as part of the trial?
How do you feel about entering a study where there is a 50% chance of getting standard radiotherapy and a 50% chance of getting proton therapy?	
What are your views on the trial pathway?	
What are your views on the trial outcomes?	

Opinions were sought on the name of the proposed study in Manchester. 'TORPEdO' was thought to be concise, easy to remember and would not deter trial participation. Existing knowledge about proton beam therapy was variable. Some described protons as a more targeted therapy with less toxicity, whilst others knew very little but had heard the term protons described in the media. In general, people considered protons to be a superior treatment and had some understanding of the differences compared with standard photon radiotherapy. There was enthusiasm to participate and be randomised in the study, to both

help future patients (inform future treatments) and have a 50% chance of receiving IMPT (the patients were informed that proton beam therapy would not be available as an NHS treatment outside the clinical trial). Some expressed that they would be disappointed if randomised to IMRT, but this would not deter them from considering the trial. Reassurance was provided that it remains uncertain in this situation whether any potential dosimetric superiority achieved with proton planning translates into clinical benefit for patients i.e., the presence of clinical equipoise. The patient pathway and trial layout were viewed positively. In Sheffield and Leeds, the timings at each stage of the pathway were discussed e.g., the timings from study enrolment to randomisation. To allow sufficient preparation for those receiving IMPT in Manchester or London, people felt the time interval between each stage of the patient pathway should be minimised (e.g., time from clinic to outcome of randomisation; time from randomisation to travel to Manchester or London). Patients reported that they would need information about the proton centres and provision of support when considering taking part in the trial.

We sought to understand peoples' views on the logistical challenges in relation to travel and accommodation if randomised to IMPT. This was particularly pertinent for those living in Sheffield and Leeds. There was a general willingness to travel to Manchester for treatment planning and delivery, which was balanced against missing family and established social networks. Feelings were mixed in relation to staying in Manchester for the duration of treatment with IMPT. The preferred option was to stay in Manchester in apartment-based accommodation close to the hospital during the week with the option of returning home at the weekends. This preference was thought to be achievable with adequate family, clinical and nursing support. Some raised the possibility and feasibility of daily travel if continuing to work or due to childcare responsibilities. Many patients felt it important for a relative or carer to stay during treatment to provide both emotional and social support. Participants felt that, as far as possible, the accommodation should be tailored to individual's needs e.g., providing reliable / fast internet access to be able to contact family or equipment / toys for children. Easy access to shuttle transport between the apartment and the hospital was highlighted to avoid potential unnecessary

anxiety caused by inadequate hospital parking. All patients and relatives emphasised the importance of ensuring adequate support for those randomised to IMPT which included: telephone or face-to-face contact with a clinical nurse specialist, speech and language therapist and dietician.

Patients' views of toxicity as the primary outcome measure were discussed. In general participants felt this to be an appropriate and useful measure. We sought to understand the side effects experienced by patients six weeks and one-year post treatment. Following this, we explored views on the patient reported outcome questionnaires that we plan to use in the study. We asked opinions about the University of Washington questionnaire as a tool to collect outcome data at different time points (see appendix 1.0). The questionnaire consists of 15 single domains divided into six physical, six socio-emotional and three global questions, which have been validated for use in head and neck cancer trials (203). To understand if the six physical questions that make up the composite score for the primary outcome measure of the trial were relevant, we compared the side effects that patients reported as the most important to them with those measured by the University of Washington questionnaire. Of those reported four of the six physical symptoms (loss of taste, oral dryness, swallowing dysfunction and problems chewing) on the University of Washington scale, were described as the most important side effects one year after treatment. The acceptability of the questionnaire was assessed using a written feedback form, which was completed by 30/33 (91%) patients. Twenty-eight (93%) of the patients thought the six physical questions were a good primary outcome measure for the trial. All patients stated the questionnaire and the scale were clear. Twenty-eight patients (93%) thought the areas covered were relevant. Hearing loss was the most common missing symptom, highlighted by 5/30 (17%) patients. Other reported missing symptoms (for the primary outcome composite score) were: psychological issues (n=2) fatigue (n=1) and bone damage (n=1). Nine patients (30%) thought free text boxes were needed. All considered the frequency of completion of questionnaires over five years follow up to be appropriate and feasible. The favoured method of completion of the questionnaire varied: on paper in clinic (12/37; 32%), on paper at home (11/37; 30%), on a tablet in clinic (8/37; 22%) and online (6/37; 16%).

This was an important piece of work to understand patients' and carers perceptions about the first proposed proton trial in the UK. The feedback on the patient pathway and logistics was encouraging for the feasibility of the study and is invaluable in shaping the trial design.

Conflict of Interest

There are no conflicts of interest.

8.2 Acknowledgements

9.0 Development of a protocol to assess the feasibility of incorporating oxygen enhanced MRI measurements in radiotherapy planning for head and neck cancer

C. Hague, M. Dubec, K. Garcez, L.W. Lee, A. McPartlin, D. Thomson, N. Slevin, A. Sykes, C. West, J. O'Connor.

Authors' contribution:

This chapter describes the development of a protocol to evaluate the feasibility of OE-MRI in head and neck radiotherapy planning. I developed and wrote the study protocol. With the help of Miss Sally Falk, I competed all relevant research and ethics documentation including the online IRAS and forms submitted to REC. I presented the study with Prof O'Connor at the local REC and made the necessary amendments following this. Prior to completing the study protocol, I organised a cancer patient panel to gain important insight from previous patients on the study design and patient pathway.

9.1 Abstract

Introduction. Tumour hypoxia is associated with treatment resistance and poor survival outcomes in head and neck squamous cell cancer (HNSCC). Non-invasive imaging techniques may be used to identify the extent of hypoxia to facilitate re-planning for adaptive radiotherapy to improve survival while minimising toxicity. The aim was to develop a protocol to assess the feasibility of incorporating oxygen-enhanced magnetic resonance (OE-MRI) biomarkers in radiotherapy planning.

Methods. A prospective study in 30 patients will evaluate the feasibility of OE-MRI in HNSCC. The primary objective is to evaluate if OE-MRI is feasible in head and neck cancer to map areas of hypoxia and normoxia during radiotherapy with a view to plan adaptation. Secondary objectives include assessment of patient tolerability of the technique by questionnaires; assessment of OE-MRI repeatability in head and neck cancer and to evaluate relationships between OE-MRI and histological measures of hypoxia in tumour biopsies.

Ethics and dissemination. North West Greater Manchester granted ethical approval (REC: 244310). The study is registered on ClinicalTrials.gov (trial registration number NCT03646747).

9.2 Introduction

The incidence of head and neck cancer is increasing with a 30% rise since the early 1990s (2). Treatment is challenging due to anatomical location and associated co-morbidities. A proportion of these cancers will have hypoxic regions. Tumour hypoxia is associated with treatment resistance (radiotherapy and chemotherapy), poor loco-regional control and survival outcomes in head and neck squamous cell cancer (HNSCC) (164,275,276). Methods have been sought to improve treatment of hypoxic tumours such as by using radiosensitisers, e.g., nimorazole (39). Dose escalation to hypoxic regions in HNSCC also appears to improve tumour cell kill but with a potential risk of increasing treatment related toxicity (277–279).

A further clinical consideration is the increasing prevalence of HPV related HNSCC, which has a better prognosis than non-viral related disease. For this subset of patients' methods to maintain tumour control while minimising treatment morbidity are of particular interest. Knowledge of the presence and extent of hypoxia in HNSCC may facilitate radiotherapy dose painting and adaptive re-planning to reduce risks of toxicity without compromising survival outcomes. Non-invasive hypoxia imaging techniques are an attractive method to facilitate this.

Oxygen enhanced magnetic resonance imaging (OE-MRI) is a technique that quantifies tissue oxygenation by utilising the paramagnetic properties of oxygen as a contrast medium. It involves the acquisition of a dynamic set of T1 weighted images whilst the patient breathes room air (21% oxygen) for short periods of time followed by a period of breathing 100% oxygen, known as an oxygen challenge. Oxygen inhalation causes a change in the concentration of molecular oxygen dissolved in the blood or tissue, which can be detected by changes in T1 relaxation termed $\Delta R1$ (188,191). By measuring the $\Delta R1$ in plasma and interstitial fluid due to an oxygen challenge, areas of hypoxia can be mapped and quantified. By combining dynamic contrast enhanced (DCE)-MRI and OE-MRI tumours can be subcategorised into regions of normoxia, i.e. perfused and oxygen enhancing (Oxy-E), regions of hypoxia, i.e. perfused and non-oxygen enhancing (Oxy-R), and non-perfused regions. The fraction of

tumour tissue refractory to an oxygen challenge defined as Oxy-R appears to act as an imaging biomarker of tumour hypoxia as evidenced by O'Connor et al (192,280). In pre-clinical studies of two xenograft models, Oxy-R was able to distinguish differences in the hypoxic fraction as shown by pimonidazole staining and provide an estimate of the hypoxic fraction. The combination of DCE-MRI and OE-MRI has recently been identified in lung cancer to identify and map hypoxia but has not been investigated in HNSCC or radiotherapy planning (193,194).

OE-MRI is an appealing addition to DCE-MRI and T2 weighted blood oxygen level dependent (BOLD) sequences, which have been studied in this setting but are limited by spatial resolution and artefact susceptibility. These limitations may be mitigated by using T1 weighted sequences. OE-MRI as an imaging technique in head and neck radiotherapy planning requires technical and biological validation in a defined patient cohort. The aim of work in this chapter was to develop a protocol to assess the feasibility of incorporating OE-MRI into radiotherapy planning.

9.3 Study aims

9.3.1 Study objectives

Primary objective

To evaluate if OE-MRI is feasible in head and neck cancer to map areas of hypoxia and normoxia during radiotherapy with a view to plan adaptation.

Secondary objectives

To assess if OE-MRI is tolerated by patients

To evaluate the repeatability of OE-MRI in head and neck cancer

To establish relationships between OE-MRI and other measures of hypoxia in head and neck cancer

9.3.2 Study endpoints

Primary endpoint

To determine if an oxygen signal can be detected in head and neck cancer by measuring the perfused Oxy-R biomarker

Secondary endpoint

To determine tolerability of OE-MRI

To assess repeatability of OE-MRI by measuring ΔOxy-R

To evaluate relationships between OE-MRI and histological measures of hypoxia using patients' diagnostic biopsies

9.4 Study design

This prospective feasibility study is the first to explore the role of OE-MRI in head and neck cancer. Pre-clinical published data in xenografts has shown the ability of the biomarker 'perfused Oxy-R' to identify, map and quantify changes in tumour hypoxia which translated into the clinic. In these studies, three different types of hypoxic tissue classification were observed: normoxia, perfused oxy-R (hypoxia) and non-perfused. The data also demonstrated the feasibility and repeatability of OE-MRI in non-small cell lung cancer (188,192,194).

As this is the first study to evaluate OE-MRI in head and neck cancer initial development work was carried out to determine the best method of delivering oxygen for patients with HNSCC to produce a suitable imaging protocol, as an expansion on published results by Salem et al. (194). This preliminary work highlighted the difficulties with obtaining an adequate signal due to the low signal to noise ratio (SNR) when using a nasal cannula compared with a Hudson tight sealed facial mask. Sequence optimisation was continued using the facial mask but due to receiver coil issues on the 1.5T Philips Achieva system based at the Wolfson Molecular Imaging Centre (WMIC), there was insufficient SNR making it difficult to quantify confidently oxygen induced signal

change. The study will now be carried out at The Christie NHS Foundation Trust using the 1.5T Philips Ingenia system.

9.5 Study outline

This study will be carried out in 30 patients with head and neck cancer. Potential participants will be identified by Dr Hague, Prof O'Connor and healthcare team physicians at the oncology clinics in the Christie NHS Foundation Trust (Dr Andrew Sykes, Dr David Thomson, Dr Kate Garcez, Dr Lip Lee, Dr Andrew McPartlin and their teams). Participants will be provided with verbal information and given an information sheet. Those who agree to participate in the study will be asked to sign the study consent form A.
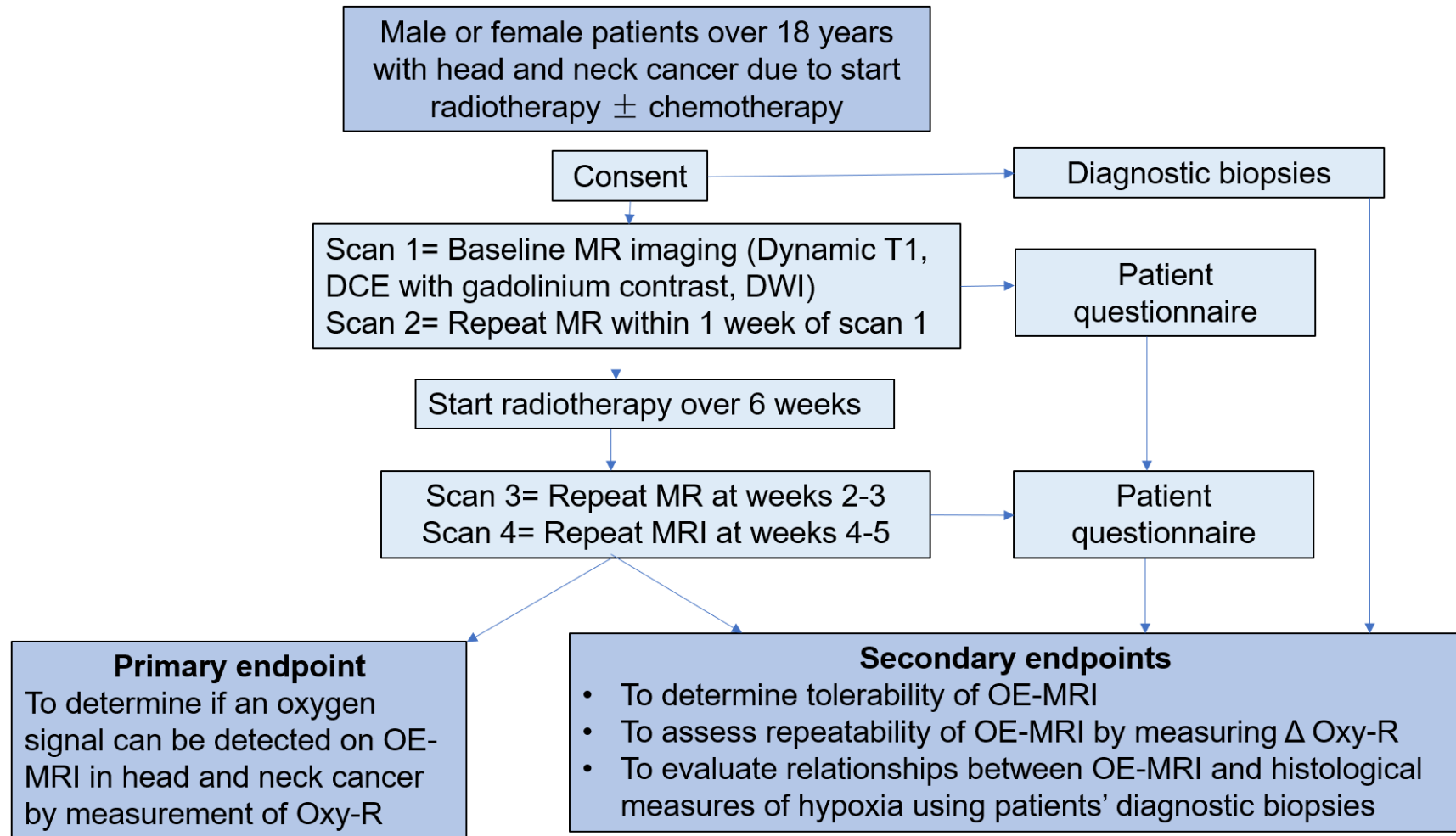
Figure 9.0. Study schema

9.5.1 Inclusion and exclusion criteria

Table 9.0. Inclusion and exclusion criteria for the study.

Inclusion Criteria
Age ≥18 years
Receive and understand verbal/written consent
ECOG performance status 0-2
Calculated creatinine clearance ≥30mL/min
Ability to lie comfortably on their back for up to 1 hour
Ability to tolerate a thermoplastic shell
Exclusion Criteria
Presence of contraindications to MRI scanning (heart valve replacement, pacemaker, metal implants not approved for MRI, neurostimulators)
History of allergy to gadolinium allergy
Female participants/patients who are pregnant, breastfeeding, surgically sterile or postmenopausal for ≥1 year

All HNSCC patients will have 6 weeks of IMRT radiotherapy +/- platinum based chemotherapy as standard of care. Treatment will involve a radiotherapy planning scan and a thermoplastic shell made for each patient, which is standard of care prior to starting radiotherapy.

MRI scanning will be done with a 1.5T whole body scanner. The MRI imaging protocol will include anatomical imaging, DCE-MRI, diffusion weighted imaging (DWI) and OE-MRI sequences. All patients will have a peripheral venous cannula inserted in the upper limb and administration of gadolinium contrast prior to scanning. For the OE-MRI acquisition, each patient will wear a sealed Hudson facial mask and will breathe medical air for a period followed by 100% oxygen (at a flow rate of 15 l/min). The signal change induced by the oxygen challenge will be combined with perfusion information from the DCE-MRI images in order to identify areas of tumour hypoxia. Data analysis and production of hypoxia maps will be produced using software developed in house. A gas-sensing probe placed inside the facemask and connected to a gas analyser will be used to confirm oxygen delivery during the acquisition.

OE-MRI baseline scans will be performed on two separate days between 24 hours and 7 days in line with previous studies to measure the perfused Oxy-R volume and the tumour hypoxic volume as baseline measurements and to assess repeatability. The two scans performed at weeks 2-3 and 4-5 from the

start of radiotherapy will evaluate changes in tumour hypoxia by measuring changes in Oxy-R fraction during the course of radiotherapy. Tumour GTVs will be outlined by a clinical oncologist using pre and post gadolinium T1W images.

On completion of each MRI session, each patient will be asked to complete a questionnaire (Figure 9.1) about their experience of having an OE-MRI scan to assess tolerability and compliance to a novel imaging protocol.

Figure 9.1. Participant questionnaire

9.5.2 Tissue-based analysis

All patients will have a diagnostic biopsy taken as standard of care at their peripheral hospital. The formalin fixed paraffin embedded (FFPE) tumour blocks will be requested and tested for hypoxia using a multi-gene hypoxia signature (172). RNA will be extracted from sections of each FFPE block and used for gene expression analysis. The previously derived hypoxia signature

(8) will then be used to provide the hypoxic status of each tumour. All patients recruited will be consented to have additional testing performed on the diagnostic biopsy, e.g., for other measures of hypoxia such as expression of the hypoxia-inducible marker CAIX.

9.5.3 Clinical and imaging data collection

Patients' demographics and imaging data will be collected. These data will include date of diagnosis, stage and primary site, pathology details, smoking status, co-morbidities, performance status, treatment details including radiotherapy dose/fractionation, surgery and addition of chemotherapy. Follow up data will be collected from the date of entry until closure of the study to include treatment response, toxicities, date of relapse or progression or death. Analysis of anonymised images will be performed via well-established in-house protocols at the University of Manchester, as part of the Cancer Research UK National Translational Cancer Imaging Network.

9.5.4 Statistical analysis

The number of patients who are eligible for chemoradiotherapy treated in our institution for HNSCC per year is approximately 400-500. Assuming 60% of these patients will proceed with radiotherapy (n=200 to 300) due to poor compliance and tolerance of MR imaging in head and neck cancer, the expected accrual rate may be only 25% (n=50 to 75) per year. In this feasibility study the target recruitment is 30 which we plan to achieve in 6 months.

To analyze patient demographics and toxicity data, summary statistics such as: mean, standard deviation, median, and range will be calculated for continuous variables. For categorical variables, the frequency and percentage in each category will be calculated.

The changes in Oxy-E, Oxy-R and non-perfused will be evaluated at different time points during radiotherapy. A perfused Oxy-R fraction >50% will be

considered to be significant, in keeping with previous imaging biomarker studies (188,194).

9.5.5 Consent procedure

Verbal and written information will be given to the healthy volunteer or patient. Participants will have given sufficient time to read the information sheet (see appendix 1.2), discuss with relatives/friends and have the opportunity to ask questions. Staff authorised to obtain consent are trained and experienced and all have good clinical practice (GCP) training. Consent will be obtained by the chief investigator (CI), co-investigators and members of staff trained in the study protocol (see appendix 1.3). Participation is voluntary and can be withdrawn by participants at any time.

9.5.6 Study limitations

The first potential limitation is participant comfort during MRI scanning and compliance of which efforts have been made to make the breathing circuit as comfortable as possible. Facial mask is a non-invasive method to deliver oxygen. Medical air (21%O_2) and oxygen will be administered throughout the breathing circuit. There is no known risk of breathing oxygen for short (<15 minute) periods. Participants will require a cannula which can cause bruising and discomfort and there is a small risk with gadolinium contrast administration, with the most common side effect of nausea being rare. Medical cover will be available at all times if required.

9.5.7 Resources and Costs

There is secured funding for this study via the NIHR Manchester Biomedical Research Centre Advanced Radiotherapy theme. This will cover the research costs relating to this application.

9.5.8 Current status

The study has been delayed in starting recruitment, due to the need to change imaging sites from the Wolfson Molecular Imaging Centre to The Christie NHS Foundation Trust. This required a new research legal agreement between parties which was approved in March 2020. Due to the coronavirus pandemic however, study recruitment has not yet begun.

10.0 Discussion and future research

There has been an increase in the number of long-term survivors with head and neck cancer. The increase in survivors is partly due to improvements in treatment, such as intensity modulated radiotherapy (IMRT) but also the rising incidence of HPV positive oropharyngeal cancer, which has superior 3 year overall survival and local control rates of >80% and >90% (11,281). The relevance of reducing normal tissue toxicities to improve quality of life is therefore more important than ever (282). There are a number of potential strategies to optimise radiotherapy and to limit normal tissue toxicities in head and neck cancer. These strategies include: standardisation of normal tissue volumes when no guidelines are available; adaptive re-planning based upon structural changes or treatment response to reduce dose to normal tissues; improved therapeutic ratio of proton beam therapy as an alternative treatment modality and evaluating if patients are willing to travel to take advantage of proton beam therapy. The main findings in this thesis were:

- Limiting dose to a block, described as a composite outline of the individual muscles of mastication may reduce trismus (Chapter 3),

- A novel muscles of mastication atlas can improve consistency of contouring by helping to standardise volumes and develop more precise dose parameters for use in avoidance planning (Chapter 4),

- A novel combined CT deep learning auto-contouring model (model$_{CT}$) can reduce time and inter-observer variability compared with manual contours in OAR delineation (Chapter 5),

- A novel MR deep learning auto-contouring model (model$_{MRI}$) has potential to improve geometric accuracy compared with CT based models but requires further sequence optimisation and validation. The model should be explored further for use in MRI-guided adaptive radiotherapy (Chapter 6),

- Multi-field optimisation (MFO) is robust to uncertainties in set up and range and has the potential to help in the development of optimisation

protocols with agreed parameters, to identify plans that may require further individualisation (Chapter 7),

- Feedback from patient and public involvement identified a general willingness to travel and stay away from home to receive proton beam therapy as part of the UK's first proton trial TORPEdO, ensuring adequate support was available (Chapter 8),

- The development of a protocol for OE-MRI in head and neck cancer, highlighted the inadequacy of nasal cannula to deliver oxygen to detect an adequate signal. Ongoing work to evaluate feasibility of OE-MRI in head and neck cancer will be performed using a facial mask (Chapter 9).

In Chapter 3, I defined a dose threshold of ≤40 Gy to the ipsilateral lateral pterygoid, masseter and block for tumours not invading the masticatory muscles, to help reduce the incidence of trismus. This was an important aim to add clarity to the wide variation in dosimetric and clinical factors associated with trismus, in order to provide an OAR avoidance structure, to improve dose delivery and functional outcome (103,212,283–285). I chose to evaluate a block, which is the composite outline of the individual masticatory muscles, to account for the close proximity of structures and to remove uncertainties regarding which individual muscles to avoid. The block included areas at the muscles interface such as nerves and fascia, which as identified by Beasley et al. using image based data mining, have a significant association with the development of trismus (224).

The results presented in Chapter 3 are hypothesis generating and are limited by the small sample size and short follow up. Despite the small sample size, valid inferences of the data could be made, due to using a continuous scale to measure pre and post mouth opening measurements to record trismus, which is a potential strength of the study. A future study which extends follow up beyond 6 months and includes a matched control cohort, will help evaluate the association of dosimetric and clinical factors with the development of trismus and has the potential to capture quality of life data, as evidenced by Karsten et al. (286). Whilst not yet evaluated in this study,

the block has the potential to reduce time and inter-observer variability and be developed into an auto contouring model to necessitate quick and efficient contouring in the adaptive radiotherapy workflow. Development of a normal tissue complication probability (NTCP) model to make further dose constraint recommendations for tumours invading the masticatory muscles, whereby limiting dose to ≤40 Gy, might not be achievable, is an area of future work.

The variation in trismus aetiologies support risk prediction models to provide a more personalised and preventative approach to treatment. In an original study by van der Geer et al. significant prognostic factors such as age, tumour site/location and treatment modality were included into a model and a risk score calculated (93). Using a calculated risk score could enable precautionary measures such as exercise prevention to be adopted earlier in the assessment and treatment phase in patients most at risk (97,287). The development of a risk prediction model to include the significant findings of chapter 3 is an area of future interest and one that may provide a more personalised approach to treatment to improve functional outcomes and overall survival.

There is a substantial lack of published guidelines and contouring atlases for the muscles of mastication to standardise volumes and improve the consistency of contours. The study presented in Chapter 4 is the first to describe a novel muscles of mastication atlas, which is now available as a free app to guide clinician contouring. Using the webpage through google chrome, https://app.bitly.com I was able to view the number and location of people using the app. To date, 153 clicks on the app have been made including in India, with the majority in the UK. The atlas significantly improved consistency between clinicians' contours. In particular this improvement was most evident at the cranio-caudal extremities of the muscles. This improvement was more apparent when trainees delineated contours as compared with consultants which demonstrates the potential benefit of the atlas as an educational tool for trainees and planning radiographers. Once used as a training tool, contours generated by planning radiographers using the atlas as a guideline can be compared against a gold standard clinician contour. This use has the potential to reduce clinician workload and improve

intra and inter-observer variability. Dosimetric relationships can be evaluated by comparing contours generated with and without the atlas, supporting the potential role of the atlas to improve the accuracy of plan dosimetry.

The atlas is available for use by other radiotherapy centres but its performance to guide contouring in a centre different to where it was developed has not yet been assessed. Validation of the atlas in a larger, heterogenous group using data from more than one institution will improve the robustness and explore use of the atlas in the development of multi-institutional trials. Future studies to develop the atlas into an auto-contouring model to reduce time and variation in adaptive radiotherapy should also be considered.

Auto contouring models reduce time and inter-observer variability and yet none have been widely adopted for routine clinical use (107,234,288,289). The lack of adoption is partly due to clinicians' concerns about the accuracy of the models and the challenges with quality assurance and validation (290). Newer techniques of artificial intelligence, such as deep learning auto contouring, have shown improved treatment plan consistency and quality compared with atlas-based models (117). The confirmed benefits of deep learning methods as evidenced by van Dikj et al. led to the development of the first combined CT auto-contouring model ($model_{CT}$) in Chapter 5 for OAR delineation. The assessment of $model_{CT}$ edited contours to reduce inter-observer variability was evaluated by four observers from a single institution, which provides a proof of principle concept but requires validation in a larger, multi-centre cohort to evaluate the external validity of the model (291).

$Model_{CT}$ was trained on data from more than one institution, reducing the potential for contouring bias compared to published models (107). Qualitative and quantitative metrics assessed the geometric similarities between $model_{CT}$ edited contours against manual contours, improving the clinical application of the model (202,292). Using both metrics however was challenging when scores conflicted for an OAR. This potential limitation could be minimised in future work by agreeing which scoring method based on clinical importance should be used prior to the study.

To increase the clinical usability of model$_{CT}$, to limit normal tissue toxicities, additional training and testing of OARs such as the pharyngeal constrictor muscles, muscles of mastication and optic apparatus is needed. The larger library of OARs will increase its use in adaptive re-planning to reduce normal tissue toxicities and improve patient outcome. There is a suggestion that deep learning auto contouring models may have a role in clinical target volume (CTV) and lymph node delineation. Due to clinician bias and the heterogeneity of clinical scenarios, auto-contouring tools are less accurate than manual contours for defining the CTV and may have a role in producing a template that requires adjustment, as evidenced by Wong et al. (106). The post-segmentation manual edits whilst time consuming are predicted to be less than manual contours alone, supporting the integration of auto-contouring models into the adaptive workflow (106).

Following evaluation of model$_{CT}$ (Chapter 5), I developed the first MR deep learning auto contouring model (model$_{MRI}$) to exploit the superior soft tissue discrimination of MR and produce contours closer to the ground truth than possible with CT auto-segmentation (69,132,133,240). The results of Chapter 6 are the first to highlight the challenges of producing one MR auto contouring model for generic use, due to the differences in MR image acquisition parameters.

The differences in slice thickness and contrast: noise ratio on the 3D MRL compared with the T2W 2D diagnostic and 2D radiotherapy planning images is one explanation for the discordance of model$_{MRI}$ automated contours with manual contours on the 3D MRL. Potential strategies to optimise the MR-images to improve the performance of model$_{MRI}$ include: (1) developing an optimised 2D and 3D sequence for the MRL and comparing the models performance on each; (2) training the model using MRL images rather than T2W diagnostic (however, due to the current shortage of MRL scans this option would not be feasible) and (3) training the model on a larger library of OARs to evaluate if the model's performance on the 3D MRL is influenced by the type of OAR.

Model$_{MRI}$ produced contours closer to the ground truth compared with model$_{CT}$ in Chapter 5, which suggests a potential development to replace CT with MR based models in head and neck planning. To support this statement, model$_{MRI}$ would need training on a larger number of OAR structures and validating using data from more than one institution to improve its external validity. An alternative to replacing CT with MR would be to co-register both imaging modalities into an auto-contouring solution. This hybrid solution would exploit the benefits and shortfalls of both and lead to future work addressing the value of fully automated models. Fully automated models have the potential to remove all human interaction and have been evaluated to critically appraise and detect errors in manual contours (14). Progressing full automation requires large datasets, an agreed and reliable ground truth, an accurate validation process and adequate image quality (115). The removal of human interaction from contouring is an area of debate, due to concerns about central decision making and ensuring a safe and good quality plan. As clinician workload and the complexity of cases increases a compromise is needed between auto contouring models to deliver a quick and efficient re-planning volume and clinician input to ensure optimal tumour coverage and de-escalation of dose to normal tissues.

Proton beam therapy can reduce doses to critical structures, but is also sensitive to inter-fraction uncertainties in patient set up and beam range (293,294). To ensure proton plans are safe and effective it is fundamental these uncertainties are addressed. The results presented in Chapter 7 are the first to evaluate the robustness of plans in post-operative oropharyngeal and oral cavity cancers. The findings demonstrate multi-field optimisation is robust to inter-fraction uncertainties in set-up and range when evaluated over multiple CT scans, without compromising mean OAR dose in post-operative oropharyngeal and oral cavity cancer. The study also highlights the association of small anatomical changes in jaw or neck position and the development of hot spots (> 110%), which often cannot be predicted.

Future studies are needed to evaluate MFO using daily cone beam CT scans (CBCTs) and in a larger cohort to perform a more thorough analysis of plan robustness. Daily CBCTs will enable a protocol to be developed for plan

robustness with agreed parameters, to identify those plans that need additional individualisation. By developing a robust analysis protocol, consistency of reporting and plan evaluation amongst radiotherapy centres will be improved and have the potential to be used in multi-institutional trials (295). Since the UK opened the first proton centre in 2018, there is a need to compare photon and proton plans in UK based trials, to gain important information about what constitutes a robust treatment plan and which patients may benefit the most from proton beam therapy (296). The important information about what constitutes a robust plan will help produce plans that are both safe and effective, even if dose variations occur.

The penultimate chapter of the thesis addresses the initial steps and early patient and public involvement in helping to inform the design of the UK's first proton trial TORPEdO (TOxicity Reduction using Proton bEam therapy for Oropharyngeal cancer) (297). Gaining patient feedback on the study design, endpoints and their willingness to travel and stay away from home, whilst undergoing treatment were invaluable, to better inform and shape the trial design, which opened to recruitment in February 2020. This piece of research is the first to highlight the willingness of patients to travel and stay away from home to participate in clinical trials, with adequate support. Understanding patients views on the logistics involved if receiving treatment away from home, such as favouring apartment based rather than hotel accommodation was important, as the need for patients to travel becomes increasingly more common, due to being limited to one proton centre in the UK. Furthermore, willingness to travel, learning about what toxicities are most important will help shape future trials and better inform clinicians when deciding on treatment. In addition having a patient-centred approach when designing a clinical trial will improve recruitment and compliance as well as providing an insight into patients thoughts and their alternative views to researchers (298,299). Patient reported outcomes can also be used electronically to capture self-reported symptoms which as we move to a more virtual way of working will help inform clinicians and clinical trials.

The final chapter describes the first study protocol which has received ethical approval to assess the tolerability and clinical feasibility of OE-MRI in head and neck radiotherapy planning. I plan to investigate the ability of OE-MRI to identify and help target dose to areas of tumour hypoxia with a view to treating smaller volumes and reducing toxicity to neighbouring OARs. Study recruitment has not yet begun due to delay incurred by needing to obtain tri-party regulatory approval to move from the Wolfson Molecular Imaging Centre to The Christie NHS Foundation Trust for imaging purposes. The study was ready to begin recruitment at the end of March 2020, but opening was delayed due to the COVID-19 pandemic.

In conclusion, the results of this thesis outline a number of novel techniques to optimise OAR delineation and reduce long term toxicities in head and neck cancer. All techniques described, on balance have a role in improving patient outcome. Adaptive re-planning has proven dosimetric benefits in head and neck cancer by increasing tumour control by targeting radio-resistant areas and reducing dose to normal tissues through smaller contoured volumes (300,301). The integration of MR based imaging techniques and auto contouring models into the adaptive radiotherapy workflow, have the potential to reduce dose to OARs, by increasing the contouring accuracy and volume definition. Furthermore auto-contouring models will help reduce clinician workload and aid larger data collection with a view to developing randomised controlled trials, evaluating the benefits of adaptive radiotherapy in clinical practice. The improved dose conformity of proton beam therapy is uniquely suited to head and neck cancer to reduce OAR toxicity owing to the large number and close proximity of healthy structures to the target tumour. Prospective randomised trials are needed to establish the clinical usefulness and toxicity reduction of proton beam therapy in head and neck cancer for wider adoption in clinical practice. The culmination of work outlined in this thesis will help to limit normal tissue toxicity whilst maintaining adequate tumour control in patients with OPC.

References

1.   Http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/head-and-neck-cancers/s. Cancer Research UK statistics.

2.   UK CR. Cancer Incidence for Common Cancers. Cancer Research UK. 2013;1.

3.   Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2018 Nov;68(6):394–424.

4.   Lee Y-CA, Al-Temimi M, Ying et al.Risk Prediction Models for Head and Neck Cancer in the US Population From the INHANCE Consortium. American Journal of Epidemiology. 2020 Apr 2;189(4):330–42.

5.   Http://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-. Cancer Research UK.

6.   Office for National Statistics. Overview of the UK Population. Overview of the UK population August 2019. 2019;

7.   Schache AG, Powell NG, Cuschieri KS, et al. HPV-related oropharynx cancer in the United Kingdom: An evolution in the understanding of disease etiology. Cancer Research. 2016;76(22):6598–606.

8.   Reddy VM, Cundall-Curry D, Bridger MWM. Trends in the incidence rates of tonsil and base of tongue cancer in England, 1985-2006. Annals of the Royal College of Surgeons of England. 2010;92(8):655–9.

9.   O'Sullivan B, Huang SH, Siu LL, et al. Deintensification candidate subgroups in human papillomavirus-related oropharyngeal cancer according to minimal risk of distant metastasis. Journal of Clinical Oncology. 2013;31(5):543–50.

10.   Posner MR, Lorch JH, Goloubeva O, et al. Survival and human papillomavirus in oropharynx cancer in TAX 324: a subset analysis from an international phase III trial. Annals of Oncology. 2011 May;22(5):1071–7.

11.   Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. The New England journal of medicine. 2010;363(1):24–35.

12.   Eriksson PO, Thornell LE. Histochemical and morphological muscle-fibre characteristics of the human masseter, the medial pterygoid and the temporal muscles. Archives of Oral Biology. 1983;28(9):781–95.

13.   Krasin MJ, Wiese KM, Spunt SL, et al. Jaw dysfunction related to pterygoid and masseter muscle dosimetry after radiation therapy in children and young adults with head-and-neck sarcomas. International Journal of Radiation Oncology Biology Physics. 2012;82(1):355–60.

14. Ramqvist T, Dalianis T. An epidemic of oropharyngeal squamous cell carcinoma (OSCC) due to human papillomavirus (HPV) infection and aspects of treatment and prevention. Anticancer research. 2011;31(5):1515–9.

15. Gillison ML. Evidence for a Causal Association Between Human Papillomavirus and a Subset of Head and Neck Cancers. Journal of the National Cancer Institute. 2000 May 3;92(9):709–20.

16. O'Sullivan B, Huang SH, Su J, et al. Development and validation of a staging system for HPV-related oropharyngeal cancer by the International Collaboration on Oropharyngeal cancer Network for Staging (ICON-S): a multicentre cohort study. The Lancet Oncology. 2016 Apr;17(4):440–51.

17. Lydiatt WM, Patel SG, O'Sullivan B, et al. Head and neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual. CA: A Cancer Journal for Clinicians. 2017 Mar;67(2):122–37.

18. de Almeida JR, Byrd JK, Wu R, et al. A systematic review of transoral robotic surgery and radiotherapy for early oropharynx cancer: A systematic review. The Laryngoscope. 2014 Sep;124(9):2096–102.

19. Parsons JT, Mendenhall WM, Stringer SP, et al. Squamous cell carcinoma of the oropharynx. Cancer. 2002 Jun 1;94(11):2967–80.

20. Pignon JP, Maître A le, Maillard E, et al. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): An update on 93 randomised trials and 17,346 patients. Radiotherapy and Oncology. 2009;92(1):4–14.

21. Bourhis J, Sire C, Graff P, et al. Concomitant chemoradiotherapy versus acceleration of radiotherapy with or without concomitant chemotherapy in locally advanced head and neck carcinoma (GORTEC 99-02): an open-label phase 3 randomised trial. The Lancet Oncology. 2012 Feb;13(2):145–53.

22. Mehanna H, Robinson M, Hartley A, et al. Radiotherapy plus cisplatin or cetuximab in low-risk human papillomavirus-positive oropharyngeal cancer (De-ESCALaTE HPV): an open-label randomised controlled phase 3 trial. The Lancet. 2019 Jan;393(10166):51–60.

23. Ferris RL, Flamand Y, Weinstein GS, et al. Transoral robotic surgical resection followed by randomization to low- or standard-dose IMRT in resectable p16+ locally advanced oropharynx cancer: A trial of the ECOG-ACRIN Cancer Research Group (E3311). Journal of Clinical Oncology. 2020 May 20;38(15_suppl):6500–6500.

24. Owadally W, Hurt C, Timmins H, et al. PATHOS: a phase II/III trial of risk-stratified, reduced intensity adjuvant treatment in patients undergoing transoral surgery for Human papillomavirus (HPV) positive oropharyngeal cancer. BMC Cancer. 2015 Dec 27;15(1):602.

25. Wang X, Eisbruch A. IMRT fTHEor head and neck cancer: Reducing xerostomia and dysphagia. In: Journal of Radiation Research. 2016. p.

i69–75.

26. The Royal College of Radiologists. Radiotherapy dose fractionation, third edition. Clinical Oncology. 2019;40.

27. Beasley M. Complications of radiotherapy: improving the therapeutic index. Cancer Imaging. 2005;5(1):78–84.

28. Dearnaley D, Syndikus I, Mossop H, et al. Conventional versus hypofractionated high-dose intensity-modulated radiotherapy for prostate cancer: 5-year outcomes of the randomised, non-inferiority, phase 3 CHHiP trial. The Lancet Oncology. 2016;17(8):1047–60.

29. Levendag PC, Teguh DN, Voet P, et al. Dysphagia disorders in patients with cancer of the oropharynx are significantly affected by the radiation therapy dose to the superior and middle constrictor muscle: A dose-effect relationship. Radiotherapy and Oncology. 2007 Oct;85(1):64–73.

30. Mohan R, Wu Q, Manning M,et al. Radiobiological considerations in the design of fractionation strategies for intensity-modulated radiation therapy of head and neck cancers. International Journal of Radiation Oncology*Biology*Physics. 2000 Feb;46(3):619–30.

31. Stromberger C, Ghadjar P, Marnitz S, et al. Comparative treatment planning study on sequential vs. simultaneous integrated boost in head and neck cancer patients. Strahlentherapie und Onkologie. 2016 Jan 13;192(1):17–24.

32. Dogan N, King S, Emami B, et al. Assessment of different IMRT boost delivery methods on target coverage and normal-tissue sparing. International Journal of Radiation Oncology*Biology*Physics. 2003 Dec;57(5):1480–91.

33. Dragan T, Beauvois S, Moreau M, et al. Clinical outcome and toxicity after simultaneous integrated boost IMRT in head and neck squamous cell cancer patients. Oral Oncology. 2019 Nov;98:132–40.

34. Verbakel WFAR, Cuijpers JP, Hoffmans D,et al. Volumetric Intensity-Modulated Arc Therapy Vs. Conventional IMRT in Head-and-Neck Cancer: A Comparative Planning and Dosimetric Study. International Journal of Radiation Oncology Biology Physics. 2009;74(1):252–9.

35. Cilla S, Deodato F, MacChia G, et al. Volumetric modulated arc therapy (VMAT) and simultaneous integrated boost in head-and-neck cancer: Is there a place for critical swallowing structures dose sparing? British Journal of Radiology. 2016;89(1059).

36. Nithya L, Raj NAN, Kumar A, et al. Comparative analysis of volumetric-modulated arc therapy and intensity-modulated radiotherapy for base of tongue cancer. Journal of medical physics / Association of Medical Physicists of India. 2014;39(2):121–6.

37. Lee TF, Fang FM. Quantitative analysis of normal tissue effects in the clinic (QUANTEC) guideline validation using quality of life questionnaire datasets for parotid gland constraints to avoid causing xerostomia during head-and-neck radiotherapy. Radiotherapy and

Oncology. 2013;106(3):352–8.

38. Anderson NJ, Rolfo M, Khoo V, et al. Dose Volume Response in Acute Dysphagia Toxicity: Validating QUANTEC Recommendations into Clinical Practice for Head and Neck Radiotherapy. International Journal of Radiation Oncology Biology Physics. 2011;81(2):S76.

39. Thomson D, Yang H, Baines H, et al. NIMRAD - A phase III trial to investigate the use of nimorazole hypoxia modification with intensity-modulated radiotherapy in head and neck cancer. Clinical Oncology. 2014;26(6):344–7.

40. 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.

41. Landberg T, Chavaudra J, Dobbs H. ICRU Report 50: Prescribing, Recording, and Reporting Photon Beam Therapy. … , Recording and Reporting …. 1993. 1–81 p.

42. Daisne J-F, Duprez T, Weynand B, et al. Tumor Volume in Pharyngolaryngeal Squamous Cell Carcinoma: Comparison at CT, MR Imaging, and FDG PET and Validation with Surgical Specimen. Radiology. 2004 Oct;233(1):93–100.

43. De Felice F, Thomas C, Barrington S, et al. Analysis of loco-regional failures in head and neck cancer after radical radiation therapy. Oral Oncology. 2015 Nov;51(11):1051–5.

44. Caudell JJ, Meredith RF, Spencer SA, et al. Margin on Gross Tumor Volume and Risk of Local Recurrence in Head-and-Neck Cancer. International Journal of Radiation Oncology*Biology*Physics. 2010 Jan;76(1):164–8.

45. Hansen CR, Johansen J, Samsøe E, et al. Consequences of introducing geometric GTV to CTV margin expansion in DAHANCA contouring guidelines for head and neck radiotherapy. Radiotherapy and Oncology. 2018 Jan;126(1):43–7.

46. Grégoire V, Evans M, Le Q-T, et al. Delineation of the primary tumour Clinical Target Volumes (CTV-P) in laryngeal, hypopharyngeal, oropharyngeal and oral cavity squamous cell carcinoma: AIRO, CACA, DAHANCA, EORTC, GEORCC, GORTEC, HKNPCSG, HNCIG, IAG-KHT, LPRHHT, NCIC CTG, NCRI, NRG Oncolog. Radiotherapy and Oncology. 2018 Jan;126(1):3–24.

47. van der Veen J, Gulyban A, Nuyts S. Interobserver variability in delineation of target volumes in head and neck cancer. Radiotherapy and Oncology. 2019 Aug;137:9–15.

48. Corkum MT, Mitchell S, Venkatesan V, et al. Does 5 + 5 Equal Better Radiation Treatment Plans in Head and Neck Cancers? Advances in Radiation Oncology. 2019 Oct;4(4):683–8.

49. Barker JL, Garden AS, Ang KK, et al. Quantification of volumetric and

geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. International Journal of Radiation Oncology Biology Physics. 2004;59(4):960–70.

50. Bhide SA, Davies M, Burke K, , et al. Weekly Volume and Dosimetric Changes During Chemoradiotherapy With Intensity-Modulated Radiation Therapy for Head and Neck Cancer: A Prospective Observational Study. International Journal of Radiation Oncology Biology Physics. 2010;76(5):1360–8.

51. Van Herk M. Errors and Margins in Radiotherapy. Vol. 14, Seminars in Radiation Oncology. 2004. p. 52–64.

52. Van Herk M, Remeijer P, Rasch C, et al. The probability of correct target dosage: Dose-population histograms for deriving treatment margins in radiotherapy. International Journal of Radiation Oncology Biology Physics. 2000;47(4):1121–35.

53. van Kranen S, Hamming-Vrieze O, Wolf A, et al. Head and Neck Margin Reduction With Adaptive Radiation Therapy: Robustness of Treatment Plans Against Anatomy Changes. International Journal of Radiation Oncology Biology Physics. 2016;96(3):653–60.

54. Jin X, Hu W, Shang H, et al. CBCT-based volumetric and dosimetric variation evaluation of volumetric modulated arc radiotherapy in the treatment of nasopharyngeal cancer patients. Radiat Oncol. 2013;8:279.

55. Nishi T, Nishimura Y, Shibata T, et al.Volume and dosimetric changes and initial clinical experience of a two-step adaptive intensity modulated radiation therapy (IMRT) scheme for head and neck cancer. Radiotherapy and Oncology. 2013;106(1):85–9.

56. Robar JL, Day A, Clancey J, et al. Spatial and Dosimetric Variability of Organs at Risk in Head-and-Neck Intensity-Modulated Radiotherapy. International Journal of Radiation Oncology Biology Physics. 2007;68(4):1121–30.

57. Hunter KU, Fernandes LL, Vineberg KA, et al. Parotid glands dose-effect relationships based on their actually delivered doses: Implications for adaptive replanning in radiation therapy of head-and-neck cancer. International Journal of Radiation Oncology Biology Physics. 2013;87(4):676–82.

58. Ho KF, Marchant T, Moore C, et al. Monitoring dosimetric impact of weight loss with kilovoltage (KV) cone beam CT (CBCT) during parotid-sparing IMRT and concurrent chemotherapy. International Journal of Radiation Oncology Biology Physics. 2012;82(3).

59. Castadot P, Geets X, Lee JA, et al. Adaptive functional image-guided IMRT in pharyngo-laryngeal squamous cell carcinoma: Is the gain in dose distribution worth the effort? Radiotherapy and Oncology. 2011;101(3):343–50.

60. Fiorentino A, Caivano R, Metallo V, et al. Parotid gland volumetric

changes during intensity-modulated radiotherapy in head and neck cancer. The British Journal of Radiology. 2012;85(1018):1415–9.

61. Brouwer CL, Steenbakkers RJHM, Langendijk JA, et al. Identifying patients who may benefit from adaptive radiotherapy: Does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help? Vol. 115, Radiotherapy and Oncology. 2015. p. 285–94.

62. Chetty IJ, Rosu-Bubulac M. Deformable Registration for Dose Accumulation. Seminars in Radiation Oncology. 2019 Jul;29(3):198–208.

63. Veiga C, McClelland J, Moinuddin S, et al. Toward adaptive radiotherapy for head and neck patients: Feasibility study on using CT-to-CBCT deformable registration for "dose of the day" calculations. Medical Physics. 2014 Feb 19;41(3):031703.

64. Broggi S, Fiorino C, Dell'Oca I, et al. A two-variable linear model of parotid shrinkage during IMRT for head and neck cancer. Radiotherapy and Oncology. 2010 Feb;94(2):206–12.

65. Brouwer CL, Steenbakkers RJHM, van der Schaaf A, et al. Selection of head and neck cancer patients for adaptive radiotherapy to decrease xerostomia. Radiotherapy and Oncology. 2016 Jul;120(1):36–40.

66. Duprez F, De Neve W, De Gersem W, et al. Adaptive Dose Painting by Numbers for Head-and-Neck Cancer. International Journal of Radiation Oncology*Biology*Physics. 2011 Jul;80(4):1045–55.

67. de Ridder M, Gouw ZAR, Sonke JJ, et al. Recurrent oropharyngeal cancer after organ preserving treatment: pattern of failure and survival. European Archives of Oto-Rhino-Laryngology. 2017 Mar 9;274(3):1691–700.

68. Mohamed ASR, Aristophanous M, Blanchard P, et al. Prospective In Silico Study of A Novel MRI-Guided Dose Adaptation Technique For Human Papillomavirus–Positive Oropharyngeal Cancer Patients. International Journal of Radiation Oncology*Biology*Physics. 2018 Apr;100(5):1362–3.

69. Li L, Sun J, Li B, et al. Computed tomography versus magnetic resonance imaging for diagnosing cervical lymph node metastasis of head and neck cancer: systematic review and meta-analysis. OncoTargets and Therapy. 2015 Jun;1291.

70. Dirix P, Vandecaveye V, De Keyzer F, et al. Diffusion-Weighted MRI for Nodal Staging of Head and Neck Squamous Cell Carcinoma: Impact on Radiotherapy Planning. International Journal of Radiation Oncology*Biology*Physics. 2010 Mar;76(3):761–6.

71. Vandecaveye V, De Keyzer F, Nuyts S, et al. Detection of head and neck squamous cell carcinoma with diffusion weighted MRI after (chemo)radiotherapy: Correlation between radiologic and histopathologic findings. International Journal of Radiation Oncology*Biology*Physics. 2007 Mar;67(4):960–71.

72. Hamming-Vrieze O, van Kranen SR, Heemsbergen WD, et al. Analysis of GTV reduction during radiotherapy for oropharyngeal cancer: Implications for adaptive radiotherapy. Radiotherapy and Oncology. 2017 Feb;122(2):224–8.

73. Klüter S. Technical design and concept of a 0.35 T MR-Linac. Clinical and Translational Radiation Oncology. 2019 Sep;18:98–101.

74. Brown E, Porceddu S, Owen R, et al. Developing an Adaptive Radiotherapy Technique for Virally Mediated Head and Neck Cancer. Journal of Medical Imaging and Radiation Sciences. 2013 Sep;44(3):134–40.

75. Chuter RW, Pollitt A, Whitehurst P, et al. Assessing MR-linac radiotherapy robustness for anatomical changes in head and neck cancer. Physics in Medicine & Biology. 2018 Jun 20;63(12):125020.

76. Bahig H, Yuan Y, Mohamed ASR, et al. Magnetic Resonance-based Response Assessment and Dose Adaptation in Human Papilloma Virus Positive Tumors of the Oropharynx treated with Radiotherapy (MR-ADAPTOR): An R-IDEAL stage 2a-2b/Bayesian phase II trial. Clinical and Translational Radiation Oncology. 2018 Nov;13:19–23.

77. Corradini S, Alongi F, Andratschke N, et al. MR-guidance in clinical reality: current treatment challenges and future perspectives. Radiation Oncology. 2019 Dec 3;14(1):92.

78. Langendijk JA, Doornaert P, Verdonck-de Leeuw IM, et al. Impact of Late Treatment-Related Toxicity on Quality of Life Among Patients With Head and Neck Cancer Treated With Radiotherapy. Journal of Clinical Oncology. 2008 Aug 1;26(22):3770–6.

79. Franzén L, Funegård U, Ericson T, et al. Parotid gland function during and following radiotherapy of malignancies in the head and neck. European Journal of Cancer. 1992 Feb;28(2–3):457–62.

80. Wijers OB, Levendag PC, Braaksma MMJ, et al. Patients with head and neck cancer cured by radiation therapy: A survey of the dry mouth syndrome in long-term survivors. Head & Neck. 2002 Aug;24(8):737–47.

81. Langendijk JA, Doornaert P, Rietveld DHF, et al. A predictive model for swallowing dysfunction after curative radiotherapy in head and neck cancer. Radiotherapy and Oncology. 2009 Feb;90(2):189–95.

82. Mortensen HR, Overgaard J, Jensen K, et al. Factors associated with acute and late dysphagia in the DAHANCA 6 & 7 randomized trial with accelerated radiotherapy for head and neck cancer. Acta Oncologica. 2013 Oct 19;52(7):1535–42.

83. Hawkins PG, Kadam AS, Jackson WC, et al. Organ-Sparing in Radiotherapy for Head-and-Neck Cancer: Improving Quality of Life. Seminars in Radiation Oncology. 2018 Jan;28(1):46–52.

84. Nutting CM, Morden JP, Harrington KJ, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. The

lancet oncology. 2011;12(2):127–36.

85. Christianen MEMC, Van Der Schaaf A, Van Der Laan HP, et al. Swallowing sparing intensity modulated radiotherapy (SW-IMRT) in head and neck cancer: Clinical validation according to the model-based approach. Radiotherapy and Oncology. 2016;118(2):298–303.

86. Roe JWG, Drinnan MJ, Carding PN, et al. Patient-reported outcomes following parotid-sparing intensity-modulated radiotherapy for head and neck cancer. How important is dysphagia? Oral Oncology. 2014;50(12):1182–7.

87. Hedström J, Tuomi L, Finizia C, et al. Identifying organs at risk for radiation-induced late dysphagia in head and neck cancer patients. Clinical and Translational Radiation Oncology. 2019 Nov;19:87–95.

88. Thomas F, Ozanne F, Mamelle G, et al. Radiotherapy alone for oropharyngeal carcinomas: the role of fraction size (2 Gy vs 2.5 Gy) on local control and early and late complications. International Journal of Radiation Oncology, Biology, Physics. 1988;15(5):1097–102.

89. Steelman R, Sokol J. Quantification of trismus following irradiation of the temporomandibular joint. MO Dent J. 1986;66:21–3.

90. Dijkstra PU, Huisman PM, Roodenburg JLN. Criteria for trismus in head and neck oncology. International Journal of Oral and Maxillofacial Surgery. 2006;35:337–42.

91. Pauli N, Johnson J, Finizia C, et al. The incidence of trismus and long-term impact on health-related quality of life in patients with head and neck cancer. Acta oncologica (Stockholm, Sweden). 2013;52(6):1137–45.

92. Dijkstra PU, Kalk WWI, Roodenburg JLN. Trismus in head and neck oncology: A systematic review. Vol. 40, Oral Oncology. 2004. p. 879–89.

93. van der Geer SJ, van Rijn P V., Kamstra JI, et al. Prevalence and prediction of trismus in patients with head and neck cancer: A cross-sectional study. Head & Neck. 2019 Jan;41(1):64–71.

94. Gebre-Medhin M, Haghanegi M, Robért L, et al. Dose-volume analysis of radiation-induced trismus in head and neck cancer patients. Acta oncologica (Stockholm, Sweden). 2016;55(11):1313–7.

95. Steiner F, Evans J, Marsh R, et al. Mouth opening and trismus in patients undergoing curative treatment for head and neck cancer. International Journal of Oral and Maxillofacial Surgery. 2015 Mar;44(3):292–6.

96. Infante-Cossio P, Torres-Carranza E, Cayuela A, et al. Impact of treatment on quality of life for oral and oropharyngeal carcinoma. International Journal of Oral and Maxillofacial Surgery. 2009 Oct;38(10):1052–8.

97. Lee R, Yeo ST, Rogers SN, et al. Randomised feasibility study to compare the use of Therabite ® with wooden spatulas to relieve and

prevent trismus in patients with cancer of the head and neck. British Journal of Oral and Maxillofacial Surgery. 2018 May;56(4):283–91.

98. Johnson J, Johansson M, Rydén A, et al. Impact of trismus on health-related quality of life and mental health. Head & Neck. 2015 Nov;37(11):1672–9.

99. Bensadoun R-J, Riesenbeck D, Lockhart PB, et al. A systematic review of trismus induced by cancer therapies in head and neck cancer patients. Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer. 2010;18(8):1033–8.

100. Lindblom U, Gärskog O, Kjellén E, et al. Radiation-induced trismus in the ARTSCAN head and neck trial. Acta Oncologica. 2014;53(5):620–7.

101. Hsieh L-C, Chen JW, Wang L-Y, et al. Predicting the severity and prognosis of trismus after intensity-modulated radiation therapy for oral cancer patients by magnetic resonance imaging. PloS one. 2014;9(3):e92561.

102. Louise Kent M, Brennan MT, Noll JL, et al. Radiation-induced trismus in head and neck cancer patients. Supportive Care in Cancer. 2008;16(3):305–9.

103. Rao SD, Saleh ZH, Setton J, et al. Dose-volume factors correlating with trismus following chemoradiation for head and neck cancer. Acta oncologica (Stockholm, Sweden). 2016;55:99–104.

104. Van Der Molen L, Heemsbergen WD, De Jong R, et al. Dysphagia after chemoradiotherapy Dysphagia and trismus after concomitant chemo-Intensity-Modulated Radiation Therapy (chemo-IMRT) in advanced head and neck cancer; Dose-effect relationships for swallowing and mastication structures. Radiotherapy and Oncology. 2013;106:364–9.

105. Teguh DN, Levendag PC, Voet P, et al. Trismus in patients with oropharyngeal cancer: Relationship with dose in structures of mastication apparatus. Head and Neck. 2008;30:622–30.

106. Wong J, Fong A, McVicar N, et al. Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. Radiotherapy and Oncology. 2020 Mar;144:152–8.

107. Lim JY, Leech M. Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck. Vol. 55, Acta Oncologica. 2016. p. 799–806.

108. Stapleford LJ, Lawson JD, Perkins C, et al. Evaluation of Automatic Atlas-Based Lymph Node Segmentation for Head-and-Neck Cancer. International Journal of Radiation Oncology*Biology*Physics. 2010 Jul;77(3):959–66.

109. Teguh DN, Levendag PC, Voet PWJ, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. International Journal of Radiation Oncology Biology Physics. 2011;81(4):950–7.

110. Hong TS, Tomé WA, Harari PM. Heterogeneity in head and neck IMRT target design and clinical practice. Radiotherapy and Oncology. 2012 Apr;103(1):92–8.

111. Loo SW, Martin WMC, Smith P, et al. Interobserver variation in parotid gland delineation: A study of its impact on intensity-modulated radiotherapy solutions with a systematic review of the literature. British Journal of Radiology. 2012;85(1016):1070–7.

112. Geets X, Daisne JF, Arcangeli S, et al. Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: Comparison between CT-scan and MRI. Radiotherapy and Oncology. 2005;77(1):25–31.

113. Walker G V., Awan M, Tao R, et al. Prospective randomized double-blind study of atlas-based organ-at-risk autosegmentation-assisted radiation planning in head and neck cancer. Radiotherapy and Oncology. 2014 Sep;112(3):321–5.

114. Ayyalusamy A, Vellaiyan S, Subramanian S, et al. Auto-segmentation of head and neck organs at risk in radiotherapy and its dependence on anatomic similarity. Radiation Oncology Journal. 2019 Jun 30;37(2):134–42.

115. Galimova RM, Buzaev IV, Ramilevich KA, et al. Artificial intelligence—Developments in medicine in the last two years. Chronic Diseases and Translational Medicine. 2019 Mar;5(1):64–8.

116. Jarrett D, Stride E, Vallis K, et al. Applications and limitations of machine learning in radiation oncology. The British Journal of Radiology. 2019 Aug;92(1100):20190001.

117. van Dijk L V., Van den Bosch L, Aljabar P, et al. Improving automatic delineation for head and neck organs at risk by Deep Learning Contouring. Radiotherapy and Oncology. 2020 Jan;142:115–23.

118. Thomson D, Boylan C, Liptrot T, et al. Evaluation of an automatic segmentation algorithm for definition of head and neck organs at risk. Radiation oncology (London, England). 2014;9:173.

119. Voet PWJ, Dirkx MLP, Teguh DN, et al. Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis. Radiotherapy and Oncology. 2011 Mar;98(3):373–7.

120. Yang J, Beadle BM, Garden AS, et al. Auto-segmentation of low-risk clinical target volume for head and neck radiation therapy. Practical Radiation Oncology. 2014 Jan;4(1):e31–7.

121. Vellayappan BA, Doody J, Vandervoort E, et al. Pre-operative versus post-operative radiosurgery for brain metastasis: Effects on treatment volume and inter-observer variability. Journal of Radiosurgery and SBRT. 2018;5(2):89–97.

122. Bi N, Wang J, Zhang T, Chen X, et al. Deep Learning Improved Clinical Target Volume Contouring Quality and Efficiency for Postoperative Radiation Therapy in Non-small Cell Lung Cancer. Frontiers in

Oncology. 2019 Nov 13;9:1192.

123. Van Heijst TCF, Van Asselen B, Pijnappel RM, et al. MRI sequences for the detection of individual lymph nodes in regional breast radiotherapy planning. British Journal of Radiology. 2016;89(1063).

124. Chung N-N, Ting L-L, Hsu W-C, et al. Impact of magnetic resonance imaging versus CT on nasopharyngeal carcinoma: primary tumor target delineation for radiotherapy. Head & Neck. 2004 Mar;26(3):241–6.

125. Liao XB, Mao YP, Liu LZ, et al. How Does Magnetic Resonance Imaging Influence Staging According to AJCC Staging System for Nasopharyngeal Carcinoma Compared With Computed Tomography? International Journal of Radiation Oncology Biology Physics. 2008;72(5):1368–77.

126. Wippold FJ. Head and neck imaging: The role of CT and MRI. Journal of Magnetic Resonance Imaging. 2007 Mar;25(3):453–65.

127. Samołyk-Kogaczewska N, Sierko E, Zuzda K, et al. PET/MRI-guided GTV delineation during radiotherapy planning in patients with squamous cell carcinoma of the tongue. Strahlentherapie und Onkologie. 2019 Sep 18;195(9):780–91.

128. Geets X, Daisne J-F, Tomsej M, et al. Impact of the type of imaging modality on target volumes delineation and dose distribution in pharyngo-laryngeal squamous cell carcinoma: comparison between pre- and per-treatment studies. Radiotherapy and Oncology. 2006 Mar;78(3):291–7.

129. Rasch CRN, Steenbakkers RJHM, Fitton I, et al. Decreased 3D observer variation with matched CT-MRI, for target delineation in Nasopharynx cancer. Radiation Oncology. 2010 Dec 15;5(1):21.

130. Bird D, Scarsbrook AF, Sykes J, et al. Multimodality imaging with CT, MR and FDG-PET for radiotherapy target volume delineation in oropharyngeal squamous cell carcinoma. BMC Cancer. 2015 Dec 4;15(1):844.

131. Doshi T, Wilson C, Paterson C, et al. Validation of a Magnetic Resonance Imaging-based Auto-contouring Software Tool for Gross Tumour Delineation in Head and Neck Cancer Radiotherapy Planning. Clinical Oncology. 2017 Jan;29(1):60–7.

132. Lin L, Dou Q, Jin Y-M, et al. Deep Learning for Automated Contouring of Primary Tumor Volumes by MRI for Nasopharyngeal Carcinoma. Radiology. 2019 Jun;291(3):677–86.

133. Wardman K, Prestwich RJD, Gooding MJ, et al. The feasibility of atlas-based automatic segmentation of MRI for H&N radiotherapy planning. Journal of Applied Clinical Medical Physics. 2016 Jul;17(4):146–54.

134. Blanchard P, Garden AS, Gunn GB, et al. Intensity-modulated proton beam therapy (IMPT) versus intensity-modulated photon therapy (IMRT) for patients with oropharynx cancer – A case matched analysis. Radiotherapy and Oncology. 2016;120(1):48–55.

135. Gunn GB, Blanchard P, Garden AS, et al. Clinical Outcomes and Patterns of Disease Recurrence after Intensity Modulated Proton Therapy for Oropharyngeal Squamous Carcinoma. International Journal of Radiation Oncology Biology Physics. 2016;95(1):360–7.

136. Romesser PB, Cahlon O, Scher E, et al. Proton beam radiation therapy results in significantly reduced toxicity compared with intensity-modulated radiation therapy for head and neck tumors that require ipsilateral radiation. Radiotherapy and Oncology. 2016 Feb;118(2):286–92.

137. McDonald MW, Liu Y, Moore MG, et al. Acute toxicity in comprehensive head and neck radiation for nasopharynx and paranasal sinus cancers: cohort comparison of 3D conformal proton therapy and intensity modulated radiation therapy. Radiation Oncology. 2016 Dec 27;11(1):32.

138. Leeman JE, Romesser PB, Zhou Y, et al. Proton therapy for head and neck cancer: expanding the therapeutic window. Vol. 18, The Lancet Oncology. 2017. p. e254–65.

139. Sio TT, Lin H-K, Shi Q, et al. Intensity modulated proton therapy versus intensity modulated photon radiation therapy for oropharyngeal cancer: First comparative results of patient-reported outcomes. International Journal of Radiation Oncology Biology Physics. 2016;95(4).

140. Stromberger C, Cozzi L, Budach V, et al. Unilateral and bilateral neck SIB for head and neck cancer patients. Strahlentherapie und Onkologie. 2016 Apr 6;192(4):232–9.

141. Jakobi A, Bandurska-Luque A, Stützer K, et al. Identification of Patient Benefit from Proton Therapy for Advanced Head and Neck Cancer Patients Based on Individual and Subgroup Normal Tissue Complication Probability Analysis. International Journal of Radiation Oncology Biology Physics. 2015;92(5):1165–74.

142. Remick JS, Schonewolf C, Gabriel P, et al. First Clinical Report of Proton Beam Therapy for Postoperative Radiotherapy for Non–Small-Cell Lung Cancer. Clinical Lung Cancer. 2017 Jul;18(4):364–71.

143. Bagley AF, Ye R, Garden AS, et al. Xerostomia-related quality of life for patients with oropharyngeal carcinoma treated with proton therapy. Radiotherapy and Oncology. 2020 Jan;142:133–9.

144. ICRU. ICRU Report 78: Prescribing, Recording and Reporting Proton-Beam Therapy. In: Journal of the ICRU. 2008. p. 131–4.

145. McDonald MW, Linton OR, Calley CSJ. Dose–Volume Relationships Associated With Temporal Lobe Radiation Necrosis After Skull Base Proton Beam Therapy. International Journal of Radiation Oncology*Biology*Physics. 2015 Feb;91(2):261–7.

146. Alahmari M, Temel Y. Skull base chordoma treated with proton therapy: A systematic review. Surgical Neurology International. 2019 Jun 7;10:96.

147. Hidehiro Hojo, Takeshi Dohmae KH et al. Difference in the relative

biological effectiveness and DNA damage repair processes in response to proton beam therapy according to the positions of the spread out Bragg peak. Radiation oncology (London, England). 2017;12:111.

148. Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. Physics in Medicine and Biology. 2012;57(11):R99–117.

149. Engelsman M, Kooy HM. Target volume dose considerations in proton beam treatment planning for lung tumors. Med Phys. 2005;32(12):3549–57.

150. Moyers MF, Miller DW, Bush DA, et al. Methodologies and tools for proton beam design for lung tumors. In: International Journal of Radiation Oncology Biology Physics. 2001. p. 1429–38.

151. Chen W, Unkelbach J, Trofimov A, et al. Including robustness in multi-criteria optimization for intensity-modulated proton therapy. Physics in Medicine and Biology. 2012;57(3):591–608.

152. Liu W, Zhang X, Li Y, et al. Robust optimization of intensity modulated proton therapy. Medical Physics. 2012;39(2):1079–91.

153. Unkelbach J, Alber M, Bangert M, et al. Robust radiotherapy planning. Physics in Medicine & Biology. 2018 Nov 12;63(22):22TR02.

154. Lomax AJ, Boehringer T, Coray A, et al. Intensity modulated proton therapy: A clinical example. Medical Physics. 2001;28(3):317–24.

155. Unkelbach J, Chan TCY, Bortfeld T. Accounting for range uncertainties in the optimization of intensity modulated proton therapy. Physics in medicine and biology. 2007;52(10):2755–73.

156. Liu W, Frank SJ, Li X, et al. Effectiveness of robust optimization in intensity-modulated proton therapy planning for head and neck cancers. Medical Physics. 2013;40(5):051711.

157. Cubillos-Mesías M, Baumann M, Troost EGC, et al. Impact of robust treatment planning on single- and multi-field optimized plans for proton beam therapy of unilateral head and neck target volumes. Radiation Oncology. 2017 Dec 28;12(1):190.

158. Baskin RM, Boyce BJ, Amdur R, et al. Transoral robotic surgery for oropharyngeal cancer: patient selection and special considerations. Cancer Management and Research. 2018 Apr;Volume 10:839–46.

159. Hutcheson KA, Holsinger FC, Kupferman ME, et al. Functional outcomes after TORS for oropharyngeal cancer: a systematic review. European Archives of Oto-Rhino-Laryngology. 2015 Feb 19;272(2):463–71.

160. Bagley HJ, Short H, Harman NL, et al. A patient and public involvement (PPI) toolkit for meaningful and flexible involvement in clinical trials – a work in progress. Research Involvement and Engagement. 2016 Dec 27;2(1):15.

161. Brizel DM, Sibley GS, Prosnitz LR, et al. Tumor hypoxia adversely

affects the prognosis of carcinoma of the head and neck. International journal of radiation oncology, biology, physics. 1997;38(2):285–9.

162. Cosse J-P, Michiels C. Tumour hypoxia affects the responsiveness of cancer cells to chemotherapy and promotes cancer progression. Anti-Cancer Agents Med Chem. 2008;8(7):790–7.

163. Shannon AM, Bouchier-Hayes DJ, Condron CM, et al. Tumour hypoxia, chemotherapeutic resistance and hypoxia-related therapies. Vol. 29, Cancer Treatment Reviews. 2003. p. 297–307.

164. Overgaard J. Hypoxic modification of radiotherapy in squamous cell carcinoma of the head and neck - A systematic review and meta-analysis. Vol. 100, Radiotherapy and Oncology. 2011. p. 22–32.

165. Nordsmark M, Bentzen SM, Rudat V, et al. Prognostic value of tumor oxygenation in 397 head and neck tumors after primary radiation therapy. An international multi-center study. Radiotherapy and Oncology. 2005 Oct;77(1):18–24.

166. Aguilera K, Brekken R. Hypoxia Studies with Pimonidazole in vivo. BIO-PROTOCOL. 2014;4(19).

167. Peridis S, Pilgrim G, Athanasopoulos I, et al. Carbonic anhydrase-9 expression in head and neck cancer: a meta-analysis. European Archives of Oto-Rhino-Laryngology. 2011 May 19;268(5):661–70.

168. Ren H-Y, Li HY, Xie T, et al. Prognostic role of hypoxia-inducible factor-1 alpha expression in osteosarcoma: a meta-analysis. OncoTargets and Therapy. 2016 Mar;1477.

169. Zhou J, Huang S, Wang L, et al. Clinical and prognostic significance of HIF-1α overexpression in oral squamous cell carcinoma: a meta-analysis. World Journal of Surgical Oncology. 2017 Dec 18;15(1):104.

170. Brooks JM, Menezes AN, Ibrahim M, et al. Development and Validation of a Combined Hypoxia and Immune Prognostic Classifier for Head and Neck Cancer. Clinical Cancer Research. 2019 Sep 1;25(17):5315–28.

171. Toustrup K, Sørensen BS, Metwally MAH, et al. Validation of a 15-gene hypoxia classifier in head and neck cancer for prospective use in clinical trials. Acta Oncologica. 2016 Oct 2;55(9–10):1091–8.

172. Eustace A, Mani N, Span PN, et al. A 26-gene hypoxia signature predicts benefit from hypoxia-modifying therapy in laryngeal cancer but not bladder cancer. Clinical Cancer Research. 2013;19(17):4879–88.

173. Mortensen LS, Johansen J, Kallehauge J, et al. FAZA PET/CT hypoxia imaging in patients with squamous cell carcinoma of the head and neck treated with radiotherapy: Results from the DAHANCA 24 trial. Radiotherapy and Oncology. 2012 Oct;105(1):14–20.

174. Overgaard J. Hypoxic Radiosensitization: Adored and Ignored. Journal of Clinical Oncology. 2007 Sep 10;25(26):4066–74.

175. Janssens GO, Rademakers SE, Terhaard CH, et al. Accelerated Radiotherapy With Carbogen and Nicotinamide for Laryngeal Cancer:

Results of a Phase III Randomized Trial. Journal of Clinical Oncology. 2012 May 20;30(15):1777–83.

176. Ling CC, Humm J, Larson S, et al. Towards multidimensional radiotherapy (MD-CRT): biological imaging and biological conformality. International Journal of Radiation Oncology*Biology*Physics. 2000 Jun;47(3):551–60.

177. Bentzen SM. Theragnostic imaging for radiation oncology: dose-painting by numbers. The Lancet Oncology. 2005 Feb;6(2):112–7.

178. Galvin JM, De Neve W. Intensity Modulating and Other Radiation Therapy Devices for Dose Painting. Journal of Clinical Oncology. 2007 Mar 10;25(8):924–30.

179. Nakano T. Carbon Beam Therapy Overcomes the Radiation Resistance of Uterine Cervical Cancer Originating from Hypoxia. Clinical Cancer Research. 2006 Apr 1;12(7):2185–90.

180. Yoshikawa K, Hasebe M, Ohashi S, et al. Cu-62-ATSM hypoxic imaging of uterine cervical cancer and outcome of carbon ion in radiotherapy. European Journal of Nuclear Medicine and Molecular Imaging. 2011 Oct 24;38(S2):260–441.

181. Elming P, Sørensen B, Oei A, et al. Hyperthermia: The Optimal Treatment to Overcome Radiation Resistant Hypoxia. Cancers. 2019 Jan 9;11(1):60.

182. Overgaard J, Sand Hansen H, Overgaard M, et al. A randomized double-blind phase III study of nimorazole as a hypoxic radiosensitizer of primary radiotherapy in supraglottic larynx and pharynx carcinoma. Results of the Danish Head and Neck Cancer Study (DAHANCA) Protocol 5-85. Radiotherapy and Oncology. 1998 Feb;46(2):135–46.

183. Hassan Metwally MA, Jansen JA, Overgaard J. Study of the Population Pharmacokinetic Characteristics of Nimorazole in Head and Neck Cancer Patients Treated in the DAHANCA-5 Trial. Clinical Oncology. 2015 Mar;27(3):168–75.

184. Bentzen J, Toustrup K, Eriksen JG, et al. Locally advanced head and neck cancer treated with accelerated radiotherapy, the hypoxic modifier nimorazole and weekly cisplatin. Results from the DAHANCA 18 phase II study. Acta Oncologica. 2015 Aug 9;54(7):1001–7.

185. Rischin D, Peters LJ, O'Sullivan B, et al. Tirapazamine, Cisplatin, and Radiation Versus Cisplatin and Radiation for Advanced Squamous Cell Carcinoma of the Head and Neck (TROG 02.02, HeadSTART): A Phase III Trial of the Trans-Tasman Radiation Oncology Group. Journal of Clinical Oncology. 2010 Jun 20;28(18):2989–95.

186. Swartz JE, Pothen AJ, Stegeman I, et al. Clinical implications of hypoxia biomarker expression in head and neck squamous cell carcinoma: a systematic review. Cancer Medicine. 2015 Jul;4(7):1101–16.

187. Toustrup K, Sørensen BS, Lassen P, et al. Gene expression classifier predicts for hypoxic modification of radiotherapy with nimorazole in

squamous cell carcinomas of the head and neck. Radiotherapy and Oncology. 2012 Jan;102(1):122–9.

188. O'Connor JPB, Jackson A, Parker GJM, et al.Dynamic contrast-enhanced MRI in clinical trials of antivascular therapies. Vol. 9, Nature Reviews Clinical Oncology. 2012. p. 167–77.

189. Young IR, Clarke GJ, Baffles DR, et al. Enhancement of relaxation rate with paramagnetic contrast agents in NMR imaging. Journal of Computed Tomography. 1981;5(6):543–7.

190. Berkowitz BA. Role of dissolved plasma oxygen in hyperoxia-induced contrast. Magnetic Resonance Imaging. 1997;15(1):123–6.

191. O'Connor JPB, Naish JH, Parker GJM, et al. Preliminary Study of Oxygen-Enhanced Longitudinal Relaxation in MRI: A Potential Novel Biomarker of Oxygenation Changes in Solid Tumors. International Journal of Radiation Oncology Biology Physics. 2009;75(4):1209–15.

192. O'Connor JPB, Boult JKR, Jamin Y, et al. Oxygen-enhanced MRI accurately identifies, quantifies, and maps tumor hypoxia in preclinical cancer models. Cancer Research. 2016;76(4):787–95.

193. Salem A, Little RA, Featherstone AK, et al. OC-0632: Oxygen enhanced-MRI is feasible, repeatable and detects radiotherapy-induced NSCLC hypoxia changes. Radiotherapy and Oncology. 2018 Apr;127:S336–7.

194. Salem A, Little RA, Latif A, et al. Oxygen-enhanced MRI Is Feasible, Repeatable, and Detects Radiotherapy-induced Change in Hypoxia in Xenograft Models and in Patients with Non–small Cell Lung Cancer. Clinical Cancer Research. 2019 Jul 1;25(13):3818–29.

195. Le Q-T, Kovacs MS, Dorie MJ, et al. Comparison of the comet assay and the oxygen microelectrode for measuring tumor oxygenation in head-and-neck cancer patients. International Journal of Radiation Oncology*Biology*Physics. 2003 Jun;56(2):375–83.

196. Sten Graffman, Peter Björk PE. Polarographic pO 2 Measurements of Intra-abdominal Adenocarcinoma in Connection with Intraoperative Radiotherapy Before and After Change of Oxygen Concentration of Anaesthetic Gases. Acta Oncologica. 2001 Jan 8;40(1):105–7.

197. McKeown SR. Defining normoxia, physoxia and hypoxia in tumours—implications for treatment response. The British Journal of Radiology. 2014 Mar;87(1035):20130676.

198. Simcock R, Simo R. Follow-up and Survivorship in Head and Neck Cancer. Clinical Oncology. 2016 Jul;28(7):451–8.

199. Lee L-Y, Chen S-C, Chen W-C, et al.Postradiation trismus and its impact on quality of life in patients with head and neck cancer. Oral Surgery, Oral Medicine, Oral Pathology & Oral Radiology. 2015;119(2):187.

200. Lee R, Molassiotis A, Rogers SN, et al. Protocol for the trismus trial—therabite versus wooden spatula in the amelioration of trismus in

patients with head and neck cancer: randomised pilot study. BMJ Open. 2018 Mar;8(3):e021938.

201. Henríquez FC, Castrillón SV. A quality index for equivalent uniform dose. Journal of medical physics / Association of Medical Physicists of India. 2011;36(3):126–32.

202. Greenham S, Dean J, Fu CKK, et al. Evaluation of atlas-based auto-segmentation software in prostate cancer patients. Journal of Medical Radiation Sciences. 2014 Sep;61(3):151–8.

203. Rogers SN, Lowe D, Yueh B, et al. The Physical Function and Social-Emotional Function Subscales of the University of Washington Quality of Life Questionnaire. Archives of Otolaryngology–Head & Neck Surgery. 2010;136(4):352.

204. van der Geer SJ, Kamstra JI, Roodenburg JLN, et al. Predictors for trismus in patients receiving radiotherapy. Acta Oncologica. 2016;55(11):1318–23.

205. Loh SY, Mcleod RWJ, Elhassan HA. Trismus following different treatment modalities for head and neck cancer: a systematic review of subjective measures. European archives of oto-rhino-laryngology. 2017;274(7):2695–707.

206. Kieser JA, Herbison GP. Anatomical knowledge and clinical evaluation of the muscles of mastication. Clinical Anatomy. 2000;13(2):94–6.

207. Johnson J, van As-Brooks CJ, Fagerberg-Mohlin B, et al.Trismus in head and neck cancer patients in Sweden: incidence and risk factors. Medical science monitor : international medical journal of experimental and clinical research. 2010;16(6):CR278-82.

208. Dhanrajani P, Jonaidel O. Trismus: aetiology, differential diagnosis and treatment. Dental Update-London-. 2002;29:88–92, 94.

209. Goldstein M, Maxymiw WG, Cummings BJ, et al. The effects of antitumor irradiation on mandibular opening and mobility: a prospective study of 58 patients. Oral surgery, oral medicine, oral pathology, oral radiology, and endodontics. 1999;88(3):365–73.

210. Lee R, Slevin N, Musgrove B, et al. Prediction of post-treatment trismus in head and neck cancer patients. British Journal of Oral and Maxillofacial Surgery. 2012;50:328–32.

211. Buglione M, Cavagnini R, Di Rosario F, et al. Oral toxicity management in head and neck cancer patients treated with chemotherapy and radiation: Dental pathologies and osteoradionecrosis (Part 1) literature review and consensus statement. Critical Reviews in Oncology/Hematology. 2015;97:131–42.

212. Pauli N, Olsson C, Pettersson N, et al. Risk structures for radiation-induced trismus in head and neck cancer. Vol. 55, Acta Oncologica. 2016. p. 788–92.

213. Lee R, Rogers SN, Caress AL, et al. RCT pilot study of Therabite vs wooden spatula in amelioration of trismus in H&N cancer patients.

Radiotherapy and Oncology. 2016;119:S298.

214. Beasley W, Thor M, McWilliam A, et al. Image-Based Data Mining for Identifying Regions Exhibiting a Dose-Response Relationship with Radiation-Induced Trismus. International Journal of Radiation Oncology*Biology*Physics. 2017;99(2):S165.

215. Gomez-Millan J, Fernández JR, Medina Carmona JA. Current status of IMRT in head and neck cancer. Reports of practical oncology and radiotherapy : journal of Greatpoland Cancer Center in Poznan and Polish Society of Radiation Oncology. 2013;18(6):371–5.

216. Anderson CM, Sun W, Buatti JM, et al. Interobserver and intermodality variability in GTV delineation on simulation CT, FDG-PET, and MR Images of Head and Neck Cancer. Jacobs journal of radiation oncology. 2014;1(1):006.

217. Segedin B, Petric P. Uncertainties in target volume delineation in radiotherapy - Are they relevant and what can we do about them? Radiology and Oncology. 2016;50(3):254–62.

218. Vinod SK, Min M, Jameson MG, et al.A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. Vol. 60, Journal of Medical Imaging and Radiation Oncology. 2016. p. 393–406.

219. Choi M, Refaat T, Lester MS, et al. Development of a standardized method for contouring the larynx and its substructures. Radiation Oncology. 2014;9(1).

220. Pirozzi S, Horvat M, Piper J, et al. SU-E-J-106: Atlas-Based Segmentation: Evaluation of a Multi-Atlas Approach for Lung Cancer. In: Medical Physics. 2012. p. 3677.

221. Daisne J-F, Blumhofer A. Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: a clinical validation. Radiation oncology (London, England). 2013;8:154.

222. Dhanrajani P, Jonaidel O. Trismus: aetiology, differential diagnosis and treatment. Dental Update-London-. 2002;29(2):88–92, 94.

223. Zeller JL. High suicide risk found for patients with head and neck cancer. Journal of the American Medical Association. 2006;296:1716–7.

224. Beasley W, Thor M, McWilliam A, et al. Image-based Data Mining to Probe Dosimetric Correlates of Radiation-induced Trismus. International Journal of Radiation Oncology*Biology*Physics. 2018 Nov;102(4):1330–8.

225. De Felice F, Musio D, Tombolini V. Mastication structures definition in head and neck cancer. Vol. 118, Radiotherapy and Oncology. 2016. p. 419.

226. Hague. C, Beasley. W GK et al. Prospective evaluation of relationships between radiotherapy dose to masticatory apparatus and trismus. Acta Oncologica. 2018;1–5.

227. Brouwer CL, Steenbakkers RJHM, van den Heuvel E, et al. 3D Variation in delineation of head and neck organs at risk. Radiation Oncology. 2012;7(1).

228. Zheng Y, Han F, Xiao W, et al. Analysis of late toxicity in nasopharyngeal carcinoma patients treated with intensity modulated radiation therapy. Radiat Oncol. 2015;10(1):17.

229. Wu VWC, Lam Y-N. Radiation-induced temporo-mandibular joint disorder in post-radiotherapy nasopharyngeal carcinoma patients: assessment and treatment. Journal of Medical Radiation Sciences. 2016;63(2):124–32.

230. Beasley WJ, McWilliam A, Aitkenhead A, et al. The suitability of common metrics for assessing parotid and larynx autosegmentation accuracy. Journal of Applied Clinical Medical Physics. 2016 Mar 8;17(2):41–9.

231. Kobayashi K, Hisamatsu K, Suzui N, et al. A Review of HPV-Related Head and Neck Cancer. Journal of Clinical Medicine. 2018 Aug 27;7(9):241.

232. Rasch CRN, Duppen JC, Steenbakkers RJ, et al. Human–Computer Interaction in Radiotherapy Target Volume Delineation: A Prospective, Multi-institutional Comparison of User Input Devices. Journal of Digital Imaging. 2011 Oct 27;24(5):794–803.

233. van Rooij W, Dahele M, Ribeiro Brandao H, et al. Deep Learning-Based Delineation of Head and Neck Organs at Risk: Geometric and Dosimetric Evaluation. International Journal of Radiation Oncology*Biology*Physics. 2019 Jul;104(3):677–84.

234. Cardenas CE, McCarroll RE, Court LE, et al. Deep Learning Algorithm for Auto-Delineation of High-Risk Oropharyngeal Clinical Target Volumes With Built-In Dice Similarity Coefficient Parameter Optimization Function. International Journal of Radiation Oncology*Biology*Physics. 2018 Jun;101(2):468–78.

235. Gooding MJ, Smith AJ, Tariq M, et al. Comparative evaluation of autocontouring in clinical practice: A practical method using the Turing test. Medical Physics. 2018 Nov 12;45(11):5105–15.

236. Pannucci CJ, Wilkins EG. Identifying and Avoiding Bias in Research. Plastic and Reconstructive Surgery. 2010 Aug;126(2):619–25.

237. Whitfield GA, Price P, Price GJ, et al. Automated delineation of radiotherapy volumes: are we going in the right direction? The British Journal of Radiology. 2013 Jan;86(1021):20110718–20110718.

238. Hague C, Beasley W, Dixon L, et al. Use of a novel atlas for muscles of mastication to reduce inter observer variability in head and neck radiotherapy contouring. Radiotherapy and Oncology. 2019 Jan;130:56–61.

239. Skitka LJ, Mosier KL, Burdick M. Does automation bias decision-making? International Journal of Human-Computer Studies. 1999 Nov;51(5):991–1006.

240. Xiangyu E, Hongmei Y, Weigang H, et al. PO-1003 A deep learning based auto-segmentation for GTVs on NPC MR images. Radiotherapy and Oncology. 2019 Apr;133:S553–4.

241. Fritscher KD, Peroni M, Zaffino P, et al. Automatic segmentation of head and neck CT images for radiotherapy treatment planning using multiple atlases, statistical appearance models, and geodesic active contours. Medical Physics. 2014 Apr 24;41(5):051910.

242. Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. Medical Physics. 2017 Feb;44(2):547–57.

243. Mukherji SK. Head and Neck MR Imaging. Neuroimaging Clinics of North America. 2004 Nov;14(4):xi.

244. Thoeny HC, De Keyzer F, King AD. Diffusion-weighted MR Imaging in the Head and Neck. Radiology. 2012 Apr;263(1):19–32.

245. Jansen JFA, Parra C, Lu Y, et al. Evaluation of Head and Neck Tumors with Functional MR Imaging. Magnetic Resonance Imaging Clinics of North America. 2016 Feb;24(1):123–33.

246. Belshaw L, Agnew CE, Irvine DM, et al. Adaptive radiotherapy for head and neck cancer reduces the requirement for rescans during treatment due to spinal cord dose. Radiation Oncology. 2019 Dec 1;14(1):189.

247. Winkel D, Bol GH, Kroon PS, et al. Adaptive radiotherapy: The Elekta Unity MR-linac concept. Clinical and Translational Radiation Oncology. 2019 Sep;18:54–9.

248. Sharp G, Fritscher KD, Pekar V, et al. Vision 20/20: Perspectives on automated image segmentation for radiotherapy. Medical Physics. 2014 Apr 24;41(5):050902.

249. Xu S, Dang H. Deep residual learning enabled metal artifact reduction in CT. In: Chen G-H, Lo JY, Gilat Schmidt T, editors. Medical Imaging 2018: Physics of Medical Imaging. SPIE; 2018. p. 132.

250. Tong N, Gou S, Yang S, et al. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. Medical Physics. 2018 Oct;45(10):4558–67.

251. Henke LE, Contreras JA, Green OL, et al. Magnetic Resonance Image-Guided Radiotherapy (MRIgRT): A 4.5-Year Clinical Experience. Clinical Oncology. 2018 Nov;30(11):720–7.

252. Chen AM, Raghavan G, Cao M, et al. Comparison between CT- and MRI-derived head and neck cancer target volumes using an integrated MRI-tri-60Co teletherapy device. Journal of Radiation Oncology. 2018 Jun 4;7(2):147–55.

253. Močnik D, Ibragimov B, Xing L, et al. Segmentation of parotid glands from registered CT and MR images. Physica Medica. 2018 Aug;52:33–41.

254. Fritscher K, Raudaschl P, Zaffino P, et al. Deep Neural Networks for

Fast Segmentation of 3D Medical Images. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2016. p. 158–65.

255. Kjaer T, Dalton SO, Andersen E, et al. A controlled study of use of patient-reported outcomes to improve assessment of late effects after treatment for head-and-neck cancer. Radiotherapy and Oncology. 2016;119(2):221–8.

256. Schneider U, Pedroni E, Lomax A. The calibration of CT Hounsfield units for radiotherapy treatment planning. Physics in Medicine and Biology. 1996 Jan 1;41(1):111–24.

257. Jäkel O, Reiss P. The influence of metal artefacts on the range of ion beams. Physics in Medicine and Biology. 2007 Feb 7;52(3):635–44.

258. Lomax AJ. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions. Physics in Medicine and Biology. 2008;53(4):1043–56.

259. McGowan SE, Albertini F, Thomas SJ, et al. Defining robustness protocols: a method to include and evaluate robustness in clinical plans. Physics in Medicine and Biology. 2015 Apr 7;60(7):2671–84.

260. Li Y, Niemela P, Liao L, et al. Selective robust optimization: A new intensity-modulated proton therapy optimization strategy. Medical Physics. 2015 Jul 28;42(8):4840–7.

261. Fredriksson A, Forsgren A, Hårdemark B. Minimax optimization for handling range and setup uncertainties in proton therapy. Medical Physics. 2011 Mar 1;38(3):1672–84.

262. Cubillos-Mesías M, Troost EGC, Lohaus F, et al. Including anatomical variations in robust optimization for head and neck proton therapy can reduce the need of adaptation. Radiotherapy and Oncology. 2019 Feb;131:127–34.

263. Yang Z, Zhang X, Wang X, et al. Multiple-CT optimization: An adaptive optimization method to account for anatomical changes in intensity-modulated proton therapy for head and neck cancers. Radiotherapy and Oncology. 2020 Jan;142:124–32.

264. Stützer K, Lin A, Kirk M, et al. Superiority in Robustness of Multifield Optimization Over Single-Field Optimization for Pencil-Beam Proton Therapy for Oropharynx Carcinoma: An Enhanced Robustness Analysis. International Journal of Radiation Oncology*Biology*Physics. 2017 Nov;99(3):738–49.

265. Knutsson H and Andersson M 2005 Morphons: segmentation using elastic canvas and paint on priors IEEE Int. Conf. on Image Proc. 2 pp II-1226-9).

266. Langius JAE, van Dijk AM, Doornaert P, et al. More Than 10% Weight Loss in Head and Neck Cancer Patients During Radiotherapy Is Independently Associated with Deterioration in Quality of Life. Nutrition and Cancer. 2013 Jan;65(1):76–83.

267. McGowan SE, Burnet NG, Lomax AJ. Treatment planning optimisation in proton therapy. The British Journal of Radiology. 2013 Jan;86(1021):20120288–20120288.

268. Sio TT, Lin H-K, Shi Q, et al. Intensity modulated proton therapy versus intensity modulated photon radiation therapy for oropharyngeal cancer: First comparative results of patient-reported outcomes. International Journal of Radiation Oncology Biology Physics. 2016;95(4).

269. Zhang W, Zhang X, Yang P, et al. Intensity-modulated proton therapy and osteoradionecrosis in oropharyngeal cancer. Radiotherapy and Oncology. 2017;123(3):401–5.

270. Nutting CM, Morden JP, Harrington KJ, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): A phase 3 multicentre randomised controlled trial. The Lancet Oncology. 2011;12:127–36.

271. Haviland JS, Agrawal R, Aird E, et al. The UK START (Standardisation of Breast Radiotherapy) Trials: 10-year follow-up results. Cancer Research. 2012;72(24a):2203.

272. Mockford C, Staniszewska S, Griffiths F, et al. The impact of patient and public involvement on UK NHS health care: a systematic review. International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua. 2012;24(1):28–38.

273. Gasson S, Bliss J, Jamal-Hanjani M, et al. The value of patient and public involvement in trial design and development. Clinical Oncology. 2015;27(12):747–9.

274. Hughes-Morley A, Hann M, Fraser C, et al. The impact of advertising patient and public involvement on trial recruitment: embedded cluster randomised recruitment trial. Trials. 2016;17(1):586.

275. Bittner M-I, Grosu A-L. Hypoxia in Head and Neck Tumors: Characteristics and Development during Therapy. Frontiers in oncology. 2013;3(August):223.

276. Janssen HL, Haustermans KM, Balm AJ, et al. Hypoxia in head and neck cancer: How much, how important? Vol. 27, Head and Neck. 2005. p. 622–38.

277. Thorwarth D, Alber M. Implementation of hypoxia imaging into treatment planning and delivery. Radiotherapy and Oncology. 2010 Nov;97(2):172–5.

278. Bassler N, Toftegaard J, Lühr A, et al. LET-painting increases tumour control probability in hypoxic tumours. Acta Oncologica. 2014 Jan 10;53(1):25–32.

279. Servagi-Vernat S, Differding S, Sterpin E, et al. Hypoxia-guided adaptive radiation dose escalation in head and neck carcinoma: A planning study. Acta Oncologica. 2015 Aug 9;54(7):1008–16.

280. Dewhirst MW, Birer SR. Oxygen-Enhanced MRI Is a Major Advance in

Tumor Hypoxia Imaging. Cancer Research. 2016 Feb 15;76(4):769–72.

281. Paver EC, Currie AM, Gupta R, et al. Human papilloma virus related squamous cell carcinomas of the head and neck: diagnosis, clinical implications and detection of HPV. Pathology. 2020 Feb;52(2):179–91.

282. You EL, Henry M, Zeitouni AG. Human papillomavirus–associated oropharyngeal cancer: review of current evidence and management. Current Oncology. 2019 Apr 29;26(2).

283. Kraaijenga SA, Hamming-Vrieze O, Verheijen S, et al. Radiation dose to the masseter and medial pterygoid muscle in relation to trismus after chemoradiotherapy for advanced head and neck cancer. Head & Neck. 2019 May 16;41(5):1387–94.

284. Astradsson T, Laurell G, Ahlberg A, et al. Trismus in patients with head and neck cancer and 5-year overall survival. Acta Octolaryngol. 2018;138 (12):1123–7.

285. Watters AL, Cope S, Keller MN, et al. Prevalence of trismus in patients with head and neck cancer: A systematic review with meta-analysis. Head & Neck. 2019 Sep 19;41(9):3408–21.

286. Karsten RT, Molen L, Hamming-Vrieze O, et al. Long-term swallowing, trismus, and speech outcomes after combined chemoradiotherapy and preventive rehabilitation for head and neck cancer; 10-year plus update. Head & Neck. 2020 Feb 29;hed.26120.

287. Pantvaidya G, Sivasanker M, Ranganathan P, et al. Prospective cross-sectional study assessing prevalence and factors affecting trismus after multimodal treatment for oral cancers. Head & Neck. 2018 Dec 12;hed.25464.

288. Tao CJ, Yi JL, Chen NY, et al. Multi-subject atlas-based auto-segmentation reduces interobserver variation and improves dosimetric parameter consistency for organs at risk in nasopharyngeal carcinoma: A multi-institution clinical study. Radiotherapy and Oncology. 2015;115(3):407–11.

289. Hu Y, Byrne M, Archibald-Heeren B, et al. Implementing user-defined atlas-based auto-segmentation for a large multi-centre organisation: the Australian Experience. Journal of Medical Radiation Sciences. 2019 Dec 28;66(4):238–49.

290. Cui S, Tseng H, Pakela J, et al . Introduction to machine and deep learning for medical physicists. Medical Physics. 2020 Jun 17;47(5).

291. Joskowicz L, Cohen D, Caplan N, et al. Inter-observer variability of manual contour delineation of structures in CT. European Radiology. 2019 Mar 7;29(3):1391–9.

292. Cardenas CE, Mohamed ASR, Tao R, et al. Prospective Qualitative and Quantitative Analysis of Real-Time Peer Review Quality Assurance Rounds Incorporating Direct Physical Examination for Head and Neck Cancer Radiation Therapy. International Journal of Radiation Oncology*Biology*Physics. 2017 Jul;98(3):532–40.

293. Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. Physics in Medicine and Biology. 2012 Jun 7;57(11):R99–117.

294. Unkelbach J, Paganetti H. Robust Proton Treatment Planning: Physical and Biological Optimization. Seminars in Radiation Oncology. 2018 Apr;28(2):88–96.

295. Burnet NG, Mackay RI, Smith E, et al. Proton beam therapy: perspectives on the National Health Service England clinical service and research programme. The British Journal of Radiology. 2020 Mar;93(1107):20190873.

296. Lowe M, Gosling A, Nicholas O, et al. Comparing Proton to Photon Radiotherapy Plans: UK Consensus Guidance for Reporting Under Uncertainty for Clinical Trials. Clinical Oncology. 2020 Jul;32(7):459–66.

297. Price J, Hall E, West C, et al. TORPEdO – A Phase III Trial of Intensity-modulated Proton Beam Therapy Versus Intensity-modulated Radiotherapy for Multi-toxicity Reduction in Oropharyngeal Cancer. Clinical Oncology. 2020 Feb;32(2):84–8.

298. Staley K. 'Is it worth doing?' Measuring the impact of patient and public involvement in research. Research Involvement and Engagement. 2015 Dec 31;1(1):6.

299. Sacristan JA, Aguaron A, Avendaño C, et al. Patient involvement in clinical research: why, when, and how. Patient Preference and Adherence. 2016 Apr;631.

300. Schwartz DL, Garden AS, Shah SJ, et al. Adaptive radiotherapy for head and neck cancer-Dosimetric results from a prospective clinical trial. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2013;106(1):80–4.

301. Surucu M, Shah KK, Roeske JC, et al. Adaptive Radiotherapy for Head and Neck Cancer: Implications for Clinical and Dosimetry Outcomes. Technology in Cancer Research and Treatment. 2017;

302. https://slideshare.net.yber55/oropharynx-cancer-practical-target-delineation.

303. www.humanatatomylibrary.com.

304. https://anatomyqa.com/important-exam-questions-head-and-neck-anatomy.

305. www.mirada-medical.com.

List of publications

The following publications have been published since my MD began in 2016:

Journals

- C. Hague, W. Beasley, K. Garcez, et al. Prospective evaluation of relationships between radiotherapy dose to masticatory apparatus and trismus. *Acta Oncologica volume 57, Issue 8, 2018*

- C. Hague, W. Beasley, L. Dixon, et al. Use of a novel muscles of mastication atlas to reduce inter observer variability in head and neck radiotherapy planning. *Radiotherapy Oncology. 2019 Jan; 130:56-61.*

- C. Hague, MC. Aznar, T. Li, et al. Inter-fraction Robustness of Multi-Field Optimized Intensity-Modulated Proton Therapy in the treatment of post-operative HPV positive oropharyngeal cancers. *British Journal of Radiology (2020) 93,1107*

- C. Hague, B. Foran, E. Hall, et al. Patient Involvement in the Design of a Phase III Trial Comparing Intensity-modulated Proton Therapy and Intensity-modulated Radiotherapy for Oropharyngeal Cancer. Clinical Oncology (R Coll Radiol) 2018; 30(5):274-276

Oral presentations

- C Hague, B Foran, E Hall, et al. Patient Involvement in the Design of a Phase III Trial Comparing Intensity-modulated Proton Therapy and Intensity-modulated Radiotherapy for Oropharyngeal Cancer.

  57th Annual PTCOG meeting, Cincinatti, May, 2018

Traditional and e-posters

- C. Hague, W. Beasley, K. Garcez, et al. Prospective evaluation of relationships between radiotherapy dose to masticatory apparatus and trismus. *ESTRO 37, Barcelona April, 2018*

- C. Hague, B. Foran, E. Hall, et al. Patient Involvement in the Design of a Phase III Trial Comparing Intensity-modulated Proton Therapy and Intensity-modulated Radiotherapy for Oropharyngeal Cancer. *ESTRO 37, Barcelona April, 2018*

- C. Hague, W. Beasley, L. Dixon, et al. Use of a novel muscles of mastication atlas to reduce inter observer variability in head and neck radiotherapy planning. *ASTRO Annual Meeting, San Antonio, 2018*

- C. Hague, MC. Aznar, T. Li, et al. Inter-fraction Robustness of Multi-Field Optimized Intensity-Modulated Proton Therapy in the treatment of post-operative HPV positive oropharyngeal cancers. 58th Annual *PTCOG meeting, Manchester, UK 2019*

- C. Hague, A.McPartlin, L.W.Lee, et al. Evaluation of a combined CT based deep learning auto-contouring model for organ at risk delineation for head and neck radiotherapy. *To be presented at ESTRO 39, 2020*

Appendices

Appendix 1.0 University of Washington Questionnaire (chapter 8)

University of Washington Quality of Life Questionnaire (UW-QOL v4)

This questionnaire asks about your health and quality of life over the past seven days.
Please answer all of the questions by ticking one box for each question.

1. Pain. (Tick one box: ☑)

I have no pain.	(100)
There is mild pain not needing medication.	(75)
I have moderate pain - requires regular medication (e.g. paracetamol).	(50)
I have severe pain controlled only by prescription medicine (e.g. morphine).	(25)
I have severe pain, not controlled by medication.	(0)

2. Appearance. (Tick one box: ☑)

There is no change in my appearance.	(100)
The change in my appearance is minor.	(75)
My appearance bothers me but I remain active.	(50)
I feel significantly disfigured and limit my activities due to my appearance.	(25)
I cannot be with people due to my appearance.	(0)

3. Activity. (Tick one box: ☑)

I am as active as I have ever been.	(100)
There are times when I can't keep up my old pace, but not often.	(75)
I am often tired and have slowed down my activities although I still get out.	(50)
I don't go out because I don't have the strength.	(25)
I am usually in bed or chair and don't leave home.	(0)

4. Recreation. (Tick one box: ☑)

There are no limitations to recreation at home or away from home.	(100)
There are a few things I can't do but I still get out and enjoy life.	(75)
There are many times when I wish I could get out more, but I'm not up to it.	(50)
There are severe limitations to what I can do, mostly I stay at home and watch TV	(25)
I can't do anything enjoyable.	(0)

5. Swallowing. (Tick one box: ☑)

I can swallow as well as ever.	(100)
I cannot swallow certain solid foods.	(70)
I can only swallow liquid food.	(30)
I cannot swallow because it "goes down the wrong way" and chokes me.	(0)

6. Chewing. (Tick one box: ☑)

I can chew as well as ever.	(100)
I can eat soft solids but cannot chew some foods.	(50)
I cannot even chew soft solids.	(0)

7. Speech. (Tick one box: ☑)

My speech is the same as always.	(100)
I have difficulty saying some words but I can be understood over the phone.	(70)
Only my family and friends can understand me.	(30)
I cannot be understood.	(0)

8. Shoulder. (Tick one box: ☑)

I have no problem with my shoulder.	(100)
My shoulder is stiff but it has not affected my activity or strength.	(70)
Pain or weakness in my shoulder has caused me to change my work / hobbies.	(30)
I cannot work or do my hobbies due to problems with my shoulder.	(0)

9. Taste. (Tick one box: ☑)

I can taste food normally.	(100)
I can taste most foods normally.	(70)
I can taste some foods.	(30)
I cannot taste any foods.	(0)

10. Saliva. (Tick one box: ☑)

My saliva is of normal consistency.	(100)
I have less saliva than normal, but it is enough.	(70)
I have too little saliva.	(30)
I have no saliva.	(0)

11. Mood. (Tick one box: ☑)

My mood is excellent and unaffected by my cancer.	(100)
My mood is generally good and only occasionally affected by my cancer.	(75)
I am neither in a good mood nor depressed about my cancer.	(50)
I am somewhat depressed about my cancer.	(25)
I am extremely depressed about my cancer.	(0)

12. Anxiety. (Tick one box: ☑)

I am not anxious about my cancer.	(100)
I am a little anxious about my cancer.	(70)
I am anxious about my cancer.	(30)
I am very anxious about my cancer.	(0)

Which issues have been the most important to you *during the past 7 days?*
Tick ☑ up to 3 boxes.

Pain	Swallowing	Taste
Appearance	Chewing	Saliva
Activity	Speech	Mood
Recreation	Shoulder	Anxiety

GENERAL QUESTIONS

Compared to the month before you developed cancer, how would you rate your health-related quality of life? (Tick one box: ☑)

Much better	(100)
Somewhat better	(75)
About the same	(50)
Somewhat worse	(25)
Much worse	(0)

In general, would you say your **health-related quality of life** during the past 7 days has been: (Tick one box: ☑)

Outstanding	(100)
Very good	(80)
Good	(60)
Fair	(40)
Poor	(20)
Very poor	(0)

Overall quality of life includes not only physical and mental health, but also many other factors, such as family, friends, spirituality, or personal leisure activities that are important to your enjoyment of life. Considering everything in your life that contributes to your personal well-being, rate your **overall quality of life** during the past 7 days. (Tick one box: ☑)

Outstanding	(100)
Very good	(80)
Good	(60)
Fair	(40)
Poor	(20)
Very poor	(0)

Please describe any other issues (medical or nonmedical) that are important to your quality of life and have not been adequately addressed by our questions (you may attach additional sheets if needed).

Appendix 1.1 Questionnaire feedback form (chapter 8)

**Manchester
Biomedical Research Centre**

**NHS
National Institute for
Health Research**

Questionnaire feedback

Please read the questionnaire and use the following questions below to help you discuss it.

1. Is it clear what the questionnaire is focusing on?

Yes	No	Not sure
☐	☐	☐

2. Are the questions worded clearly? Is there any wording that you would change?

Yes	No
☐	☐

3. Do you think the 6 chosen questions are a good measure to compare standard radiotherapy with proton beam therapy?

Yes	No	Not sure
☐	☐	☐

4. Do you feel that all of the areas covered in the questionnaire are relevant?

Yes	No	Not sure
☐	☐	☐

**Public Programmes
People | Research | Dialogue**
at Central Manchester University Hospitals
NHS Foundation Trust

5. Are there any areas that the questionnaire doesn't cover which you think it should? (Please write the areas below)

6. Do you think that the questionnaire is…

Too long	Too short	About right
☐	☐	☐

7. What do you think about the 1-100 scale?

8. How would you prefer to complete the questionnaire?

IPAD	
Paper	
Online	
Post	

9. Any other comments about the questionnaire

Appendix 1.2 Participant information leaflet (chapter 9)

Participant information sheet

Oxygen enhanced MRI measurement in head and neck cancer: validation and efficacy of response

We invite you to take part in a research study for patients with head and neck cancer which will involve up to four extra MRI scans whilst breathing air or oxygen gas. Before you decide whether to take part, it is important for you to understand why the research is being conducted and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for taking the time to read this.

Who will conduct the research?

- The study will be carried out by researchers based at The Christie NHS Foundation Trust and at the Wolfson Molecular Imaging Centre (WMIC) which is part of The University of Manchester.

- The researchers include; Dr Christina Hague, Prof James O'Connor, Dr Andrew McPartlin, Dr David Thomson, Prof Catharine West, Prof Marcel van Herk and Prof Nick Slevin. This work is being supported by the National Institute for Health Research Manchester Biomedical Research Centre.

What is the purpose of the research?

- To improve the ways, we treat cancer we are trying to discover new ways to reliably detect and measure the amount of oxygen in cancer cells. The amount of oxygen can affect the behaviour and survival of cancer cells following treatment. This information may help us to understand why some cancers are more aggressive and are able to survive following treatment. To try and achieve this aim, we will be using a special technique called magnetic resonance imaging (MRI) to measure the amount of oxygen in a tumour. MRI is a non-invasive imaging technique.

Why have I been chosen?

- You have been chosen as we are inviting patients with cancers from the head and neck who are due to start treatment with radiotherapy to participate in this research study.

What would I be asked to do if I took part?

- If you agree to take part, we will ask you to sign a consent form. By signing this consent form, you agree to having the MRI scans with the breathing circuit and injection of a medical dye (contrast medium).

- Normally you do not have MRI scans as part of your treatment but as part of the study you will have up to four extra MRI scans. The first two MRI scans will be performed before your treatment with radiotherapy starts and will be done between 24 hours and 7 days of each other. The third MRI scan will be performed at the end of week 2 and the fourth at the end of week 4 during treatment.

- You will need to have a shell made prior to the scan in the radiotherapy department at the Christie NHS Foundation Trust. All patients undergoing radiotherapy to the head and neck region will require a shell to improve the accuracy of the treatment as NHS standard of care. The shell is a thin piece of plastic that is heated and moulded to the shape of your head and shoulders. The shell will clip into the couch that you will lie flat on and be moved slowly through the imaging tube. We will ask you to wear the shell during the scan.

- During the scan we will ask you to wear a facial mask to breathe air and oxygen gas through similar to that worn in hospital. During the scan the gas will be alternated so measurements of tumour oxygen levels can be taken.

- All female patients of childbearing age will be asked to have a blood pregnancy test prior to treatment as part of NHS standard of care.

- You will have an injection of a medical dye (contrast medium) into your vein via a cannula at each of the 4 scans. The research team will ask your permission to perform some extra tests on your initial biopsy (medical procedure during which a small sample of tissue is removed from a part of your body) which was performed when you were diagnosed. You will not require an extra biopsy. We will be performing additional gene testing on your tissue sample as part of the study. Residual tissue samples will be stored in a metal block store cabinet in a secure location within the MCRC building.

- Following each scan, you will be asked to complete a questionnaire about your experience of having an MRI scan whilst breathing a mixture of air and oxygen.

What are the possible benefits of taking part?

- There are no direct benefits to you, **your treatment will not change.** This work will provide information to help future patients with head and neck cancer.

What are the possible disadvantages and risks of taking part?

- Prior to the scan you will be asked to remove all metal work i.e. jewellery, which will be kept safe.

- The MRI scan will take approximately 60 minutes but may be divided into shorter sections if needed. This scan does not involve x-rays and is very safe. The shell can feel tight but is required to keep you still so the scan is done with the most accuracy. The shell and scan can make people feel claustrophobic. While inside the scanner you will hear a loud drilling noise. This is due to the magnetic field being switched on and off at a high frequency. You will be given earplugs and possibly music in the background to make you feel more relaxed.

- You should not experience any discomfort with the facial mask and both gases will allow you to breathe normally.

- You will also be asked to have an injection of a medical dye through a cannula into a vein. There is a small risk you may develop an allergic reaction to this. This is often mild and rare but there will be medical staff on hand to see you if required. More commonly patients may experience mild nausea or discomfort/ bruising at the cannula site which is often mild and self-limiting.

What will happen to my personal information?

- All data collected about you will be anonymous (all information that could identify you will be removed). The scans will be labelled with a unique subject number to anonymise the data. Other identifiable information will be available to the research team only. Anonymised scan images may be published following completion of the study.

- We are collecting and storing this personal information in accordance with the General Data Protection Regulation (GDPR) and Data Protection Act 2018 which legislate to protect your personal information. The legal basis upon which we are using your personal information is "public interest task" and "for research purposes" if sensitive information is collected. For more information about the way we process your personal information and comply with data protection law please see our Privacy Notice for Research Participants.

- After the study has finished, all the data collected as part of the study will be kept and stored according to research regulations and the University of Manchester and The Christie's standard procedures. All study data will be kept at The Christie for a period of at least 15 years, in case it is needed for audit or inspection. All the data will be kept securely and access will be restricted to authorised personnel. The images collected as part of this study may be used in future research with appropriate approvals. At the end of the 15year period, all research data will be destroyed.

- You have a number of rights under data protection law regarding your personal information. For example, you can request a copy of the information we hold about you. This is known as a Subject Access Request. If you would like to know more about your different rights, please consult our privacy notice for research and if you wish to contact us about your data protection rights, please email dataprotection@manchester.ac.uk or write to The Information Governance Office, Christie Building, University of Manchester, Oxford Road, M13 9PL at the University and we will guide you through the process of exercising your rights.

- You also have a right to complain to the Information Commissioner's Office, Tel 0303 123 1113.

Will my participation in the study be confidential?

- Yes, your participation in this study will be kept confidential to the study team and those with access to your personal information as listed above. Your name will not appear in any publications. Your medical records (which contain personal information) will be reviewed by the clinical care team to assess if you are able to enter the study.

- Once you have joined the study you will be assigned a unique reference number that will be used to label the images and samples we collect, to preserve your anonymity. The research team will be able to link the collected information back to you with this reference number. The research team will also look at the routine scans that you have had during your treatment and any follow-up scans and procedures you have after treatment for a period of 12 months. All study data will be kept secure with restricted access.

- Individuals from the University of Manchester, The Christie or regulatory authorities may need to look at your medical records and the data collected for this study to make sure the project is being carried out as planned. This may involve looking at identifiable data but all individuals involved in auditing and monitoring the study, will have a strict duty of confidentiality to you as a research participant.

What happens if I do not want to take part or if I change my mind?

- It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form.

- If you decide to take part you are still free to withdraw at any time without giving a reason and without detriment to yourself. However, it

will not be possible to remove your data from the project once it has been anonymised and forms part of the dataset as we will not be able to identify your specific data. This does not affect your data protection rights.

Will my data be used for future research?

- When you agree to take part in a research study, the information about your health and care may be provided to researchers running other research studies in this organisation. The future research should not be incompatible with this research project. These organisations may be universities, NHS organisations or companies involved in health and care research in this country or abroad. Your information will only be used by organisations and researchers to conduct research in accordance with the UK Policy Framework for Health and Social Care Research**.**

- This information will not identify you and will not be combined with other information in a way that could identify you. The information will only be used for the purpose of health and care research and cannot be used to contact you regarding any other matter or to affect your care. It will not be used to make decisions about future services available to you.

Will I be paid for participating in the research?

- You will be reimbursed for any additional travel costs and there will be no cost to you. The scans will take place at the Wolfson Molecular Imaging Centre, 27 Palatine Road, Withington, Manchester.

What is the duration of the research?
- Each MRI scan will take up to 1 hour and we will ask you to undergo up to 4 scans. Attempts will be made for the scan to occur on the same day as you attend for your radiotherapy treatment to minimise travel and inconvenience.

- You will be involved in the study for up to 3 months (consent to first follow up post radiotherapy).

Where will the research be conducted?

- The research will be conducted at the Wolfson Molecular Imaging Centre, 27 Palatine Road, Withington, Manchester, M20 3LJ and at The Christie NHS Foundation Trust, Wilmslow Road, Manchester, M20 4BX.

Will the outcomes of the research be published?

- The results of this study will be used by Dr Christina Hague as part of her Research Doctorate or MD thesis.

- We will also communicate the results of this study to the participants through publications. We expect the results to be available 12 months after the last patient has finished participating in the study. Your name will not appear in any of these reports. The research team will ask if you would like us to send you a copy of the publications or a summary of the research when the study has finished. Your name and hospital number will be kept on record securely (on a password protected hospital database) to enable us to check your records for up-to-date contact details in order to contact you. Alternatively, you may contact Dr Christina Hague (christina.hague@christie.nhs.uk) for a copy of the results.

Who has reviewed the study?

- All research in the NHS involving patients is reviewed and approved by an independent group of people called a Research Ethics Committee to protect your safety, rights, wellbeing, and dignity. The local research and development department at The Christie has also given approval. This study has also been independently peer-reviewed.

What if I want to make a complaint?

- Minor Complaints - If you have a minor complaint then you need to contact the researcher(s) in the first instance via the details below who will do their best to answer your questions.

Prof James O'Connor
Phone: 0161 446 3896
Email: James.O'Connor@manchester.ac.uk

- Formal complaints – If you wish to make a formal complaint or if you are not satisfied with the response you have gained from the researchers in the first instance the please contact the Research Governance and Integrity Manager, Research Office, Christie Building, University of Manchester, Oxford Road, Manchester, M13 9PL, by emailing:

  research.complaints@manchester.ac.uk or by telephoning 0161 275 2674 or 275 2046.

- Harm- In the unlikely event that something does go wrong, and you are harmed during the research you may have grounds for a legal

action for compensation against the University of Manchester or the NHS Trust, but you may have to pay your legal costs. The normal National Health Service complaints mechanisms will still be available to you.

What Do I Do Now?

If you have any queries about the study or if you are interested in taking part then please contact your clinical team or the researcher(s) details below

Prof James O'Connor

Phone: 0161 446 3896

Email: James.O'Connor@manchester.ac.uk

Thank you for your participation

Appendix 1.3 Participant Consent Form (chapter 9)

Participant Consent Form
Oxygen enhanced MRI measurement in head and neck cancer: validation and efficacy of response

Participant Identification Number for this trial:_____

	Activities	Initials
1	I confirm that I have read the attached information sheet (Version 3.0, 04/10/2018) for the above study and have had the opportunity to consider the information and ask questions and had these answered satisfactorily.	
2	I understand that my participation in the study is voluntary and that I am free to withdraw at any time without giving a reason and without detriment to myself. I understand that it will not be possible to remove my data from the project once it has been anonymised and forms part of the data set. I agree to take part on this basis	
3	I agree to my GP being informed of my participation in this study.	
4	I understand that relevant sections of my imaging scans and data collected during the study may be looked at by individuals from the University of Manchester, The Christie NHS Foundation Trust and regulatory authorities for auditing and Monitoring purposed. I give permission for these individuals to have access to my records.	
5	I understand that if I withdraw from the study or lose the capacity to give consent to continue on the study or I am withdrawn by the research team, the research team will keep the anonymised data, scan images and tissue samples already collected and use them confidentially in connection with the study.	
6	I agree for tissue samples collected during my surgery to be accessed, stored and further testing performed in the MCRC by the research team.	
7	I agree that any data collected may be published in anonymous form in journals	
8	I agree that the researchers/ researchers at other institutions may contact me in future about other research projects.	
9	I agree that the researchers may retain my contact details in order to provide me with a summary of the findings for this study.	
10	I agree to take part in this study	

Data Protection

The personal information we collect and use to conduct this research will be processed in accordance with data protection law as explained in the Participant Information Sheet and the [Privacy Notice for Research Participants](#).

_____   _____

Name of Participant   Signature   Date

_____   _____

Name of the person taking consent Signature   Date

When completed, original form to be kept in Investigator Site file, one copy to be kept in the medical notes and one copy to be given to the participant

Thank you very much for agreeing to participate in this research

Appendix 2 Published Papers

C. Hague, W. Beasley, K. Garcez, et al. Prospective evaluation of relationships between radiotherapy dose to masticatory apparatus and trismus. *Acta Oncologica volume 57, Issue 8, 2018*

C. Hague, W. Beasley, L. Dixon, et al. Use of a novel muscles of mastication atlas to reduce inter observer variability in head and neck radiotherapy planning. *Radiotherapy Oncology. 2019 Jan; 130:56-61.*

C. Hague, MC. Aznar, T. Li, et al. Inter-fraction Robustness of Multi-Field Optimized Intensity-Modulated Proton Therapy in the treatment of post-operative HPV positive oropharyngeal cancers. *British Journal of Radiology (2020) 93,1107*

C. Hague, B. Foran, E. Hall, et al. Patient Involvement in the Design of a Phase III Trial Comparing Intensity-modulated Proton Therapy and Intensity-modulated Radiotherapy for Oropharyngeal Cancer. Clinical Oncology (R Coll Radiol) 2018; 30(5):274-276

Taylor & Francis
Taylor & Francis Group

Check for updates

ORIGINAL ARTICLE

Prospective evaluation of relationships between radiotherapy dose to masticatory apparatus and trismus

Christina Hague[a], William Beasley[b], Kate Garcez[a], Lip W. Lee[a], Andrew McPartlin[a], Alan McWilliam[b,c], David Ryder[d], Andrew J. Sykes[a], David Thomson[a], Marcel van Herk[b,c], Catharine West[e] and Nick J. Slevin[a]

[a]Department of Head and Neck Clinical Oncology, The Christie NHS Foundation Trust, Manchester, UK; [b]Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Heath, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK; [c]Department of Radiotherapy Related Research, The Christie NHS Foundation Trust, Manchester, UK; [d]Statistics Unit, The Christie NHS Foundation Trust, Manchester, UK; [e]Translational Radiobiology Group, Division of Cancer Sciences, Manchester Academic Health Science Centre, University of Manchester, The Christie NHS Foundation Trust, Manchester, UK

ABSTRACT

Aims: This feasibility study aimed to identify relationships between radiation doses to the masticatory apparatus as a combined block or as individual subunits with changes in trismus following radiotherapy.

Material and methods: Twenty patients from a single center were recruited prospectively as part of a randomized trial comparing proactive exercises in the management of trismus. Patients with stage III/IV oral cavity or oropharyngeal squamous cell cancers received intensity-modulated radiotherapy with concurrent systemic therapy. All patients had trismus prior to radiotherapy. Maximal inter-incisor distance (MID) was measured pre- and 6 months from the start of radiotherapy. Bilateral muscles of mastication: medial and lateral pterygoids (MP and LP), masseters (M), temporalis (T), temporomandibular joint (TMJ) were contoured on CT images. The block comprised all muscles excluding the TMJ below the orbital floor. Mean dose, equivalent uniform dose (EUD) and V35–V60 Gy were compared with change in MID.

Results: In six patients, the MID deteriorated at 6 months from the start of radiotherapy compared with 14 whose MID improved. No significant association was observed between age, gender, smoking, alcohol status, exercise compliance, cisplatin, tumor site, stage, V35–V60 Gy or EUD with change in MID. A clinical outlier was excluded. Without the outlier ($n = 19$), a significant association was seen between mean dose and change in MID at 6 months for the ipsilateral block ($p = .01$), LP ($p = .04$) and M ($p < .01$). All patients where trismus deteriorated at 6 months received mean doses >40 Gy to the block.

Conclusion: Higher mean radiation doses to the ipsilateral block, LP and M were significantly associated with deterioration in trismus. Limiting dose to these structures to ≤ 40 Gy for tumors not invading the masticatory muscles may improve treatment-related sequelae. The ipsilateral block, LP and M should be studied further as possible alternative avoidance structures in radiotherapy treatment planning.

Introduction

Trismus or 'locked jaw' is defined as an 'inability to fully open the mouth'. It is a common treatment-related effect in patients with head and neck cancer resulting in significant morbidity. The prevalence of trismus varies considerably. This variability is due to patient and physician under-reporting, study differences in clinicopathologic and treatment factors, and the lack of a universally agreed definition [1–3].

Trismus is caused by impaired function of the muscles of mastication secondary to benign or malignant processes and their associated treatment. The muscles include: medial (MP) and lateral pterygoids (LP), masseter (M), temporalis (T) as well as the temporo-mandibular joint (TMJ). The paired muscles of mastication assist with chewing through their attachment onto the mandible. The MP, M and T close the jaw whilst the LP opens the jaw. Injury to these muscles will result in reduced function and range of mandibular motion leading to reduced nutrition, impaired oral hygiene and difficulty speaking [4]. The most widely used definition for trismus is a maximal inter-incisor distance (MID) of ≤ 35 mm [5].

There are a number of patient, tumor and treatment-related factors which contribute to trismus [1,6]. The radiation-induced pathogenesis involves fibrosis and atrophy of the mastication muscles secondary to ischemia [7]. The effect of radiation is not immediate but one that progresses over months to years following treatment. The severity of radiation-induced trismus appears to be related to the total dose received and the volume of tissues within the radiation field [8]. Significant dose–response relationships have been found

but the literature varies considerably on what dose constraints should be used. There is also no consensus as to which of the mastication muscles should be defined as an organ at risk with a view to avoidance planning [6,9,10]. Rao and van der Molen [11,12] reported that doses to the pterygoids and masseter muscles were the most robust predictors for the development of trismus whilst other studies showed doses to the TMJ and the pterygoids were important [11–13]. A limitation of published studies is the use of retrospectively collected data and mouth opening measurements based on dichotomized data [14,15].

Given these limitations a feasibility study was carried out using prospectively collected data to identify and compare different dose parameters to a combination of these muscles defined as a block and as individual mastication muscles. Mouth opening measurements on a continuous scale were used to increase statistical power.

Material and methods

Patients

The study population comprised patients recruited in Manchester into the multi-center phase 3 randomized controlled Trismus trial [16]. The trial compared proactive exercises using therabite (platon medical) versus standard wooden spatulas in patients with stage III/IV oral cavity/oropharyngeal cancers to improve trismus. Patients randomized to either intervention were asked to follow a protocol of exercises beginning 2–3 weeks prior to starting radiotherapy. These included five sessions per day of five openings/closing per session with a 30 s stretch for each opening [17]. Patients had stage III/IV squamous cell carcinomas of the oral cavity/oropharynx and were treated with concurrent radiotherapy and systemic therapy to the primary tumor from February 2013 to January 2015. Of the 22 patients enrolled in Manchester, two patients were excluded due to undergoing primary surgery with mandibulectomy. The remaining 20 underwent ipsilateral, bilateral neck dissections or hemi-glossectomies as shown in Supplementary Table S1.

Patients received 60–66 Gy in 30 fractions (2.0–2.2 Gy per fraction) over 6 weeks. A rotational-based intensity-modulated radiotherapy treatment (IMRT) plan was calculated using the pinnacle treatment planning system (Pinnacle version 9.6; Philips Radiation Oncology Systems, Andover, MA) with target delineation performed on axial CT images. No dose constraints were applied to the muscles of mastication. All patients had concurrent chemotherapy with cisplatin or cetuximab if cisplatin was contraindicated.

Evaluation of trismus and dose

Patients enrolled in the trial had trismus defined as a sense of jaw tightening self-reported by each patient prior to radiotherapy. For this study, MID measurements were taken prospectively at baseline and 6 months from the start of radiotherapy. The averages of two MID measurements were taken at baseline and at 6 months. The change in MID for each patient was calculated.

The block was defined as the MP, LP, M and T muscles below the orbital floor to exclude the cranial component of the T. Mean dose, equivalent uniform dose (EUD) and V35–V60 Gy were calculated for the contoured volumes of the block and each individual muscle on the ipsilateral and contralateral side. The EUD is a method of reporting radiotherapy dose distributions taking account of non-linearity of tissue dose–response whilst not attempting to make predictions of absolute outcome [18].

An in-house contouring atlas was designed for the masticatory apparatus to aid contouring. Absorbed doses were recalculated for the block and masticatory apparatus using dose volume histograms. Data were extracted via a script to calculate dose volume histograms using in-house software.

Statistics

Analysis was performed using graph pad prism version 6 (Graph pad software), SPSS version 23 (IBM SPSS statistics) and Microsoft Office excel version 2010 . Data analysis compared patient, tumor and treatment-related factors with percent change in MID at 6 months from baseline. Non-parametric tests of Fisher's exact, chi-squared and chi-squared test for trend were used. Correlations between the different muscles and dose parameters were calculated using Spearman's rank and linear regression models. A p value of $<.05$ was considered statistically significant.

Results

The 20 patients reviewed all had established trismus at baseline. In six, the change in MID deteriorated from a median baseline of 29 mm (17–34) pre-radiotherapy to 18 mm (6–29) 6 months from the first radiation treatment. In 14 patients, the MID improved from a median baseline of 16 mm (8–34) to 31 mm (9–39) at 6 months. At the time of the analysis no patients had recurred with residual disease. Supplementary Table S1 shows the distribution of patients in the deterioration and improvement groups in relation to a number of parameters. Patients with a deterioration in MID had more involved lymph nodes ($p = .04$). However, the distribution of exercise frequency did not differ between the deterioration and improvement group ($p = 1.00$), and there were no other statistically significant differences in relation to patient, tumor and treatment-related factors (Supplementary Table S1). Although not statistically significant, mean doses received by the block and individual ipsilateral muscles were higher in the deterioration group with the exception of the temporalis muscle (Supplementary Table S1). A clinical outlier was detected within the improvement group. Further analyses were performed with and without the outlier.

Negative correlations were observed between mean doses to the block and individual ipsilateral muscles with changes in MID. The correlation improved once the outlier was excluded across the masticatory apparatus (Table 1). The size of this effect increased with the outlier excluded and reached statistical significance for the ipsilateral block ($\beta = 6.26$, $p < .01$), MP ($\beta = 6.61$, $p < .01$), LP ($\beta = 4.15$, $p < .01$) and M

($\beta = 4.58$, $p = .02$) muscles. Figure 1 illustrates a change of 4–7% in mouth opening per gray of dose to the relevant structures.

Discussion

This feasibility study using prospective data showed an association between mean dose to the ipsilateral block and individual masticatory muscles with the development of trismus. All patients had established trismus prior to radiotherapy measured on a continuous scale. The change in MID

Table 1. Correlations between the mean radiation dose to the muscles of mastication with change in maximal inter-incisor distance.

	With outlier ($n = 20$)		Without outlier ($n = 19$)	
	r^s value[a]	p value	r^s value	p value
Ipsilateral block	−0.42	.07	−0.58	.01
Ipsilateral MP	−0.27	.25	−0.42	.09
Ipsilateral LP	−0.32	.16	−0.41	.04
Ipsilateral M	−0.44	.05	−0.61	<.01
Ipsilateral temporalis	−0.23	.65	−0.52	.39

[a]Calculated using Spearman rank correlation.
MP: medial pterygoids; LP: lateral pterygoids; M: masseters.

6 months from the start of radiotherapy is a similar metric to that used by Joyce van der Geer et al. [2] who observed a peak in the prevalence of trismus at 6 months.

In our study, there were no statistical significant associations observed between changes in trismus and age, gender, tumor stage or site, concurrent cisplatin or frequency of proactive exercises during and after treatment. Even though not statistically significant ($p = 1.00$), a relationship was observed between the use of proactive exercises during and after treatment, in those whom the MID improved the frequency of daily exercises performed was higher. Mean doses to the ipsilateral LP ($r^s = -0.41$, $p = .04$) and M ($r^s = -0.61$, $p < .01$) were significantly associated with change in MID. There was also a correlation between mean dose to the ipsilateral MP but this did not reach statistical significance ($r^s = -0.42$, $p = .09$). The relationship between trismus and mean dose to the ipsilateral MP is similar to published data by Kent et al. [13]. In a study of 40 patients, Kent et al. [13] identified a 45% prevalence in trismus with mean doses to the pterygoids of ≥ 55 Gy [13]. In our study, mean dose to the ipsilateral M was most correlated with trismus. This is similar to other retrospective studies. In a study of 139 patients, a significant relationship between mean doses to the ipsilateral
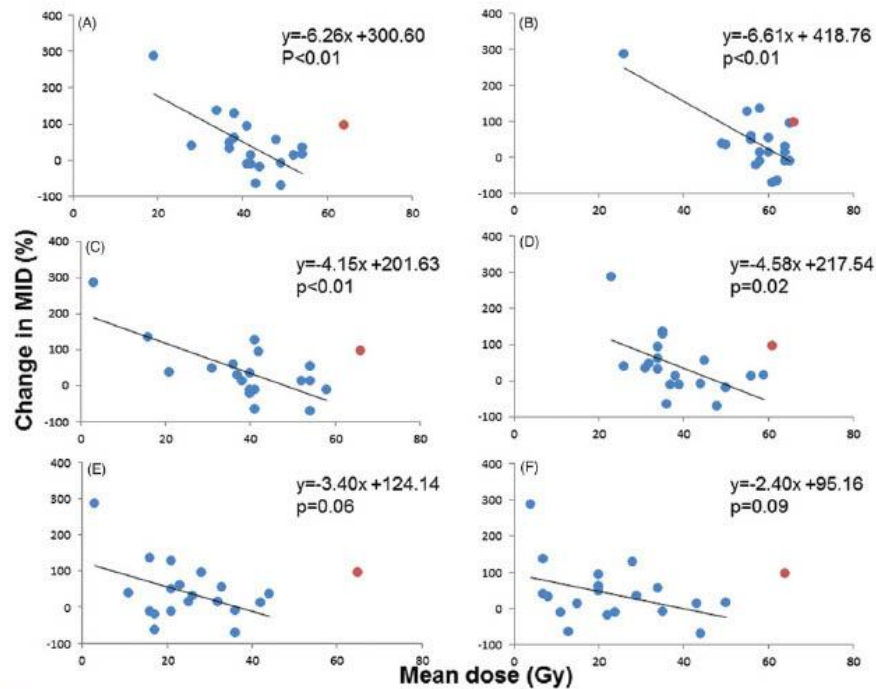

Figure 1. Scatter plots showing the relationships between the percentage change in maximal inter-incisor distance (MID) 6 months following the start of radiotherapy and mean doses to the muscles of mastication. The clinical outlier is shown by the red point. A (ipsilateral block), B (ipsilateral medial pterygoid), C (ipsilateral lateral pterygoid), D (ipsilateral masseter), E (ipsilateral temporalis) and F (ipsilateral TMJ). Data is given for 19 patients.

masseter of 60 Gy was found [19] contrasting with a study of 55 patients where doses of V20–V40 Gy to the masseter were associated with the development of trismus [12].

In one patient referred to as a clinical outlier, the MID improved at 6 months despite mean doses in excess of 60 Gy received by all masticatory muscles. This outlier was the only HPV positive, stage T4b tumor in the cohort. The large primary base of tongue tumor had invaded into the pterygoid muscles resulting in trismus at diagnosis. The contoured GTV was much larger at 106 cm^3 than the rest of the cohort (median 15 cm^3, range 3–106 cm^3). This patient had very few risk factors for developing late effects being a nonsmoker, absence of concurrent platinum-based chemotherapy and no surgery. Post-treatment scans confirmed a complete response with no current evidence of recurrence 12 months following treatment. The improvement in trismus can be explained by tumor response and improved pterygoid muscle function.

Although there are known differences in muscle architecture and function between the pterygoid muscles, an association was found between mean dose to the MP and LP with trismus. This can be explained by the close proximity of the pterygoid muscles and synergistic function to open and close the jaw [20]. The mean dose received by the ipsilateral LP was lower than the MP in those where by trismus improved (39 Gy versus 57 Gy). This is in keeping with Hsieh et al. [21] who suggested a mean dose constraint of <42 Gy to the LP to reduce the risk of trismus. The mean doses to the ipsilateral MP, LP, M and TMJ excluding the clinical outlier were greater in those where trismus deteriorated at 6 months from the start of radiotherapy compared to those where trismus improved. There was no significant correlation observed between V35 and V60 Gy with changes in trismus across all paired muscles and the block. Our observation contrasts with another study in 124 patients where doses of V40–V60 Gy to the ipsilateral masseter were associated with statistically significant changes in trismus [14]. The difference in the dose–response relationships between the two studies may be attributed to the small sample size and short follow-up period of 6 months in our study compared with a median follow-up of 66 months in the study by Lindblom et al. [14] In our study there was a weaker and non-significant association seen between the EUD and change in MID compared with the mean dose. The EUD was calculated to evaluate potential non-linear dose effects within a muscle, but the results are less strongly correlated with changes in trismus than the mean dose.

To our knowledge, only the effect of dose on the masticatory muscles as individual subunits has been explored previously. An avoidance dose for the masticatory apparatus has not been agreed due to disparity within the literature. Our study is the first to describe and evaluate the concept of a block as an alternative OAR. The block includes the MP, LP, M and T below the floor of the orbit to exclude the cranial component of the T. These muscles were chosen due to their similarities in function and anatomical locations. There was a modest significant correlation ($r^2 = -0.58$, $p = .01$) between mean dose to the ipsilateral block and change in MID. Patients with a deterioration in trismus 6 months from the start of radiotherapy received mean doses >40 Gy to the ipsilateral block, MP, LP and M. Significant correlations were seen between mean dose and changes in MID with the ipsilateral block, LP and M. Limiting mean dose ≤40 Gy to the ipsilateral block, LP and M for tumors not invading masticatory muscles in those with established trismus prior to radiotherapy and whom underwent proactive exercises could be considered. Mean dose ≤40 Gy is considerably lower than that reported for each individual muscle both in our study and in the published literature. The lower mean dose constraint for the ipsilateral block can partly be explained by the larger volume of muscle within the block compared to the muscles as separate entities. The concept of grouping muscles together as a block would help remove uncertainties as to which masticatory muscles should be avoided. It would also improve our understanding of the clinical significance of radiotherapy to the regions in the muscle interface such as fat, fascia or nerves of which little is known. Image based data mining has illustrated this point through highlighting an area adjacent to the masseter that has a dose–response relationship [22]. Avoidance structures such as the block may help reduce the severity of radiation-induced trismus but require validation in larger studies.

Limitations to the study

Although this is the first article to our knowledge that uses prospective pre- and post-radiotherapy MID measurements on a continuous scale to evaluate dose effects on organs at risk, the study is small. All patients had subjective jaw tightening prior to the start of radiotherapy and underwent proactive exercises during and after treatment. The results are hypothesis generating and require validation in a larger study. Our findings need to be verified in a larger sample and matched control cohort with longer follow-up. A larger study would enable development of a normal tissue complications probability (NTCP) model and recommendations for mean dose constraints to use in radiotherapy planning. However, our work does show that further studies using MID methodology would be useful.

Conclusions

In this prospective study, higher mean radiation doses to the ipsilateral block, LP and M were significantly associated with trismus. The findings suggest that limiting mean dose to the ipsilateral block, LP and M to ≤40 Gy for tumors not invading the masticatory muscles may reduce treatment-related morbidity. The ipsilateral block, LP and M should be studied further as alternative OARs and possible avoidance structures in radiotherapy planning in future studies. This suggestion requires validation in a larger study with longer follow-up.

Disclosure statement

No potential conflict of interest was reported by the authors.
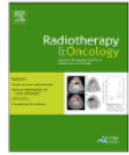
Original article

Use of a novel atlas for muscles of mastication to reduce inter observer variability in head and neck radiotherapy contouring

Christina Hague [a,*], William Beasley [b], Lynne Dixon [a], Simona Gaito [a], Kate Garcez [a], Andrew Green [b,c], Lip W. Lee [a], Massimo Maranzano [d], Andrew McPartlin [a], Hitesh Mistry [b], Damian Mullan [e], Andrew J. Sykes [a], David Thomson [a], Marcel Van Herk [b,c], Catharine M. West [f], Nick Slevin [a]

[a] Department of Head and Neck Clinical Oncology, The Christie NHS Foundation Trust, Manchester; [b] Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Heath, University of Manchester, Manchester Academic Health Science Centre; [c] Department of Radiotherapy Related Research, The Christie NHS Foundation Trust, Manchester; [d] Department of Oral-Maxillo-Facial and Plastic Reconstructive Surgery, Central Manchester University Hospitals; [e] Department of Radiology, The Christie NHS Foundation Trust, Manchester; and [f] Translational Radiobiology Group, Division of Cancer Sciences, Manchester Academic Health Science Centre, University of Manchester, The Christie NHS Foundation Trust, UK

ABSTRACT

Purpose/objective(s): Trismus is caused by injury to the masticatory muscles resulting from cancer or its treatment. Contouring these muscles to reduce dose and radiation related trismus can be problematic due to interobserver variability. This study aimed to evaluate the reduction in interobserver variability achievable with a new contouring atlas.

Materials/methods: The atlas included: medial and lateral pterygoids (MP, LP), masseter (M) and temporalis (T) muscles, and the temporo-mandibular joint (TMJ). Seven clinicians delineated five paired structures on CT scans from 5 patients without the atlas. After ≥5 weeks, contouring was repeated using the atlas. Using contours generated by the clinicians on the same 5 CT scans as reference, dice similarity coefficient (DSC), mean distance-to-agreement (DTA) and centre of mass (COM) difference were compared with and without the atlas. Comparison was also performed split by training grade. Mean and standard deviation (SD) values were measured.

Results: The atlas reduced interobserver variability for all structures. Mean DTA significantly improved for MP ($p = 0.01$), M ($p < 0.01$), T ($p < 0.01$) and TMJ ($p < 0.01$). Mean DTA improved using the atlas for the trainees across all muscles, with the largest reduction in variability observed for the T (4.3 ± 7.1 v 1.2 ± 0.4 mm, $p = 0.06$) and TMJ (2.1 ± 0.7 v 0.8 ± 0.3 mm, $p < 0.01$). Distance between the COM and interobserver variability reduced in all directions for MP and T.

Conclusion: A new atlas for contouring masticatory muscles during radiotherapy planning for head and neck cancer reduces interobserver variability and could be used as an educational tool.

Radiotherapy to the head and neck is challenging due to complex anatomy and large number of organs at risk (OARs). Current radiotherapy techniques such as Intensity Modulated Radiotherapy (IMRT) increase dose conformity allowing improved loco-regional tumour control as well as reduced normal tissue effects [1,2]. To fully exploit the advantages of IMRT, accurate and consistent target delineation is required. Manual target volume and OAR delineation are affected by clinician variability [3]. Minimising interobserver variability will improve the accuracy of the dose delivered, maximise tumour control, limit toxicities and increase knowledge of organ at risk (OAR) dose [4–6]. Methods to standardise OAR volumes have been developed including superior imaging techniques, peer review and the development of contouring atlases [7].

Contouring atlases can help standardise volumes, reduce interobserver variability and improve normal tissue sparing in daily clinical practice [8–11]. Atlases agreed by an expert panel may reduce inconsistencies between radiotherapy centres and facilitate multi-institutional clinical trials [12]. There are a number of atlases for head and neck cancer: the current Danish Head and Neck

* Corresponding author.
E-mail addresses: Christina.hague@christie.nhs.uk (C. Hague), William.beasley@christie.nhs.uk (W. Beasley), lynne.dixon3@nhs.net (L. Dixon), simona.gaito@christie.nhs.uk (S. Gaito), Kate.garcez@christie.nhs.uk (K. Garcez), Andrew.green-2@manchester.ac.uk (A. Green), lipwai.lee@christie.nhs.uk (LW. Lee), massimo.maranzano@postgrad.manchester.ac.uk (M. Maranzano), Andrew.mcpartlin@christie.nhs.uk (A. McPartlin), hitesh.mistry@manchester.ac.uk (H. Mistry), damian.mullan@christie.nhs.uk (D. Mullan), Andrew.sykes@christie.nhs.uk (A.J. Sykes), david.thomson@christie.nhs.uk (D. Thomson), marcel.vanherk@manchester.ac.uk (M. Van Herk), Catharine.west@manchester.ac.uk (C.M. West), nick.slevin@christie.nhs.uk (N. Slevin).

Cancer Group (DAHANCA), European Organization for Research and Treatment of Cancer (EORTC) and Radiation Therapy Oncology Group (RTOG) [12]. One limitation of current published atlases is the absence of delineation guidelines for the masticatory muscles.

Trismus is defined as a maximum inter-incisor distance of ≤35 mm and is caused by impaired function of the masticatory muscles [13]. Trismus can manifest in poor dental hygiene, impaired chewing, malnutrition and psychological difficulties including low self-esteem, depression and suicidal intentions, which all reduce health-related quality-of-life [14,15]. Clinical assessment of patients is challenging due to a restricted ability to assess disease status. There are several patient, tumour and treatment related factors for trismus of which radiotherapy is a known contributor with an incidence in advanced oropharyngeal cancers of 35–55% [16–20]. Mouth opening is a complex action controlled by the synergistic actions of the paired muscles of mastication. These include: medial and lateral pterygoids (MP, LP), masseter (M), temporalis (T) as well as the temporo-mandibular joint (TMJ) [21]. The origin, insertion and function of each of the muscles of mastication are summarised in Supplementary Tables 1a and 1b. Identifying the masticatory apparatus as an OAR with a view to avoidance planning will aim to reduce toxicities and improve quality-of-life. Dosimetric studies showed a relationship between the severity of trismus with dose and volume of muscle treated [22]. Despite this there is no standardised defined OAR or dose threshold for the masticatory muscles for radiotherapy planning [23]. Within the current literature, inconsistences exist regarding proposed dose parameters as summarised in Supplementary Table 2. For example, the largest most recent study by Rao et al of 421 patients suggested limiting the high dose volume of the ipsi-lateral MP to V_{68} Gy < 10 cm^3 [16].

Few studies have evaluated the use of delineation guidelines to improve interobserver variability in contouring OARs. A paper by Brouwer et al. of 6 head and neck clinicians showed poor compliance with delineation guidelines for the spinal cord, parotid and submandibular glands was associated with an increase in interobserver variability [24]. Currently there are no standardised delineation guidelines for contouring the masticatory muscles [12]. Accurate delineation of these muscles with a validated atlas is a prerequisite for high quality radiotherapy planning to improve consistency, standardise contours and reduce radiation related trismus. This study aimed to evaluate a novel muscles of mastication atlas to aid clinician contouring, reduce interobserver variability, support training and the development of multi-institutional clinical trials.

Methods

A muscles of mastication atlas was developed by a multi-disciplinary expert team consisting of a consultant radiologist, maxillo-facial surgeon and clinical oncologists. Using the Pinnacle (Pinnacle version 9.6, Philips Radiation Oncology Systems, Andover, MA) treatment planning system, the muscles of mastication (MP, LP, M, T) and TMJ were contoured on computed tomography (CT) slices. All muscles were delineated using the soft tissue window with the exception of the TMJ which was contoured on a bone window. The contours were extracted as DICOM files and converted into an app using in-house software. The atlas app is shown in Fig. 1, with a link attached www.bit.ly/trismusatlas (to access the webpage please open with google chrome version 56 or opera version 43). Included in the atlas app is a table explaining the anatomical boundaries of each component of the muscles of mastication.

Seven head and neck clinicians (five consultants and two trainees not included in the multi-disciplinary team) delineated the paired muscles of mastication on randomly selected CT scans from five patients without the atlas. After a minimum gap of five weeks each clinician was given the atlas and re-contoured the same structures on the same five CT scans.

Contours were created for each patient by the same multi-disciplinary team that produced the atlas and used as the reference. In-house software was used to compare clinician-drawn structures with the reference. Dice similarity coefficient (DSC), mean distance to agreement (DTA) and the centre of mass
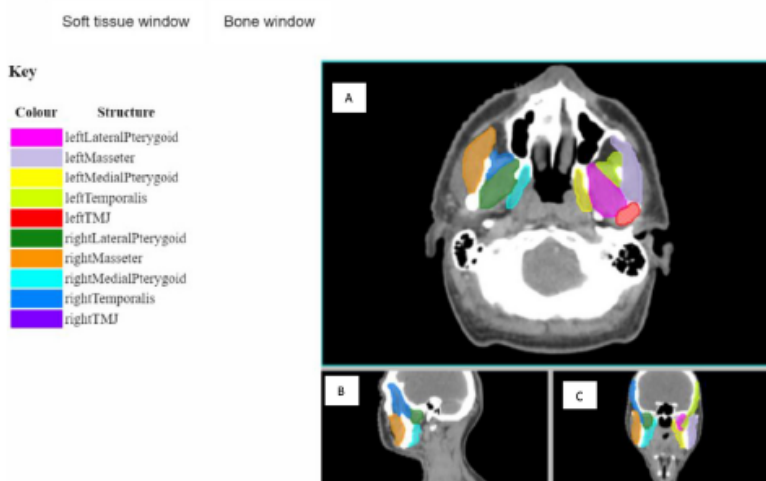

Fig. 1. Overview of the masticatory muscles atlas showing axial (A), sagittal (B) and coronal (C) slices. The right TMJ is not visible on the CT slice shown in Figure. www.bit.ly/trismusatlas (to access the webpage please open with google chrome version 56 or opera version 43).

225

Table 1
Comparison of the contoured volumes of the muscles of mastication without and with the atlas.

	Volume Mean ± SD (cm³)		*P* value
	No atlas	Atlas	
Lateral pterygoids	6.7 ± 1.5	6.8 ± 1.2	0.70
Medial pterygoids	7.7 ± 1.9	8.5 ± 1.3	0.02
Masseters	19.4 ± 2.1	20.1 ± 1.7	0.04
Temporalis*	20.3 ± 6.6	28.0 ± 4.3	<0.01
TMJs	1.7 ± 1.1	1.3 ± 0.3	<0.01

Abbreviations: SD = standard deviation; TMJ = temporo-mandibular joint.

difference (COM) were evaluated and compared to the reference for each volume without and with the atlas. DTA was calculated by measuring the distance from each point on the reference surface to the closest point on the clinician-drawn surface and combining into a DTA histogram. The mean DTA was then calculated from this DTA histogram. The mean and standard deviation (SD) across all patients were compared without and with the atlas to assess inter-observer variability. Comparison was also performed split by training grade. Standard deviation maps were produced to illustrate the variation in structure delineation between clinicians at each voxel without and with the atlas. The Standard deviation

(SD) per voxel is calculated in 3D by combining all clinician contours. All voxels inside a contour are given a value of 1 and all outside the contour are given a value of 0. The standard deviation of each voxel is calculated to illustrate regions in which there is little agreement between clinicians: the larger the standard deviation the larger the disagreement.

Statistical analysis

Analysis was performed using Graph pad prism version 6 (Graph pad software) and Microsoft Office Excel 2010. A paired t-test was used to compare mean DSC, mean DTA and distance to COM without and with the atlas. Statistical significance was defined as $p \leq 0.05$.

Results

The median (range) time between contouring without and with the atlas across all seven clinicians was 66 (35–145) days. Using the atlas there was an increase in the mean delineated volumes for all muscles excluding the TMJ (Table 1). The SD of the contoured volumes significantly reduced using the atlas for the MP ($p = 0.01$), T ($p = 0.05$) and TMJ ($p < 0.01$). The difference in distance between the COM and SD reduced significantly in all directions
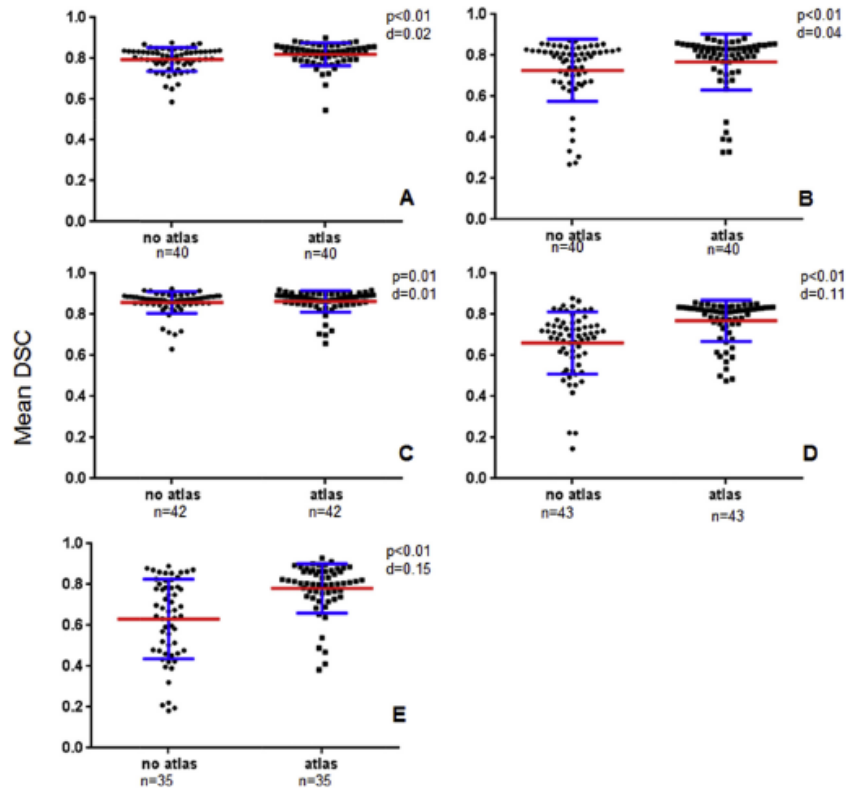

Fig. 2. Comparison of the Dice similarity coefficient (DSC) for individual clinician manual contours without and with the atlas across the 5 pairs of masticatory muscles. (A) Lateral pterygoid, (B) Medial pterygoid, (C) Masseter, (D) Temporalis, (E) Temporo-mandibular Joint. Red horizontal bar illustrates the mean, blue bars illustrate the standard deviation. d equals the difference in absolute mean values.
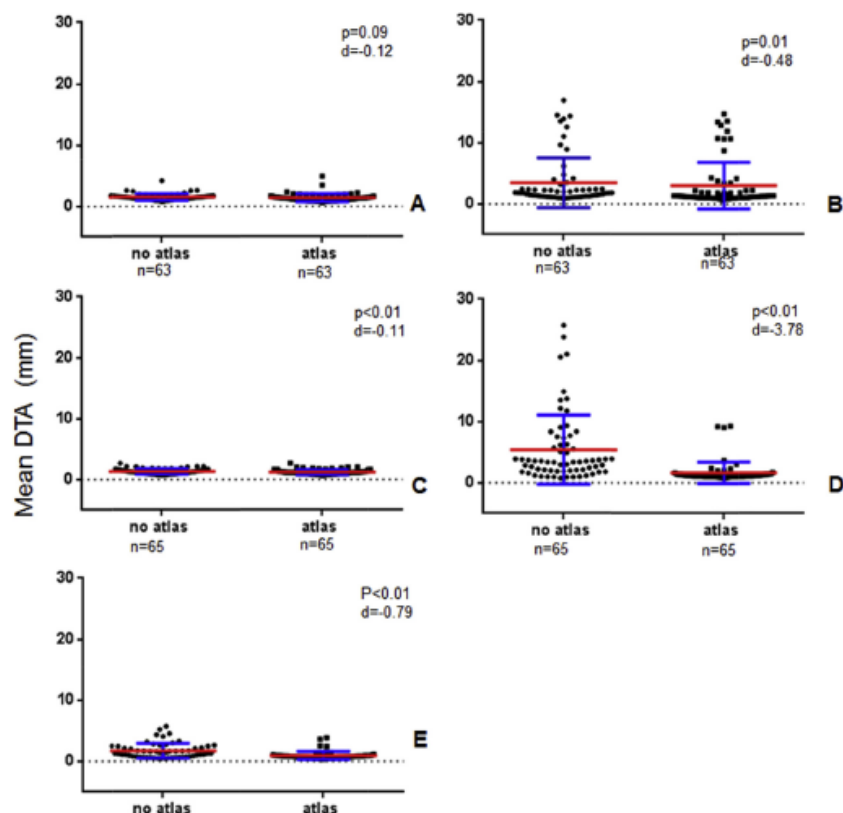
Fig. 3. Comparison of the mean Distance to agreement (DTA) of all individual clinician manual contours without and with the atlas for the 5 pairs of masticatory muscles. Lateral pterygoid, (B). Medial pterygoid, (C). Masseter, (D). Temporalis, (E). Temporo-mandibular Joint. Red horizontal bar illustrates the mean, blue bars illustrate the standard deviation. d equals the difference in absolute mean values.

with the atlas for the T: anterior-posterior 2.1 ± 1.4 vs 4.7 ± 4.7 mm, $p = 0.03$; left–right 4.5 ± 3.0 vs 8.7 ± 5.9 mm, $p = 0.03$; superior-inferior 3.4 ± 2.8 vs 7.0 ± 4.7 mm, $p = 0.05$. No significant difference in the COM distance was observed with the atlas for the LP, MP and M.

Mean DSC significantly improved using the atlas for the LP (0.8 ± 0.1 vs 0.8 ± 0.1 $p < 0.01$), MP (0.7 ± 0.2 vs 0.7 ± 0.2, $p < 0.01$), T (0.7 ± 0.2 vs 0.8 ± 0.1, $p < 0.01$) and TMJ (0.6 ± 0.2 vs 0.8 ± 0.1, $p < 0.01$). No significant improvement in mean DSC was observed using the atlas for the M (0.9 ± 0.1 vs 0.9 ± 0.0, $p = 0.27$) (Fig. 2).

Mean DTA improved using the atlas for all muscles, reaching significance for the MP (3.5 ± 4.1 mm vs 3.0 ± 3.8, $p = 0.01$), M (1.4 ± 0.4 vs 1.2 ± 0.4, $p < 0.01$), T (5.4 ± 5.6 vs 1.6 ± 1.7 mm, $p < 0.01$) and TMJ (1.7 ± 1.2 vs 0.9 ± 0.7 mm, $p < 0.01$), see Fig. 3. Using the atlas, the mean DTA improved for the LP but the variability increased, however this was not significant (1.5 ± 0.5 vs 1.4 ± 0.7 mm, $p = 0.09$).

Fig. 4 illustrates standard deviation maps on representative slices for the M and T for a single patient. Regions in which there is variation between clinician contours are shown in varying degrees of blue – the darker the shade of blue the larger the variation. The atlas reduced the variation between clinicians for the T particularly at the cranial and caudal aspects of the muscle. The reduction in variability with the atlas was smaller for the M.

Table 2 shows the analysis of the mean ± SD DTA without and with the atlas performed according to clinician training grade. An improvement in mean DTA using the atlas was observed by the trainees across all masticatory muscles, with the largest improvement and reduction in variability noted for the T (4.3 ± 7.1 v 1.2 ± 0.4 mm, $p = 0.06$) and TMJ (2.1 ± 0.7 v 0.8 ± 0.3 mm, $p < 0.01$).

Discussion

This prospective study is the first to test the feasibility of a novel atlas of muscles of mastication for contouring in head and neck radiotherapy. Feasibility was defined as a reduction in inter-observer variability. This study showed the atlas: (i) improved spatial overlap and alignment of contours for the LP, MP, T and TMJ; and (ii) improved consistency in contouring of all masticatory muscles by the trainees.

Radiation induced trismus is a significant cause of treatment related morbidity [25]. The main application of contouring the masticatory apparatus as an avoidance structure is for tumours not infiltrating the muscles in order to spare normal healthy tissue. Clinician variation in OAR contouring may lead to over dosage of the muscles of mastication and as a consequence trismus and poor quality of life. There is overlap between the development of trismus and other health-related quality of life variables. In a paper

227

in consistency across all muscles by the trainees implies the benefit of the atlas as an educational tool for trainees including dosimetrists. The atlas may also be used by other radiotherapy centres to improve consistency, knowledge and establish collaborations to aid the development of multi-institutional clinical trials.

Development of NTCP models can be facilitated by generating agreement in dose constraint parameters, facilitated by a greater consistency in contouring the muscles of mastication. The reduction in variability in contouring the muscles of mastication may translate into a reduction in variability in reported dose to these structures [30]. It is beyond the scope of the present study however to determine the dosimetric effects of reduced clinician inter-observer variation. Future studies to integrate the atlas into an auto-contouring model to reduce inter and intraobserver variability and minimise time constraints should also be considered [31]. Improving consistency of contours of the MP, T and TMJ will help standardise volumes, develop more precise dosimetric parameters which can be implemented into avoidance radiotherapy planning to potentially improve radiation related trismus and quality of life [21].

Whilst this is the first paper to our knowledge that uses a novel atlas for the muscles of mastication to evaluate interobserver variability, the study did not explore intraobserver variability or time constraints. The study only involved CT scan images of five patients, however as seven clinicians contoured each plan, a good measure of interobserver variability was obtained. Established dose–response relationships will be facilitated by a more consistent clinician approach to contouring the masticatory muscles. Future studies with greater consistency in contouring and larger numbers are required to further evaluate dose constraints.

A novel atlas has been developed to contour the masticatory muscles during head and neck radiotherapy planning. The atlas has been shown to significantly reduce interobserver variability for the MP, T and TMJ. The atlas could be considered as an education tool to improve knowledge amongst trainees and provide contouring consistency to aid the development of multi-institutional clinical trials. The atlas has been developed into an app for wider distribution amongst radiotherapy centres. Reducing interobserver variability and standardising treatment volumes will improve the accuracy of avoidance planning and potentially reduce radiation related trismus.

Conflicts of interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2018.10.030.

References

[1] Gomez-Millán J, Fernández JR, Medina Carmona JA. Current status of IMRT in head and neck cancer. Reports Pract Oncol Radiother J Gt Cancer Cent Pozn Polish Soc Radiat Oncol 2013;18:371–5.
[2] Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide SA, Clark C, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. Lancet Oncol 2011;12:127–36.
[3] Lim JY, Leech M. Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck. Acta Oncol 2016;55:799–806.
[4] Loo SW, Martin WMC, Smith P, Cherian S, Roques TW. Interobserver variation in parotid gland delineation: A study of its impact on intensity-modulated radiotherapy solutions with a systematic review of the literature. Br J Radiol 2012;85:1070–7.
[5] Anderson CM, Sun W, Buatti JM, Maley JE, Policeni B, Mott SL, et al. Interobserver and intermodality variability in GTV delineation on simulation CT, FDG-PET, and MR Images of Head and Neck Cancer. Jacobs J Radiat Oncol 2014;1:006.
[6] Segedin B, Petric P. Uncertainties in target volume delineation in radiotherapy – Are they relevant and what can we do about them? Radiol Oncol 2016;50:254–62.
[7] Vinod SK, Min M, Jameson MG, Holloway LC. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. Vol. 60. J Med Imaging Radiat Oncol 2016;393–406.
[8] Choi M, Refaat T, Lester MS, Bacchus I, Rademaker AW, Mittal BB. Development of a standardized method for contouring the larynx and its substructures. Radiat Oncol 2014;9.
[9] Pirozzi S, Horvat M, Piper J, Nelson A. SU-E-J-106: Atlas-based segmentation: evaluation of a multi-atlas approach for lung cancer. Med Phys 2012:3677.
[10] Daisne J-F, Blumhofer A. Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: a clinical validation. Radiat Oncol 2013;8:154.
[11] Hansen CR, Johansen J, Samsøe E, Andersen E, Petersen JBB, Jensen K, et al. Consequences of introducing geometric GTV to CTV margin expansion in DAHANCA contouring guidelines for head and neck radiotherapy. Radiother Oncol 2017.
[12] Brouwer CL, Steenbakkers RJHM, Bourhis J, Budach W, Grau C, Grégoire V, 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. Radiother Oncol 2015;117:83–90.
[13] Dhanrajani P, Jonaidel O. Trismus: aetiology, differential diagnosis and treatment. Dent Updat 2002;29:88–92. 94.
[14] Zeller JL. High suicide risk found for patients with head and neck cancer. J Am Med Assoc 2006;296:1716–7.
[15] Lee R, Yeo ST, Rogers SN, Caress AL, Molassiotis A, Ryder D, et al. Randomised feasibility study to compare the use of Therabite with wooden spatulas to relieve and prevent trismus in patients with cancer of the head and neck. Br J Oral Maxillofac Surg 2018.
[16] Rao SD, Saleh ZH, Setton J, Tam M, McBride SM, Riaz N, et al. Dose-volume factors correlating with trismus following chemoradiation for head and neck cancer. Acta Oncol 2016;55:99–104.
[17] Teguh DN, Levendag PC, Voet P, Van Der Est H, Noever I, De Kruijf W, et al. Trismus in patients with oropharyngeal cancer: Relationship with dose in structures of mastication apparatus. Head Neck 2008;30:622–30.
[18] Van Der Molen L, Heemsbergen WD, De Jong R, Van Rossum MA, Smeele LE, Rasch CRN, et al. Dysphagia after chemoradiotherapy Dysphagia and trismus after concomitant chemo-Intensity-Modulated Radiation Therapy (chemo-IMRT) in advanced head and neck cancer; Dose-effect relationships for swallowing and mastication structures. Radiother Oncol 2013;106:364–9.
[19] Pauli N, Johnson J, Finizia C, Andréll P. The incidence of trismus and long-term impact on health-related quality of life in patients with head and neck cancer. Acta Oncol 2013;52:1137–45.
[20] Beasley W, Thor M, McWilliam A, Green A, Mackay R, Slevin N, et al. Image-based data mining to probe dosimetric correlates of radiation-induced trismus. Int J Radiat Oncol Biol Phys 2018.
[21] De Felice F, Musio D, Tombolini V. Mastication structures definition in head and neck cancer. Vol. 118. Radiother Oncol 2016:419.
[22] Hague C, Beasley W, Garcez K, Lee LW, McPartlin A, McWilliam A, et al. Prospective evaluation of relationships between radiotherapy dose to masticatory apparatus and trismus. Acta Oncol (Madr) 2018:1–5.
[23] Goldstein M, Maximiw WG, Cummings BJ, Wood RE. The effects of antitumor irradiation on mandibular opening and mobility: a prospective study of 58 patients. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 1999;88:365–73.
[24] Brouwer CL, Steenbakkers RJHM, van den Heuvel E, Duppen JC, Navran A, Bijl HP, et al. 3D Variation in delineation of head and neck organs at risk. Radiat Oncol 2012;7.
[25] Loh SY, Mcleod RWJ, Elhassan HA. Trismus following different treatment modalities for head and neck cancer: a systematic review of subjective measures. Eur Arch Otorhinolaryngol 2017;274:2695–707.
[26] Lee R, Slevin N, Musgrove B, Swindell R, Molassiotis A. Prediction of post-treatment trismus in head and neck cancer patients. Br J Oral Maxillofac Surg 2012;50:328–32.
[27] Gebre-Medhin M, Haghanegi M, Robért L, Kjellén E, Nilsson P. Dose-volume analysis of radiation-induced trismus in head and neck cancer patients. Acta Oncol 2016;55:1313–7.
[28] Zheng Y, Han F, Xiao W, Xiang Y, Lu L, Deng X, et al. Analysis of late toxicity in nasopharyngeal carcinoma patients treated with intensity modulated radiation therapy. Radiat Oncol 2015;10:17.
[29] Wu VWC, Lam Y-N. Radiation-induced temporo-mandibular joint disorder in post-radiotherapy nasopharyngeal carcinoma patients: assessment and treatment. J Med Radiat Sci 2016;63:124–32.
[30] Beasley W, McWilliam A, Aitkenhead A, Mackay RI, Rowbottom CG. The suitability of common metrics for assessing parotid and larynx autosegmentation accuracy. J Appl Clin Med Phys 2016.
[31] Teguh DN, Levendag PC, Voet PWJ, Al-Mamgani A, Han X, Wolf TK, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. Int J Radiat Oncol Biol Phys 2011;81:950–7.

PROTON THERAPY SPECIAL FEATURE: FULL PAPER

Inter-fraction robustness of intensity-modulated proton therapy in the post-operative treatment of oropharyngeal and oral cavity squamous cell carcinomas

[1,2]CHRISTINA HAGUE, MRCP, FRCR, [3,4]MARIANNE AZNAR, PhD, [5]LEI DONG, PhD, [5]ALIREZA FOTOUHI-GHIAM, MD, [1]LIP WAI LEE, FRCR, [5]TAORAN LI, PhD, [5]ALEXANDER LIN, MD, [6]MATTHEW LOWE, PhD, [5]JOHN N LUKENS, MD, [1]ANDREW MCPARTLIN, MD, [5]SHANNON O'REILLY, PhD, [1,2]NICK SLEVIN, PhD, [5]SAMUEL SWISHER-MCCLURE, MD, [1,2]DAVID THOMSON, MD, [7,8]MARCEL VAN HERK, PhD, [9,10]CATHARINE WEST, PhD, [5]WEI ZOU, PhD and [5]BOON-KENG KEVIN TEO, PhD

[1]Department of Clinical Oncology, The Christie NHS Foundation Trust, Manchester, UK
[2]University of Manchester, Manchester Academic Health Sciences Centre, Manchester, UK
[3]Manchester Cancer Research Centre, Division of Molecularand Clinical Cancer Science, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic HealthSciences Centre, Manchester, UK
[4]The Christie NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, UK
[5]Department of Radiation Oncology, University of Pennsylvania, Philadelphia, USA
[6]Christie Medical Physics and Engineering, The Christie NHS Foundation Trust, Manchester, UK
[7]Division Clinical Cancer Science, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
[8]Department of Radiotherapy Related Research, The Christie NHS Foundation Trust, Manchester, UK
[9]Translational Radiobiology Group, Division of Cancer Sciences, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK
[10]The Christie NHS Foundation Trust, Manchester, UK

Address correspondence to: Dr Christina Hague
E-mail: christinahague@doctors.org.uk

Objective: To evaluate dosimetric consequences of inter-fraction setup variation and anatomical changes in patients receiving multifield optimised (MFO) intensity modulated proton therapy for post-operative oropharyngeal (OPC) and oral cavity (OCC) cancers.

Methods: Six patients receiving MFO for post-operative OPC and OCC were evaluated. Plans were robustly optimised to clinical target volumes (CTVs) using 3 mm setup and 3.5% range uncertainty. Weekly online cone beam CT (CBCT) were performed. Planning CT was deformed to the CBCT to create virtual CTs (vCTs) on which the planned dose was recalculated. vCT plan robustness was evaluated using a setup uncertainty of 1.5 mm and range uncertainty of 3.5%. Target coverage, $D_{95\%}$ and hotspots, $D_{0.03cc}$ were evaluated for each uncertainty along with the vCT-calculated nominal plan. Mean dose to organs at risk (OARs) for the vCT-calculated nominal

plan and relative % change in weight from baseline were evaluated.

Results: Robustly optimised plans in post-operative OPC and OCC patients are robust against inter-fraction setup variations and range uncertainty. $D_{0.03cc}$ in the vCT-calculated nominal plans were clinically acceptable across all plans. Across all patients D95% in the vCT-calculated nominal treatment plan was at least 100% of the prescribed dose. No patients lost ≥10% weight from baseline. Mean dose to the OARs and max dose to the spinal cord remained within tolerance.

Conclusion: MFO plans in post-operative OPC and OCC patients are robust to inter-fraction uncertainties in setup and range when evaluated over multiple CT scans without compromising OAR mean dose.

Advances in knowledge: This is the first paper to evaluate inter-fraction MFO plan robustness in post-operative head and neck treatment.

INTRODUCTION

Radiotherapy for head and neck cancers is challenging due to the proximity of normal structures. Irradiation of healthy tissue can lead to long-term toxicities such as xerostomia,

dysphagia and dysgeusia that have a negative impact on quality of life.[1] The finite range of high energy protons makes proton beam therapy an attractive treatment option for squamous cell cancers of the head and neck by limiting

dose to normal tissues beyond the target volume. However, the sharp distal fall-off of proton beams makes the dose distribution sensitive to setup, motion and range uncertainties. In proton treatment planning, potential sources of error resulting from uncertainty in the location of the Bragg peak need to be considered. Potential sources of error include: daily changes in the patients' position or anatomy, organ motion, delineation variation, image artefacts, inaccurate conversion of CT Hounsfield units to the proton stopping power and changes in the beam path due to variable tissue densities.[2,3] Patient immobilisation, image guidance, expansion margins and use of a dual energy CT may reduce errors but in clinical practice centre-specific setup uncertainties and ±3.5% uncertainty on stopping powers are typically used to account for residual uncertainties.[4,5] Despite attempts to reduce setup uncertainties, some daily variations are unavoidable and may result in under- or overdosage of the target volume and organs at risk (OARs) respectively.

Robustness to uncertainty may be attained in many ways. One such approach is through robust plan optimisation whereby uncertainties are accounted for in the optimisation process. The optimisation process can reduce dose gradients within the plan and consequently, the occurrence of hot and cold spots due to uncertainties in the proton beam range. In robust optimisation, patient setup errors and the corresponding perturbations in the dose distribution are simulated by shifts in the isocentre position on a single CT scan. Range uncertainties are simulated by scaling the CT to stopping power ratio calibration. Setup variation, however, is more complex than a rigid translation of the CT image and can involve soft tissue deformation, variations in spine alignment, neck tilt and shoulder position. In the literature, different methods of robust plan optimisation exist, including probabilistic treatment planning, the use of worst-case scenarios, selective robust and minimax optimisation.[6–8] Probabilistic treatment planning described by Unkelbach incorporates both range and setup as random variables into the optimisation process. This contrasts with worst-case optimisation described by Liu at al whereby the optimisation is based on the worst case scenario.[9,10] However, the robust optimisation process only accounts for positional and range uncertainties, and does not evaluate robustness against anatomical changes during the treatment course. Commercial systems using multiple patient images prior to treatment are being studied for anatomical robust optimisation but are yet to be adopted in routine clinical practice.[11,12] Using pencil beam scanning, two methods for plan optimisation can be used: single field optimisation (SFO) and multifield optimisation (MFO). In the latter, all beam spots are optimised together using multiple fields and beam angles to modulate and shape the dose. A superior dose distribution can be delivered compared with passively scattered proton beams or SFO, in particular in cases of complex geometry. MFO, however, can be prone to inter-fractional uncertainties due to anatomic changes, such as from weight loss or tumour response, resulting from increased in-field dose gradients. These non-linear changes in patient's anatomy due to tissue deformations are not explicitly modelled in robust optimisation; it is unknown if robustly optimised MFO can still accommodate these unplanned range variations.

Conventionally, uncertainties are accounted for with the use of a planning target volume (PTV). In a study of fourteen patients by Liu et al, a robust clinical target volume (CTV) based optimisation approach produced a superior plan for targets and OARs compared with a PTV approach.[5] Stutzer et al showed MFO to be superior to SFO in original plan robustness in the treatment of oropharyngeal cancer.[13]

MFO is potentially more sensitive to tissue deformation during the treatment course, which is likely to be different along different beam angles. The effectiveness of robust plan optimisation in MFO has not been studied in post-operative head and neck treatment planning. There is also no agreed consensus as to which optimisation method is superior in evaluating MFO plans partly due to variable literature and lack of standardised protocols. In this study, we aim to evaluate how robust MFO plans are to uncertainties in set up and range error for patients treated post-operatively for oropharyngeal and oral cavity cancers using weekly cone beam CT (CBCT) scans. Such CBCT scans capture realistic inter-fraction setup variations as well as tissue deformations during the entire treatment course.

METHODS

Patient selection
Six head and neck patients treated between July 2017 and April 2018 and planned with multifield robust optimisation in Eclipse (Eclipse v. 13.7, Varian Medical systems, Palo Alto, CA) were retrospectively selected. Inclusion criteria were: patients with oropharyngeal or oral cavity cancers requiring post-operative proton beam therapy to the primary site and elective neck, and the availability of weekly CBCT images. All patients were immobilised in a 5-point thermoplastic mask and positioned using daily orthogonal kV imaging. The relative percentage change in each patient's weight was recorded weekly during treatment which was subsequently correlated with CTV and OAR coverage.

Planning approach
Each patient was treated with two or three CTV dose levels. CTV1 was defined as the surgical bed with CTV2 and CTV3 defining "at-risk sites". CTV1 received 60–63 Gy (RBE), CTV2 received 54–60 Gy (RBE) and CTV3 received 54 Gy (RBE) all treated with a dose of between 1.8 and 2.1 Gy (RBE) per fraction. Each treatment plan was based on a three-field beam arrangement consisting of two posterior obliques and one anterior field with range shifters (7.2 cm anterior 8.0 cm posterior) *in situ* as shown in Figure 1(a) , and (b). The posterior oblique fields cover the superior portion of the target while the anterior field covers the inferior target. The posterior oblique and anterior fields overlap in the superior–inferior direction over a 2 cm region and are optimised to be robust against 3 mm longitudinal isocentre perturbations between the posterior oblique and anterior fields. In this way, smooth dose gradients are achieved in the superior–inferior directions for each field at the overlap region without the need to feather the match line. Range shifters are positioned close to the patient to minimise the post-range shifter air gap and subsequent increase in spot size.

The total prescribed dose to the original treatment nominal plan was defined such that 95% of each target volume received

230

Figure 1. Three field IMPT beam arrangement (a) and (b) with two posterior oblique fields and an anterior field indicated by red arrows. *In situ* range shifters located anterior and posterior to the patient are used to minimise the air gap between the range shifter and the patient. Representative lateral (c) and anterior (d) DRRs of one planning CT and maximum intensity projection of 5 vCTs (e) and (f) with red contours representing inter-fraction setup variation of cervical vertebrae. DRR, digitally reconstructed radiograph; IMPT, intensity modulated proton therapy; vCTs, virtual CTs.
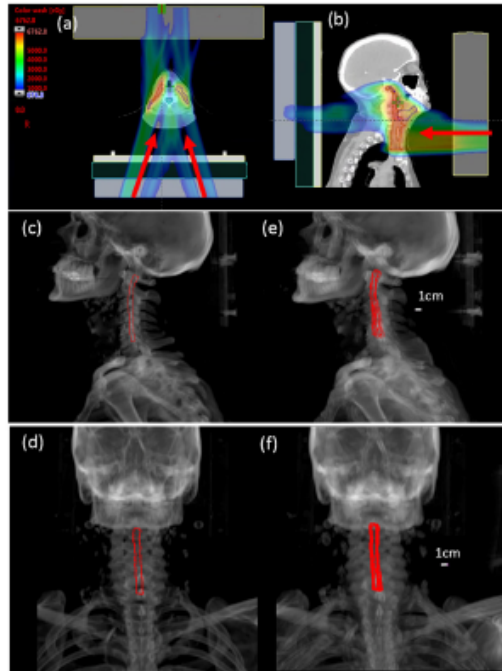
a minimum of 100% of the prescribed dose to this CTV level. Clinically acceptable plans require 95% of the target volume in the worst-case scenario to receive at least 95% of the prescribed dose and $D_{0.03cc}$ to receive ≤110% of the prescribed dose.

Analysis of robustness

Plans were optimised using MFO. CTVs were robustly optimised using a setup uncertainty of 3 mm and a range uncertainty of 3.5%. Setup perturbations were simulated along three orthogonal directions and combined with range uncertainties in each position to produce 12 scenarios used in the robust optimisation, $(x \pm 3\,\text{mm})$, $(y \pm 3\,\text{mm})$, $(z \pm 3\,\text{mm})$ and CT Hounsfield unit (HU) scaling of ±3.5%.

Daily positioning and setup correction were accomplished using orthogonal kV imaging. Each patient underwent weekly online CBCT, which were then used to create reliable virtual CTs (vCTs). If a CBCT was acquired, a 3D–3D match was performed followed by a confirmatory kV imaging without additional couch

motion. The vCTs were generated by deforming the planning CT onto the CBCT using a diffeomorphic implementation of the Morphons algorithm.[14] The deformable registration provided a high-quality image in the treatment geometry on which the planned dose could be evaluated to estimate the delivered dose. Each target CTV on the vCT was compared with the original treatment nominal plan and manually edited to modify the superior and inferior extension to ensure consistency of anatomical landmarks. The vCT plan robustness evaluation was performed using a residual setup uncertainty of 1.5 mm and range uncertainty of 3.5%. The 1.5 mm setup robustness evaluation considers the uncertainty in the coincidence between the imaging and radiation isocentres, intra-fraction motion, as well as variations in user dependent choice of region of interest for evaluating image registration between the vCT and the planning CT. Actual setup variability of the cervical neck vertebrae is illustrated in the digitally reconstructed radiographs depicted in Figure 1(e), and (f) for Patient 5. A video depicting the setup variation is available in the supplementary data.

Dose calculation

All doses and dose–volume histograms (DVHs) were calculated using proton convolution superposition (PCS) v. 13.7 within the Eclipse™ treatment planning system (Varian Medical Systems, Palo Alto, CA). Plan metrics were extracted from the calculated DVHs using a MATLAB (Mathworks, Natick, MA) script. Maximum and minimum values under uncertainty for $D_{95\%}$ for each CTV dose level and $D_{0.03cc}$ for the high dose CTV were extracted and compared with the original treatment nominal value. Mean dose in the vCT-calculated nominal case to ipsilateral and contralateral parotid glands, oral cavity, pharyngeal constrictor muscles, larynx and maximum dose to the spinal cord were calculated. Relative % change in weight from baseline to end of treatment was recorded.

RESULTS

The demographics of the six patients are outlined in Table 1. In all patients, the vCT-calculated nominal treatment plan $D_{95\%}$ across the vCTs was at least 100% of the prescribed dose. The median time from baseline radiotherapy planning scan to first CBCT within week 1 of treatment was 29 (range 22–52) days.

Across the six patients, robustness evaluation of CTV coverage using inter-fraction uncertainties of 3 mm setup and 3.5% range error for the initial plan (day 0) and 1.5 mm for the residual setup and 3.5% range for the vCT-calculated plans are shown in Figure 2a and b. The dose to the vCT-calculated nominal plan is close to the maximum CTV dose band across all six plans. Clinically acceptable $D_{0.03cc}$ values in the high dose CTV levels are shown in Figure 2a and b with the exception of Patient 6. In Patient 6, the highest $D_{0.03cc}$ value for the vCT-calculated nominal plan is 113% (71 vs. 63 Gy) of the target dose. The hot spot in patient six is shown in Figure 3 and can be explained by a shift in the jaw position and variation of the posterior neck. When uncertainty analysis is included, the highest prescribed $D_{0.03cc}$ for this fraction was 116% of the target dose (73 vs. 63 Gy if scaled to the full treatment dose). Analysis of daily kV setup image revealed that this relatively large variation in jaw position

231

Table 1. Patients' demographics

Patient number	1	2	3	4	5	6
Age (years)	73	50	67	70	55	74
Gender	M	M	M	F	M	M
Site	Base of tongue	Tonsil	Oral tongue	Tonsil	Vallecula	Oral tongue
Laterality of tumour	Right	Right	Right	Right	Left	Left
Stage	T2N2bM0	T2N2aM0	T2N1M0	T2N1M0	T4aN1M0	T2N0M0
HPV status	Positive	Positive	Negative	Positive	Positive	Negative
Chemotherapy	None	Weekly Cisplatin	None	Weekly Cisplatin	Weekly Cisplatin	Weekly Cisplatin
Prescription dose/Gy RBE	60	63	60	63	63	63
Baseline weight	78	92	79	78	99	109
Weight loss in kg at the end of treatment (relative % change)	0	3 (3%)	0	5 (6%)	1 (1%)	7 (6%)

F, Female; M, Male; RBE, relative biological effectiveness.

occurred only once during the entire course of treatment. The $D_{0.03cc}$ values for the other two vCT-calculated nominal plans of Patient 6 in Figure 2b were within the uncertainty bands of the initial plan.

None of the six patients lost >10% weight relative to their baseline throughout treatment. The largest absolute (relative %) weight loss were noted in patients 4 and 6 both of whom lost 6% by the end of treatment. Two patients lost no weight and in the remaining two patients their relative % weight loss remained

within 5%. Figure 4 illustrates the mean and maximum doses to the OARs with time.

The impact of inter-fraction setup, anatomical variation and changes in patient weight on the OAR doses is depicted in Figure 5. While vCT-calculated nominal doses do vary for each patient, dose constraints did not exceed planning objectives for any OARs.

Figure 2. Figure shows an overview of the change in $D_{95\%}$, $D_{0.03cc}$ in the high-risk CTV and patient weight from baseline to the end of treatment. Uncertainties over all error scenarios are shown as a light band, CTV 63 (green), CTV 60 (red), CTV 54 (blue). The solid line shows the nominal dose. Dashed lines indicate weight changes of ±5% from the baseline weight recorded at the time of the planning CT scan. CTV, clinical target volume.
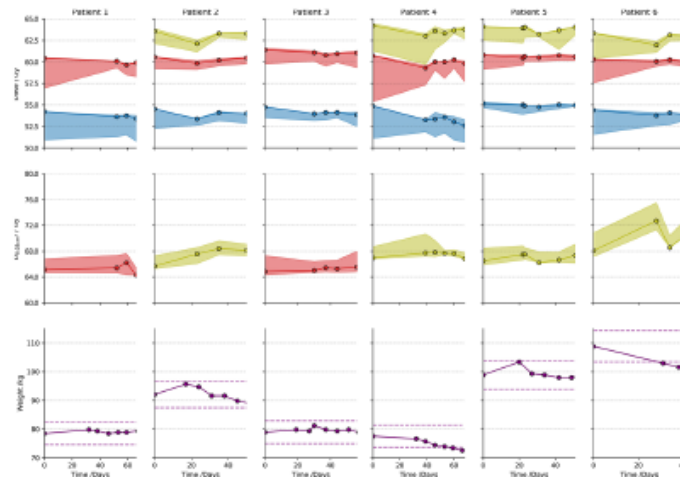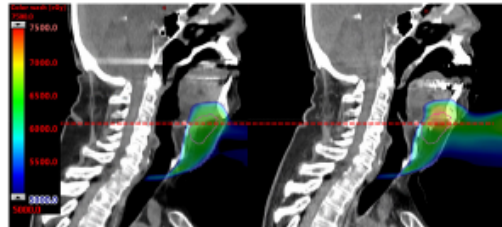
232

Figure 3. (a) Sagittal view of planning CT patient six with CTV1 contour and (b) corresponding slice of the first vCT-calculated nominal plan. The hotspot is attributed to differences in the jaw position (shown by the dashed red reference line) and the setup variation of the posterior neck region. CTV, clinical target volume; vCT, virtual CT.

DISCUSSION

This novel pilot study to evaluate dosimetric consequences of anatomical variations and range errors in MFO plans in the post-operative setting has demonstrated:

(i) plans are robust to uncertainties in setup and range, (ii) weight loss within 6% does not negatively impact dosimetric coverage, (iii) OAR exposure for all vCT-calculated nominal plans are within tolerance.

MFO plans have been shown to be robust to uncertainties early in the treatment course enabling plan adaptation if necessary. Weight loss of >10% from baseline in head and neck radio-therapy is influenced by radiation related toxicities and significantly associated with global poor health-related quality of life.[15] Changes in weight can affect the positioning of OARs such as the parotid gland resulting in potential overdosage. The use of proactive enteral feeding is the cause of much debate. In this study, none of the six patients lost ≥10% weight throughout treatment and required enteral feeding. Of the two patients who lost the largest amount of weight (6%) from baseline, plans remained robust to uncertainties. This highlights the potential importance of aggressive and proactive symptom and supportive management during treatment, so as to minimise weight loss.

Proton planning, particularly in head and neck radiotherapy is challenging due to uncertainties in setup and range on a

Figure 4. Figure illustrates the changes in mean dose to the oral cavity, pharyngeal constrictor muscles, larynx, ipsilateral parotid gland and maximum dose to the spinal cord with time. Dose constraints to each OAR are; oral cavity mean dose 20 Gy or ALARA, pharyngeal constrictor muscles mean dose 50 Gy or ALARA, larynx mean dose 20 Gy or ALARA, ipsilateral parotid gland mean dose 20 Gy or ALARA, maximum spinal cord 45 Gy. ALARA as low as reasonably achievable; OAR, organ at risk.
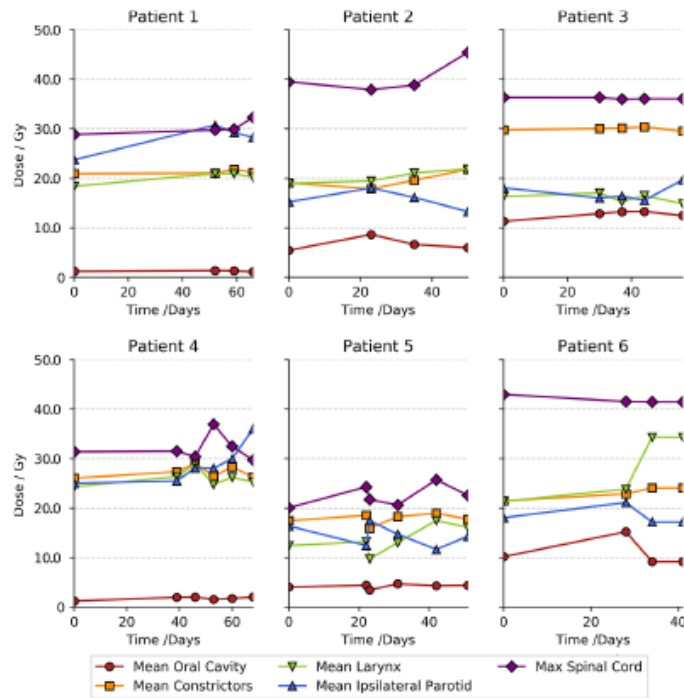
233

Figure 5. Axial (a) and sagittal (b) views of setup variation in the neck region between planned and actual treatment positions for Patient 4. Blended images of planning CT and vCT shown demonstrating differences in the neck angle as indicated by red arrows. Nominal dose distribution on the planning CT is shown in (c) with CTV 54 contour in blue and the corresponding vCT dose. Green body contour of the planning CT is shown in (c) and (d). Red dotted circle in (d) indicates region of hotspot outside of CTV 54 and cold spot within CTV 54. Red arrow in (d) indicates region with variation in setup position. CTV, clinical target volume; vCT, virtual CT.
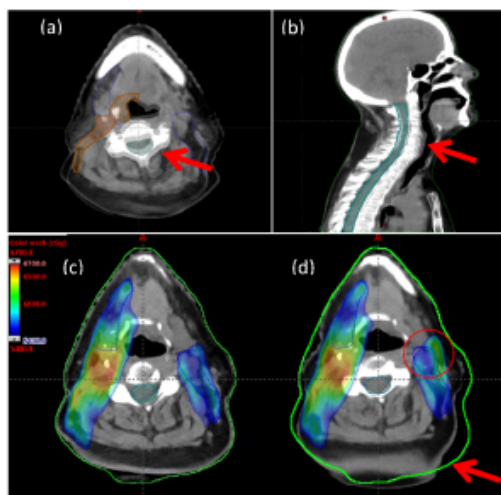
background of complex anatomy and close proximity of tumours with normal tissues. Optimisation techniques may help minimise the effect of uncertainties, accepting the trade-off between delivering an acceptable integral dose and normal tissue sparing.[16] In this study, dose constraints to the oral cavity, pharyngeal constrictor muscles, larynx and parotid glands were achieved without compromising plan robustness.

MFO proton beam plans in head and neck radiotherapy are susceptible to the development of hot and cold spots due to anatomical motion and changes in the position of the head and neck. During radiotherapy, as patients relax, the neck and jaw position can move, which potentially alters the dose distribution. In all six patients, the distribution of hot spots in the vCT-calculated nominal plan as defined by $D_{0.03cc}$ were clinically

acceptable (within 110% of the prescribed dose and 115% for the worst case). In Patient 6, $D_{0.03cc}$ of 113% in the vCT-calculated nominal plan was noted at Day 34 from baseline. The development of a hot spot correlated with a shift in the jaw position as shown in Figure 3. The dose distribution corrected to within the clinically acceptable range on subsequent weekly CBCTs. Changes in the patients' position is a random variable unlike weight loss which is more systematic throughout treatment. More regular imaging with daily CBCTs could be considered in those who experience hot spots, $D_{0.03cc}$ >110% and weight loss >10% to ensure an adequate dose distribution to the target and maintain OAR constraints. Another example of the impact of setup variation of the neck and its impact on dose distribution of MFO plans is depicted in Figure 5 where the hotspot is seen outside of the CTV. Current robust optimisation techniques do not take into account anatomical deformation.

Although this is the first study to evaluate optimisation of MFO plans in post-operative oropharyngeal and oral cavity patients, the sample size is small and not all patients underwent the same number of CBCT scans during treatment. In addition, the mean doses at each vCT-calculated nominal plan reflect dose to the full prescription but the measurement point is for a single fraction. A more thorough analysis could be performed with daily CBCT using deformable dose accumulation on the vCTs. Unfortunately, daily CBCT was not available for this study. Challenges of maintaining robustness may be different in other anatomical sites in the head and neck, such as in the paranasal sinuses and skull base, where random, daily changes in sinus filling may occur. Despite these limitations this is the first paper to our knowledge to evaluate MFO robust optimisation in post-operative oropharyngeal and oral cavity cancer patients and may help in the process of developing optimisation protocols with agreed parameters to help identify plans that require additional individualisation.

CONCLUSION
In this novel pilot study, MFO plans in post-operative oropharyngeal and oral cavity patients were observed to be robust to uncertainties in setup and range in regard to CTV coverage when evaluated over multiple CBCT scans. Dose constraints to OARs remained within tolerance and changes in weight from baseline did not appear to affect CTV coverage or plan quality. Development of a robust analysis protocol for MFO plans may improve consistency of reporting and plan evaluation amongst radiotherapy centres.

ACKNOWLEDGMENT

REFERENCES

1. Kjaer T, Dalton SO, Andersen E, Karlsen R, Nielsen AL, Hansen MK, et al. A controlled study of use of patient-reported outcomes to improve assessment of late effects after treatment for head-and-neck cancer. *Radiother Oncol* 2016; 119: 221–8. doi: https://doi.org/10.1016/j.radonc.2016.04.034

2. Schneider U, Pedroni E, Lomax A. The calibration of CT Hounsfield units for radiotherapy treatment planning. *Phys Med*

234

Biol 1996; 41: 111–24. doi: https://doi.org/10.1088/0031-9155/41/1/009

3. Jäkel O, Reiss P. The influence of metal artefacts on the range of ion beams. *Phys Med Biol* 2007; 52: 635–44. doi: https://doi.org/10.1088/0031-9155/52/3/007

4. Lomax AJ. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions. *Phys Med Biol* 2008; 53: 1043–56. doi: https://doi.org/10.1088/0031-9155/53/4/015

5. Liu W, Frank SJ, Li X, Li Y, Park PC, Dong L, et al. Effectiveness of robust optimization in intensity-modulated proton therapy planning for head and neck cancers. *Med Phys* 2013; 40: 051711. doi: https://doi.org/10.1118/1.4801899

6. McGowan SE, Albertini F, Thomas SJ, Lomax AJ. Defining robustness protocols: a method to include and evaluate robustness in clinical plans. *Phys Med Biol* 2015; 60: 2671–84. doi: https://doi.org/10.1088/0031-9155/60/7/2671

7. Li Y, Niemela P, Liao L, Jiang S, Li H, Poenisch F, et al. Selective robust optimization: a new intensity-modulated proton therapy optimization strategy. *Med*

Phys 2015; 42: 4840–7. doi: https://doi.org/10.1118/1.4923171

8. Fredriksson A, Forsgren A, Hårdemark B. Minimax optimization for handling range and setup uncertainties in proton therapy. *Med Phys* 2011; 38: 1672–84. doi: https://doi.org/10.1118/1.3556559

9. Unkelbach J, Chan TCY, Bortfeld T. Accounting for range uncertainties in the optimization of intensity modulated proton therapy. *Phys Med Biol* 2007; 52: 2755–73. doi: https://doi.org/10.1088/0031-9155/52/10/009

10. Liu W, Zhang X, Li Y, Mohan R. Robust optimization of intensity modulated proton therapy. *Med Phys* 2012; 39: 1079–91. doi: https://doi.org/10.1118/1.3679340

11. Cubillos-Mesías M, Troost EGC, Lohaus F, Agolli L, Rehm M, Richter C, et al. Including anatomical variations in robust optimization for head and neck proton therapy can reduce the need of adaptation. *Radiother Oncol* 2019; 131: 127–34. doi: https://doi.org/10.1016/j.radonc.2018.12.008

12. Yang Z, Zhang X, Wang X, Zhu XR, Gunn B, Frank SJ, et al.;in press Multiple-CT optimization: an adaptive optimization method to account for anatomical changes

in intensity-modulated proton therapy for head and neck cancers. *Radiotherapy and Oncology* 2019; 2019. doi: https://doi.org/10.1016/j.radonc.2019.09.010

13. Stützer K, Lin A, Kirk M, Lin L. Superiority in robustness of Multifield optimization over Single-Field optimization for Pencil-Beam proton therapy for oropharynx carcinoma: an enhanced robustness analysis. *Int J Radiat Oncol Biol Phys* 2017; 99: 738–49. doi: https://doi.org/10.1016/j.ijrobp.2017.06.017

14. Knutsson H, Andersson M. Morphons: segmentation using elastic canvas and paint on priors IEEE int. *Conf. on Image Proc* 2005; 2: 1226–9.

15. Langius JAE, van Dijk AM, Doornaert P, Kruizenga HM, Langendijk JA, Leemans CR, et al. More than 10% weight loss in head and neck cancer patients during radiotherapy is independently associated with deterioration in quality of life. *Nutr Cancer* 2013; 65: 76–83. doi: https://doi.org/10.1080/01635581.2013.741749

16. McGowan SE, Burnet NG, Lomax AJ. Treatment planning optimisation in proton therapy. *Br J Radiol* 2013; 86: 20120288. doi: https://doi.org/10.1259/bjr.20120288

235

Contents lists available at ScienceDirect

Clinical Oncology

journal homepage: www.clinicaloncologyonline.net

Editorial

Patient Involvement in the Design of a Phase III Trial Comparing Intensity-modulated Proton Therapy and Intensity-modulated Radiotherapy for Oropharyngeal Cancer

C. Hague [*], B. Foran [†], E. Hall [‡], S. Guild [§], O. Joseph [¶], R. Moule [||], C. Nutting [**], S. Parsons [¶], R. Prestwich [††], N. Slevin [*], C. West [‡‡], D. Thomson [*]

[*] The Christie NHS Foundation Trust, Manchester, UK
[†] Weston Park Hospital, Sheffield, UK
[‡] The Institute of Cancer Research, Clinical Trials and Statistics Unit, London, UK
[§] Patient Representative, Leeds, UK
[¶] Manchester University NHS Foundation Trust, The University of Manchester, Public Programmes Team, Manchester, UK
[||] The University College London Hospital, London, UK
[**] The Royal Marsden NHS Foundation Trust, London, UK
[††] St James' University Hospital, Leeds, UK
[‡‡] Division of Cancer Science, The University of Manchester, Manchester Academic Health Science Centre, Christie Hospital, Manchester, UK

For patients with favourable risk human papilloma virus-associated oropharyngeal cancer, local control and survival outcomes are excellent [1,2]. However, despite the use of highly conformal intensity-modulated radiotherapy (IMRT) severe acute and late side-effects are common, adversely affecting quality of life. Compared with photons, the superior dosimetric properties of protons with sharp lateral penumbra and distal fall-off reduce the radiation dose beyond the target volume and may lessen treatment-related toxicities such as oral mucositis, dryness, taste disturbance, swallowing dysfunction and osteoradionecrosis. However, there are only preliminary observational data to support the clinical advantage of proton beam therapy for oropharyngeal cancer [3–6] and prospective randomised trials are needed. An ongoing phase II/III study (NCT01893307) from the MD Anderson primarily aims to compare rates of late grade 3–5 toxicities between IMRT and intensity-modulated proton therapy (IMPT) for oropharyngeal cancer [7]. The UK proposes to open a multicentre phase III study (TORPEdO, TOxicity Reduction using Proton bEam therapy for Oropharyngeal cancer) to assess the benefit of IMPT in terms of patient-reported toxicities and quality of life and, as a secondary objective, cost-effectiveness.

About 700 of the 1500 funded annual capacity for two planned UK National Health Service (NHS) proton beam centres (The Christie Hospital in Manchester, opening August 2018, and University College London Hospital, opening 2020) will be available for either clinical trials or evaluative commissioning. The UK, with an established strong track record in delivering major practice-changing clinical trials in radiotherapy, e.g. PARSPORT, START and CHHiP [8–10], aims to be at the forefront in establishing the evidence base for the use of proton beam therapy.

Patient and public involvement in the early stages of trial design increases the success of a study in terms of its feasibility and acceptability to patients, thereby supporting recruitment [11–13]. We conducted three focus groups in Manchester, Leeds and Sheffield to understand patients' views about the proposed TORPEdO trial, including acceptability of randomisation, the patient pathway when enrolled in the trial, willingness to travel and stay in Manchester or London for proton beam therapy and the trial design and end points. Fifteen patients with favourable risk oropharyngeal cancer who had completed radiotherapy ≥ 1 year ago were identified and invited to participate from each centre. Overall, 33 of the 45 invited patients and eight relatives attended the focus groups between September and October 2017. Each session lasted 2 h and consisted of presentations and discussions structured around a series of questions (Table 1). Information was recorded on pre-

Author for correspondence: D. Thomson, The Christie NHS Foundation Trust, Manchester, UK.
E-mail address: david.thomson@christie.nhs.uk (D. Thomson).

Table 1
Questions asked during the focus groups

All centres	Additional questions for Leeds and Sheffield
What do you think are the differences between standard radiotherapy and proton therapy?	What are your views on travelling and staying in Manchester if you are offered proton beam therapy as part of the trial?
How do you feel about entering a study where there is a 50% chance of getting standard radiotherapy and a 50% chance of getting proton therapy?	
What are your views on the trial pathway?	
What are your views on the trial outcomes?	

prepared laminates, patient questionnaires, audio-recordings and by telephone or e-mail contact with a sample of patients after each meeting. Data were interpreted using thematic analysis.

Opinions were sought on the name of the proposed study in Manchester. 'TORPEdO' was thought to be concise, easy to remember and would not deter trial participation. Existing knowledge about proton beam therapy was variable. Some described protons as a more targeted therapy with less toxicity, whereas others knew very little but had heard the term 'protons' described in the media. In general, people considered protons to be a superior treatment and had some understanding of the differences compared with standard photon radiotherapy. There was enthusiasm to participate and be randomised in the study, to both help future patients (inform future treatments) and have a 50% chance of receiving IMPT (the patients were informed that proton beam therapy would not be available as an NHS treatment outside the clinical trial). Some expressed that they would be disappointed if randomised to IMRT, but this would not deter them from considering the trial. Reassurance was provided that it remains uncertain in this situation whether any potential dosimetric superiority achieved with proton planning translates into clinical benefit for patients, i.e. there is clinical equipoise. The patient pathway and trial layout were viewed positively. In Sheffield and Leeds, the timings at each stage of the pathway were discussed, e.g. the timings from study enrolment to randomisation. To allow sufficient preparation for those receiving IMPT in Manchester or London, people felt the time interval between each stage of the patient pathway should be minimised (e.g. time from clinic to outcome of randomisation; time from randomisation to travel to Manchester or London). Patients reported that they would need information about the proton centres and provision of support when considering taking part in the trial.

We sought to understand peoples' views on the logistical challenges in relation to travel and accommodation if randomised to IMPT. This was particularly pertinent for those living in Sheffield and Leeds. There was a general willingness to travel to Manchester for treatment planning and delivery, which was balanced against missing family and established social networks. Feelings were mixed in relation to staying in Manchester for the duration of treatment with IMPT. The preferred option was to stay in Manchester in apartment-based accommodation close to the hospital during the week with the option of returning home at the weekends. This preference was thought to be achievable with adequate family, clinical and nursing support. Some raised the possibility and feasibility of daily travel if continuing to work or due to childcare responsibilities. Many patients felt it important for a relative or carer to stay during treatment to provide both emotional and social support. Participants felt that, as far as possible, the accommodation should be tailored to individual's needs, e.g. providing reliable/fast internet access to be able to contact family or equipment/toys for children. Easy access to shuttle transport between the apartment and the hospital was highlighted to avoid potential unnecessary anxiety caused by inadequate hospital parking. All patients and relatives emphasised the importance of ensuring adequate support for those randomised to IMPT, which included telephone or face-to-face contact with a clinical nurse specialist, speech and language therapist and dietician.

Patients' views of toxicity as the primary outcome measure were discussed. In general, participants felt this to be an appropriate and useful measure. We sought to understand the side-effects experienced by patients 6 weeks and 1 year post-treatment. Following this we explored views on the patient-reported outcome questionnaires that we plan to use in the study. We asked opinions about the University of Washington questionnaire as a tool to collect outcome data at different time points. The questionnaire consists of 15 single domains divided into six physical, six socio-emotional and three global questions, which have been validated for use in head and neck cancer trials [14]. To understand if the six physical questions that make up the composite score for the primary outcome measure of the trial were relevant, we compared the side-effects that patients reported as the most important to them with those measured by the University of Washington questionnaire. Of those reported, four of the six physical symptoms (loss of taste, oral dryness, swallowing dysfunction and problems chewing) on the University of Washington scale were described as the most important side-effects 1 year after treatment. The acceptability of the questionnaire was assessed using a written feedback form, which was completed by 30/33 (91%) patients. Twenty-eight (93%) of the patients thought the six physical questions were a good primary outcome measure for the trial. All patients stated

the questionnaire and the scale were clear. Twenty-eight patients (93%) thought the areas covered were relevant. Hearing loss was the most common missing symptom, highlighted by 5/30 (17%) patients. Other reported missing symptoms (for the primary outcome composite score) were: psychological issues ($n = 2$), fatigue ($n = 1$) and bone damage ($n = 1$). Nine patients (30%) thought free text boxes were needed. All considered the frequency of completion of questionnaires over 5 years of follow-up to be appropriate and feasible. The favoured method of completion of the questionnaire varied: on paper in clinic (12/37; 32%), on paper at home (11/37; 30%), on a tablet in clinic (8/37; 22%) and online (6/37; 16%).

This was an important piece of work to understand patients' and carers' perceptions about the first proposed proton trial in the UK. The feedback on the patient pathway and logistics was encouraging for the feasibility of the study and is invaluable in shaping the trial design.

References

[1] Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tân PF, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med* 2010;363:24—35. https://doi.org/10.1056/NEJMoa0912217.

[2] O'Sullivan B, Huang SH, Siu LL, Waldron J, Zhao H, Perez-Ordonez B, et al. Deintensification candidate subgroups in human papillomavirus-related oropharyngeal cancer according to minimal risk of distant metastasis. *J Clin Oncol* 2013;31: 543—550. https://doi.org/10.1200/JCO.2012.44.0164.

[3] Blanchard P, Garden AS, Gunn GB, Rosenthal DI, Morrison WH, Hernandez M, et al. Intensity-modulated proton beam therapy (IMPT) versus intensity-modulated photon therapy (IMRT) for patients with oropharynx cancer — a case matched analysis. *Radiother Oncol* 2016;120:48—55. https://doi.org/10.1016/j.radonc.2016.05.022.

[4] Gunn GB, Blanchard P, Garden AS, Zhu XR, Fuller CD, Mohamed AS, et al. Clinical outcomes and patterns of disease recurrence after intensity modulated proton therapy for oropharyngeal squamous carcinoma. *Int J Radiat Oncol Biol Phys* 2016;95:360—367. https://doi.org/10.1016/j.ijrobp.2016.02.021.

[5] Sio TT, Lin H-K, Shi Q, Gunn GB, Cleeland CS, Lee JJ, et al. Intensity modulated proton therapy versus intensity modulated photon radiation therapy for oropharyngeal cancer: first comparative results of patient-reported outcomes. *Int J Radiat Oncol Biol Phys* 2016;95:1107—1114. https://doi.org/10.1016/j.ijrobp.2016.02.044.

[6] Zhang W, Zhang X, Yang P, Blanchard P, Garden AS, Gunn B, et al. Intensity-modulated proton therapy and osteoradionecrosis in oropharyngeal cancer. *Radiother Oncol* 2017; 123:401—405. https://doi.org/10.1016/j.radonc.2017.05.006.

[7] https://clinicaltrials.gov/ct2/show/NCT01893307. [Accessed December 2017].

[8] Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide SA, Clark C, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. *Lancet Oncol* 2011;12:127—136. https://doi.org/10.1016/S1470-2045(10)70290-4.

[9] Haviland JS, Agrawal R, Aird E, Barrett J, Barrett-Lee P, Brown J, et al. The UK START (Standardisation of Breast Radiotherapy) Trials: 10-year follow-up results. *Cancer Res* 2012;72:2203. https://doi.org/10.1158/0008-5472.SABCS12-S4-1.

[10] Dearnaley D, Syndikus I, Mossop H, Khoo V, Birtle A, Bloomfield D, et al. Conventional versus hypofractionated high-dose intensity-modulated radiotherapy for prostate cancer: 5-year outcomes of the randomised, non-inferiority, phase 3 CHHiP trial. *Lancet Oncol* 2016;17:1047—1060. https://doi.org/10.1016/S1470-2045(16)30102-4.

[11] Mockford C, Staniszewska S, Griffiths F, Herron-Marx S. The impact of patient and public involvement on UK NHS health care: a systematic review. *Int J Qual Health Care* 2012;24: 28—38. https://doi.org/10.1093/intqhc/mzr066.

[12] Gasson S, Bliss J, Jamal-Hanjani M, Krebs M, Swanton C, Wilcox M. The value of patient and public involvement in trial design and development. *Clin Oncol* 2015;27:747—749. https://doi.org/10.1016/j.clon.2015.06.020.

[13] Hughes-Morley A, Hann M, Fraser C, Meade O, Lovell K, Young B, et al. The impact of advertising patient and public involvement on trial recruitment: embedded cluster randomised recruitment trial. *Trials* 2016;17:586. https://doi.org/10.1186/s13063-016-1718-1.

[14] Rogers SN, Lowe D, Yueh B, Weymuller Jr EA. The physical function and social-emotional function subscales of the University of Washington Quality of Life Questionnaire. *Arch Otolaryngol Neck Surg* 2010;136:352. https://doi.org/10.1001/archoto.2010.32.